



**HAL**  
open science

# Maximum likelihood estimation in partially observed Markov models with applications to time series of counts

Tepmony Sim

► **To cite this version:**

Tepmony Sim. Maximum likelihood estimation in partially observed Markov models with applications to time series of counts. Statistics Theory [stat.TH]. Télécom ParisTech, 2016. English. NNT : 2016ENST0020 . tel-01458087

**HAL Id: tel-01458087**

**<https://pastel.hal.science/tel-01458087>**

Submitted on 6 Feb 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



EDITE - ED 130

## Doctorat ParisTech

### THÈSE

pour obtenir le grade de docteur délivré par

**TELECOM ParisTech**

**Spécialité « Signal et Images »**

*présentée et soutenue publiquement par*

**Tepmony SIM**

le 8 mars 2016

## **Estimation du maximum de vraisemblance dans les modèles de Markov partiellement observés avec des applications aux séries temporelles de comptage**

Directeur de thèse : **François ROUEFF**  
Co-encadrement de la thèse : **Randal DOUC**

#### Jury

**M. Randal DOUC**, Professeur, CITI, Télécom SudParis

**Mme Valentine GENON-CATALOT**, Professeur, MAP5, Université Paris Descartes

**M. François LE GLAND**, Directeur de recherche, ASPI, IRISA/INRIA Rennes

**M. Stéphane ROBIN**, Directeur de recherche, MIA Paris, INRA/AgroParisTech

**M. François ROUEFF**, Professeur, LTCI, Télécom ParisTech

**M. Philippe SOULIER**, Professeur, SEGMI, Université Paris Ouest Nanterre

**M. Jean-Michel ZAKOIAN**, Professeur, LFA, CREST

Co-directeur de thèse

Examinatrice

Rapporteur

Examinateur

Directeur de thèse

Examinateur

Rapporteur

**TELECOM ParisTech**

école de l'Institut Mines-Télécom - membre de ParisTech



# Maximum Likelihood Estimation in Partially Observed Markov Models with Applications to Time Series of Counts

Tepmony Sim

Dissertation Submitted to  
Télécom ParisTech  
in the Fulfilment of the Requirement of the Degree of  
Doctor of Philosophy

Year 2016



# Abstract

Maximum likelihood estimation is a widespread method for identifying a parametrized model of a time series from a sample of observations. Under the framework of well-specified models, it is of prime interest to obtain consistency of the estimator, that is, its convergence to the true parameter as the sample size of the observations goes to infinity. For many time series models, for instance hidden Markov models (HMMs), such a “strong” consistency property can however be difficult to establish. Alternatively, one can show that the maximum likelihood estimator (MLE) is consistent in a weakened sense, that is, as the sample size goes to infinity, the MLE eventually converges to a set of parameters, all of which associate to the same probability distribution of the observations as for the true one. The consistency in this sense, which remains a preferred property in many time series applications, is referred to as *equivalence-class consistency*. The task of deriving such a property generally involves two important steps: 1) show that the MLE converges to the maximizing set of the asymptotic normalized log-likelihood; and 2) show that any parameter in this maximizing set yields the same distribution of the observation process as for the true parameter.

In this thesis, our primary attention is to establish the equivalence-class consistency for time series models that belong to the class of partially observed Markov models (PMMs) such as HMMs and observation-driven models (ODMs). We first establish the result of **Step 2**), that is identifying the maximizing set of the asymptotic normalized log-likelihood, under very mild assumptions and under a general framework of PMMs. The novel contribution of our approach is that, in contrast with previous approaches, the identifiability result is addressed by relying on the uniqueness of the invariant distribution of the Markov chain associated to the complete data, regardless of its rate of convergence to the equilibrium. This result then applies, for example, to HMMs and ODMs in which specific models, namely, the polynomially ergodic Markov chain, the negative binomial integer-valued GARCH (NBIN-GARCH) and the normal mixture GARCH (NM-GARCH) models are illustrated.

Then, we establish the convergence of the MLE in **Step 1)** for two separate classes of models, namely, the class of fully dominated PMMs and the class of ODMs by generalizing the existing results. For fully dominated PMMs, formulating this convergence property is carried out through the approach which consists in establishing a key property—the exponential forgetting of the filtering distribution by using coupling method. For ODMs, this is performed by proving that the process is stationary and ergodic, and by approximating the conditional likelihood by an appropriately defined stationary version of it, which is shown to converge using classical ergodic theory arguments. The stationary and ergodic solutions for the family of ODMs are obtained via the theory of Markov chains without irreducibility assumption. All of these results are addressed under assumptions that are quite direct and easy to check in concrete examples. As illustration, we examine three examples, namely the NBIN-GARCH(1, 1), the NM-GARCH(1, 1) and the threshold IN-GARCH(1, 1) models.

Finally, extension of ODMs to general-order versions is also considered. By incorporating relevant lagged variables into the model, a general-order ODM can be embedded in the classical first-order version of it. In this extended class, the stationarity and ergodicity of the process are studied. The consistency of the MLE is considered. The asymptotic normality of the MLE is also treated in this general framework. All the results are reported under sufficient and easy-to-check conditions. We apply these results to the GARCH( $p, q$ ), the log-linear Poisson autoregression of order ( $p, q$ ) and the NBIN-GARCH( $p, q$ ) models. An empirical study completes this part.

**Keywords:** consistency, ergodicity, hidden Markov models, maximum likelihood, observation-driven models, partially observed Markov models, time series of counts

# Résumé

L'estimation du maximum de vraisemblance est une méthode répandue pour l'identification d'un modèle paramétré de série temporelle à partir d'un échantillon d'observations. Dans le cadre de modèles bien spécifiés, il est primordial d'obtenir la consistance de l'estimateur, à savoir sa convergence vers le vrai paramètre lorsque la taille de l'échantillon d'observations tend vers l'infini. Pour beaucoup de modèles de séries temporelles, par exemple les modèles de Markov cachés ou « hidden Markov models » (HMM), la propriété de consistance « forte » peut cependant être difficile à établir. On peut alors s'intéresser à la consistance de l'estimateur du maximum de vraisemblance (EMV) dans un sens faible, c'est-à-dire que lorsque la taille de l'échantillon tend vers l'infini, l'EMV converge vers un ensemble de paramètres qui s'associent tous à la même distribution de probabilité des observations que celle du vrai paramètre. La consistance dans ce sens, qui reste une propriété privilégiée dans beaucoup d'applications de séries temporelles, est dénommée *consistance de classe d'équivalence*. L'obtention de la consistance de classe d'équivalence exige en général deux étapes importantes : 1) montrer que l'EMV converge vers l'ensemble qui maximise la log-vraisemblance normalisée asymptotique ; et 2) montrer que chaque paramètre dans cet ensemble produit la même distribution du processus d'observation que celle du vrai paramètre.

Cette thèse a pour objet principal d'établir la consistance de classe d'équivalence pour des modèles de séries temporelles qui appartiennent à la classe des modèles de Markov partiellement observés, ou « partially observed Markov models » (PMM), comme les HMM et les modèles « observation-driven » (ODM). Nous établissons le résultat de l'étape 2), qui identifie l'ensemble qui maximise la log-vraisemblance normalisée asymptotique, sous des hypothèses assez faibles et dans un cadre général de PMM. La nouveauté de notre approche est que, contrairement aux approches précédentes, le résultat d'identifiabilité est traité en se basant sur l'unicité de la distribution stationnaire de la chaîne de Markov associée aux données entières, sans tenir compte de son taux de convergence vers l'équilibre. Nous appliquons ensuite



ce résultat, par exemple, aux HMM et ODM, dans lesquels des modèles spécifiques, à savoir la chaîne de Markov ergodique polynomialement, le modèle « negative binomial integer-valued GARCH » (NBIN-GARCH) et le modèle GARCH gaussien mélangé ou « normal mixture GARCH » (NM-GARCH), sont détaillés.

Nous établissons ensuite la convergence de l'EMV dans l'étape 1) pour deux classes de modèles distinctes, d'une part la classe des PMMs entièrement dominés et d'autre part la classe des ODM, en généralisant les résultats obtenus. Pour les PMMs entièrement dominés, l'établissement de cette propriété de convergence est faite selon l'approche qui consiste à établir une propriété clé : l'oubli exponentiel de la distribution de filtrage en utilisant la méthode de couplage. Pour la classe des ODM, ceci est fait en montrant que le processus est stationnaire et ergodique, et en donnant une approximation de la vraisemblance conditionnelle par sa version stationnaire convenablement définie, dont nous montrons qu'elle converge selon des arguments classiques relevant de la théorie ergodique. Les solutions stationnaires et ergodiques pour la famille des ODM sont obtenues via la théorie des chaînes de Markov sans condition d'irréductibilité. Tous ces résultats sont traités sous des hypothèses relativement directes et faciles à vérifier dans des exemples concrets. Pour illustration, nous examinons trois exemples, qui sont les modèles NBIN-GARCH(1, 1), NM-GARCH(1, 1) et IN-GARCH à seuil de l'ordre (1, 1).

L'extension des ODM à des versions d'ordre général est aussi considérée. En intégrant des variables décalées pertinentes dans le modèle, un ODM d'ordre général peut avoir une représentation classique d'ODM de premier ordre. Dans cette classe élargie, la stationnarité et l'ergodicité du processus sont étudiées. La consistance de l'EMV est considérée. La normalité asymptotique de l'EMV est également traitée dans ce cadre général. Tous les résultats sont obtenus sous des conditions suffisantes facilement vérifiables. Nous appliquons ces résultats au modèle GARCH( $p, q$ ), à l'autorégression log-linéaire de Poisson d'ordre ( $p, q$ ), et au modèle NBIN-GARCH( $p, q$ ). Une étude empirique complète cette partie.

**Mots-clés :** consistance, ergodicité, maximum de vraisemblance, modèles de Markov cachés, modèles de Markov partiellement observés, observation-driven, séries temporelles de comptage

*To my parents and relatives, those who educate me and those who bring  
peace.*

I hereby declare that this thesis entitled “*Maximum Likelihood Estimation in Partially Observed Markov Models with Applications to Time Series of Counts*” is the result of my own research except as cited in the references. This thesis has not been accepted for any degree and is not concurrently submitted in candidature for any other degree.

# Acknowledgements

I would like to express my heartfelt gratitude to both of my profound Ph.D. advisors, François Roueff, Professor at Télécom ParisTech, and Randal Douc, Professor at Télécom SudParis, for their constant encouragement, kindness, support, patience, caring and excellent guidance during these three years. François and Randal have not only been great thesis advisors, but also great mentors and companies. I have learned a lot from them so that I can still greatly benefit for the years to come. I will never forget their efforts in seeking for financial grant for my Ph.D. thesis and finding a very beautiful and comfortable accommodation in Paris for me.

I would like to greatly thank Professor François Le Gland and Professor Jean-Michel Zakoian for agreeing to review and to write the report on this manuscript. Many thanks go to Professor Valentine Genon-Catalot, Professor Philippe Soulier and Dr. Stéphane Robin for being the members of the jury.

I am grateful to my current employer in Cambodia, Dr. Romny Om, Director General of Institute of Technology of Cambodia (ITC), for encouraging and granting me the permission to pursue my Ph.D. degree during these three years. Without his permission, I will not be able to resume my old position at ITC after finishing my study. I am also thankful to the four deputy directors of ITC, Dr. Thavarith Chunhieng, Mr. Sothan Nuth, Mr. Norith Phol and Mr. San Penh, for their kind encouragement and support.

I am indebted with H.E. Mr. Sophann Ket, Delegation Permanent of the Kingdom of Cambodia to UNESCO, and Director of la Maison du Cambodge, for his encouragement and support on my accommodation issue.

I would like to thank all friends and colleagues from numerous places in Cambodia (at ITC, in Phnom Penh and elsewhere), especially Mrs. Sopheap Khvay, and in France (at Télécom ParisTech, in Paris and elsewhere) for their helps, encouragement, care, sharing funs and helpful discussions. I would like to specially thank Marine for her French language support.

Last but most importantly, I want to express my deepest thanks to my parents and relatives, for their tolerance, understanding and love.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Background . . . . .	1
1.2	Motivation and Problem Statements . . . . .	3
1.2.1	Consistency . . . . .	3
1.2.2	General-order Observation-driven Models . . . . .	5
1.3	Approaches and Main Results . . . . .	5
1.3.1	Identification of the Maximizing Set $\Theta_*$ . . . . .	6
1.3.2	Consistency of the MLE for Fully Dominated and Partially Observed Markov Models . . . . .	6
1.3.3	Consistency of the MLE for Observation-driven Models . . . . .	7
1.3.4	Generalizations of Observation-driven Models . . . . .	8
1.4	Organization of the Thesis . . . . .	10
<b>2</b>	<b>Partially Observed Markov Chains: Identification of the Maximizing Set of the Asymptotic Normalized Log-likelihood</b>	<b>11</b>
2.1	Introduction . . . . .	11
2.2	A General Approach to Identifiability . . . . .	15
2.2.1	General Setting and Notation: Partially Dominated and Partially Observed Markov Chains . . . . .	15
2.2.2	Main Result . . . . .	18
2.2.3	Construction of the Kernel $\Phi^{\theta, \theta'}$ as a Backward Limit . . . . .	22
2.3	Application to Hidden Markov Models . . . . .	25
2.3.1	Definitions and Assumptions . . . . .	25
2.3.2	Equivalence-class Consistency . . . . .	28
2.3.3	A Polynomially Ergodic Example . . . . .	29
2.4	Application to Observation-driven Models . . . . .	32
2.4.1	Definitions and Notation . . . . .	32
2.4.2	Identifiability . . . . .	34
2.4.3	Examples . . . . .	35
2.5	Postponed Proofs . . . . .	36

2.5.1	Proof of Eq. (2.10)	36
2.5.2	Proof of Lemma 2.2.9	37
2.5.3	Proof of Lemma 2.3.9	38
<b>3</b>	<b>Consistency of the MLE for the Class of Fully Dominated Markov Chains</b>	<b>41</b>
3.1	Introduction	41
3.2	Consistency of the Block-type Maximum Likelihood Estimators	43
3.2.1	Notation and Definitions	43
3.2.2	Forgetting of Initial Distribution for the Conditional $L$ -likelihood	51
3.2.3	Convergence of the Maximum $L$ -likelihood Estimator	53
3.3	Consistency of the MLE for Fully Dominated Markov Models	55
3.4	Postponed Proofs	67
3.4.1	Proof of Proposition 3.2.11	67
3.4.2	Proof of Proposition 3.2.13	72
3.4.3	Proof of Lemma 3.3.5	75
3.4.4	Proof of Proposition 3.3.9	76
3.4.5	Some Useful Lemmas	80
<b>4</b>	<b>Observation-driven Models: Handy Sufficient Conditions for the Convergence of the MLE</b>	<b>82</b>
4.1	Introduction	82
4.2	Definitions and Notation	85
4.3	Main Results	88
4.3.1	Preliminaries	88
4.3.2	Convergence of the MLE	90
4.3.3	Ergodicity	91
4.4	Examples	94
4.4.1	NBIN-GARCH Model	94
4.4.2	NM-GARCH Model	96
4.4.3	Threshold IN-GARCH Model	101
4.5	Numerical Experiments	103
4.5.1	Numerical Procedure	103
4.5.2	Simulation Study	104
4.6	Postponed Proofs	108
4.6.1	Convergence of the MLE	108
4.6.2	Ergodicity	110
4.6.3	Proof of Lemma 4.6.1.	113

<b>5</b>	<b>General-order Observation-driven Models: Ergodicity, Consistency and Asymptotic Normality of the MLE</b>	<b>117</b>
5.1	Introduction . . . . .	117
5.2	Definitions and Notation . . . . .	120
5.3	Main Results . . . . .	123
5.3.1	Preliminaries . . . . .	123
5.3.2	Convergence of the MLE . . . . .	125
5.3.3	Asymptotic Normality of the MLE . . . . .	127
5.3.4	Ergodicity . . . . .	133
5.4	Examples . . . . .	137
5.4.1	GARCH( $p, p$ ) Model . . . . .	138
5.4.2	Log-linear Poisson GARCH( $p, p$ ) Model . . . . .	142
5.4.3	NBIN-GARCH( $p, p$ ) Model . . . . .	156
5.5	Empirical Study . . . . .	168
5.5.1	Sharpness of Ergodicity Condition (5.56) for the Log-linear Poisson GARCH Model . . . . .	168
5.5.2	Data Example . . . . .	171
5.6	Postponed Proofs . . . . .	175
5.6.1	Proof of Lemma 5.3.4 . . . . .	175
5.6.2	Proof of Theorem 5.3.12 . . . . .	181
5.6.3	Proof of Lemma 5.4.2 . . . . .	186
5.6.4	Proof of Lemma 5.4.11 . . . . .	187
<b>6</b>	<b>Conclusions and Future Perspectives</b>	<b>188</b>
	<b>Résumé en français</b>	<b>190</b>
	<b>Bibliography</b>	<b>244</b>

## List of Figures

2.1	Graphical Representation of the Partially Observed Markov Model . . . . .	16
2.2	Graphical Representation of the Hidden Markov Model . . . . .	25
2.3	Graphical Representation of the Observation-driven Model . . . . .	33
3.1	Graphical Representation of the Hidden Markov Model . . . . .	45

3.2	Graphical Representation of the Markov Chain in Random Environment . . . . .	47
3.3	Graphical Representation of the Markov Switching Model . . .	48
3.4	Graphical Representation of the Fully Dominated and Partially Observed Markov Model . . . . .	56
4.1	Graphical Representation of the Observation-driven Model. . .	87
4.2	Boxplots of the Difference of Log-likelihoods . . . . .	106
4.3	Boxplots of the Estimated MLE for Model (M.1) . . . . .	107
4.4	Boxplots of the Estimated MLE for Model (M.2) . . . . .	108
5.1	Simulated Time Series for Log-linear Poisson GARCH of Order (1, 1) . . . . .	170
5.2	Simulated Time Series for Log-linear Poisson GARCH of Order (2, 2) . . . . .	171
5.3	Time Series of Number of Earthquakes in 2014 (Hourly) . . .	173
5.4	Serial Autocorrelation for Earthquake Data . . . . .	174
5.5	Prediction of NBIN-GARCH models . . . . .	174
5.6	Prediction of Log-linear Poisson GARCH Models . . . . .	175

## List of Tables

4.1	Mean of Estimates and MADEs for the NBIN-GARCH(1, 1) .	105
5.1	AIC, BIC and ASPE Values for the NBIN-GARCH and the Log-linear Poisson GARCH with Different Orders . . . . .	173



# Chapter 1

## Introduction

### 1.1 Background

A partially observed Markov model (PMM) is a bivariate Markov process such that only one part of the complete process is observed—hence “partially observed” and the other component, the state variable, is “hidden” or “unobserved”—and any statistical inference has to be carried out by means of using a sample of realizations of the observation process. Notably, the class of PMMs contains two prominent classes of statistical models commonly used in examining serial-dependence sequences and time series data, namely, the class of hidden Markov models (HMMs) and the class of observation-driven models (ODMs). A common yet important feature of these two particular subclasses lies on the fact that the hidden processes of the state variables are Markov chains in their own right. This feature permits establishing some important statistical properties such as stationarity and ergodicity for the complete chains or models by relying on those inherited statistical properties of the unobserved chains.

HMMs are applied widely in various areas ranging from biology ([Churchill \[1992\]](#), [Krogh et al. \[1994\]](#)), econometrics ([Hull and White \[1987\]](#), [Mamon and Elliott \[2007\]](#)), finance ([Zhang \[2004\]](#)), neurophysiology ([Fredkin and Rice \[1987\]](#)), signal processing ([Crouse et al. \[1998\]](#), [Krishnamurthy and Chung \[2007\]](#)), speech recognition ([Rabiner \[1989\]](#), [Huang et al. \[1990\]](#), [Juang and Rabiner \[1991\]](#)) to time series analysis ([MacDonald and Zucchini \[1997\]](#)). Theory and inference for this class of models were developed by Baum and coworkers in a series of contributions [Baum and Petrie \[1966\]](#), [Baum and Eagon \[1967\]](#), [Baum and Sell \[1968\]](#), [Petrie \[1969\]](#), [Baum et al. \[1970\]](#), [Baum \[1972\]](#) (see also [Ephraim and Merhav \[2002\]](#) for a brief historical note on HMMs). These studies have later been further considered and developed

in a series of works by [Finesso \[1991\]](#), [Leroux \[1992\]](#), [Mével \[1997\]](#), [Bickel et al. \[1998\]](#), [Jensen and Petersen \[1999\]](#), [Le Gland and Mével \[2000\]](#), [Douc and Matias \[2001\]](#), [Bickel et al. \[2002\]](#), [Mével and Finesso \[2004\]](#), [Genon-Catalot and Laredo \[2006\]](#), [Douc et al. \[2009, 2011\]](#), [Douc and Moulines \[2012\]](#) (see also [Cappé et al. \[2005\]](#) and the references therein for a comprehensive discussion). Although there have been substantial studies for this important class, so far some theoretical results have not become standard yet and thus still need to be improved.

ODMs were introduced by [Cox \[1981\]](#) to describe statistical models with a key feature that the hidden state variable can be represented explicitly as a deterministic function of the past observation variables. This property is of practical interest, since the associated (conditional) likelihood function and its derivatives are easy to derive, the parameter estimation is hence relatively simple, and the prediction, which is of prime importance in many time series applications, is straightforward. The celebrated generalized autoregressive conditional heteroskedasticity of order (1,1) model (well known as GARCH(1,1) model) introduced by [Bollerslev \[1986\]](#), as well as most of the models derived from this one are typical examples of ODMs (see [Bollerslev \[2008\]](#) for a list of some of them). In addition, these models are particularly useful for dealing with time series data with discrete nature, while the classical autoregressive models are no longer appropriate (see [[Kedem and Fokianos, 2002](#), Chapter 4] for a comprehensive account and [Tjøstheim \[2012\]](#) for a recent survey). The applications of ODMs can be found in a wide variety of disciplines such as economics ([Pindyck and Rubinfeld \[1998\]](#)), environmental study ([Bhaskaran et al. \[2013\]](#)), epidemiology and public health study ([Zeger \[1988\]](#), [Davis et al. \[1999\]](#), [Ferland et al. \[2006\]](#)), finance ([Liesenfeld and Richard \[2003\]](#), [Rydberg and Shephard \[2003\]](#), [Fokianos and Tjøstheim \[2011\]](#), [Francq and Zakoian \[2011\]](#)) and population dynamics ([Ives et al. \[2003\]](#)). Theory and inference for this class of models when the observations are integer-valued have been developed in a series of contributions including [Streett \[2000\]](#), [Davis et al. \[2003\]](#), [Heinen \[2003\]](#), [Ferland et al. \[2006\]](#), [Fokianos et al. \[2009\]](#), [Franke \[2010\]](#), [Woodard et al. \[2010\]](#), [Fokianos and Tjøstheim \[2011\]](#), [Henderson et al. \[2011\]](#), [Neumann \[2011\]](#), [Davis and Liu \[2012\]](#), [Doukhan et al. \[2012\]](#), [Douc et al. \[2013\]](#), [Fokianos et al. \[2013\]](#), [Christou and Fokianos \[2014, 2015\]](#) and [Douc et al. \[2015\]](#). As in the case of HMMs, some of existing theoretical results for these models have not become standard yet and thus still need to be ameliorated.

Besides HMMs and ODMs, the class of PMMs include many other interesting models that have potential applications in time series and other fields such as Markov switching models (see [Hamilton \[1988, 1989\]](#)), many of Markov chains in random environment introduced by [Cogburn \[1980\]](#) and

Bourgin and Cogburn [1981], pairwise Markov random fields (see Pieczynski and Tebbache [2000]) and other instances considered e.g. by Pieczynski [2003] and Ephraim and Mark [2012]. However, the theory and inference for this general setting still remains meagre in the literature.

## 1.2 Motivation and Problem Statements

This thesis aims at developing theory and inference for parametric models concerning in particular time series of counts under general frameworks of PMMs and other variants. It is mainly driven by the following issues encountered in HMMs and ODMs in particular (and hence in PMMs in general).

### 1.2.1 Consistency

Maximum likelihood estimation is a natural and widespread method for identifying a parametric model of a time series from a sample of realizations of the observation process. Under a well-specified model setting, that is when the law of the observations is fully described by a parametrized member of the class of PMMs, it is of prime interest to obtain the consistency of the estimator, namely, its convergence to the true parameter, say  $\theta_*$ , as the sample size goes to infinity. The proof generally involves two important steps:

1. the maximum likelihood estimator (MLE) converges to the maximizing set  $\Theta_*$  of the asymptotic normalized log-likelihood, and
2. this set  $\Theta_*$  indeed reduces to the singleton of the true parameter  $\theta_*$ .

The second step is usually referred to as solving the *identifiability* problem, but it can actually be split in two sub-problems:

- 2.1 show that any parameter in the maximizing set  $\Theta_*$  yields the same distribution for the observations as for the true parameter  $\theta_*$ , and
- 2.2 show that for a sufficiently large sample size, the set of such parameters reduces to the singleton of  $\theta_*$ .

However, at least in the case of HMMs, Problem 2.2 above appears to be difficult to solve; see Allman et al. [2009], Gassiat et al. [2016] and the references therein for recent advances in the case of HMMs. Indeed, Problem 2.1 can be solved independently, and with Step 1 above, this directly yields that the MLE is consistent in a weakened sense, namely, that the estimated parameter converges to the set of all the parameters associated to the same distribution as the one of the observed sample. The consistency in this sense is referred to as *equivalence-class consistency*, as introduced by Leroux [1992].

### *Consistency issue in HMMs*

For HMMs, the consistency of the MLE is of primary importance; it arises either as a subject of study (see [Baum and Petrie \[1966\]](#), [Petrie \[1969\]](#), [Leroux \[1992\]](#), [Le Gland and Mevel \[2000\]](#), [Le Gland and Mevel \[2000\]](#), [Douc and Matias \[2001\]](#), [Douc et al. \[2004\]](#), [Genon-Catalot and Laredo \[2006\]](#), [Douc et al. \[2011\]](#)) or as a basic assumption (see [Bickel et al. \[1998\]](#), [Jensen and Petersen \[1999\]](#)). Among these contributions, the consistency result obtained later in [Douc et al. \[2011\]](#) appears to be quite general and can be applied to many versions of HMMs with some general state and observation spaces. In this contribution, identifying the maximizing set  $\Theta_*$  of the asymptotic normalized log-likelihood is carried out by using the so-called *exponential separation of measures* condition, which is apparently easy to apply when the underlying model is geometrically ergodic. However, for non-geometrically ergodic models such as the polynomially ergodic Markov chain (see [Tuominen and Tweedie \[1994\]](#) and [Järner and Roberts \[2002\]](#) for the existence and some studies of such an example), checking this condition can be laborious and involved. Hence the characterization of the maximizing set  $\Theta_*$  of the asymptotic normalized log-likelihood (and hence the equivalence-class consistency of the MLE) remains a delicate question for HMMs.

### *Consistency issue in ODMs*

In ODMs, the (conditional) likelihood function and its derivatives are easy to compute, making the estimation at least numerically feasible. The consistency of the MLE can however be cumbersome and is often derived using computations specific to the studied model. Especially, when the observed variable is discrete, general consistency results have been obtained only recently in [Davis and Liu \[2012\]](#) or [Douc et al. \[2013\]](#) (see also in [Henderson et al. \[2011\]](#) for the existence of stationary and ergodic solutions to some ODMs). Nevertheless, in these contributions, the way of proving that the maximizing set  $\Theta_*$  reduces to the singleton of the true parameter  $\theta_*$  requires checking specific conditions in each given example and seems difficult to assert the result in a more general context, for instance when the distribution of the observations given the hidden variable also depends on an unknown parameter. The examples of such models include the negative binomial integer-valued GARCH (NBIN-GARCH) model introduced in [Zhu \[2011\]](#) and the normal mixture GARCH (NM-GARCH) model proposed by [Haas et al. \[2004\]](#) and [Alexander and Lazar \[2006\]](#). To our best knowledge, (complete) consistency of the MLE has not been treated for the NBIN-GARCH model (see [Davis and Liu \[2012\]](#) for some partial treatment); and

the usual consistency proof of the MLE for the GARCH model cannot be directly adapted to the NM-GARCH model.

It is thus of great interest if this consistency issue can be tackled under the general spirit of PMMs, and if the present approach is able to account for the intractable versions of HMMs and ODMs mentioned hereinabove.

### 1.2.2 General-order Observation-driven Models

In many time series, solving Problem 1 and Problem 2.1 in Section 1.2.1, that is, establishing the consistency of the MLE, usually requires assuming or proving that the observation process is stationary and ergodic. Under HMM framework, stationary and ergodic solutions for the model are inherited from the stationarity and ergodicity of the underlying hidden process and are usually studied by using  $\varphi$ -irreducibility assumption. However, it may not be at all the same fortune when the studied model is under ODM setting. In this latter case, it turns out that a more subtle approach is required, at least in the circumstance where the observed process assumes integer values. The difficulty in this case, though in its own right it is a Markov chain, may arise from the degeneracy of the hidden state process from its state equation. Up to this point, several methods toward the solution of this problem has been proposed. These include the perturbation technique (see Fokianos and Tjøstheim [2011]), the contractivity approach (see Neumann [2011]), the weak dependence approach (see Doukhan et al. [2012]) and the approach based on the theory of Markov chains without irreducibility assumption introduced in Douc et al. [2013]. Among these, the result obtained recently in Douc et al. [2013] appears to be able to cope with many models of interest lying in the class of *first-order* ODMs, regardless of whether the observation process is discrete or continuous. However, whether this same result applies to a more general and flexible context of *general-order* ODMs or general-order GARCH type models has not been known so far. The same question is for the asymptotic properties of the MLE for the class of general-order ODMs as posed by Douc et al. [2013] and recently addressed in Tjøstheim [2015].

## 1.3 Approaches and Main Results

Throughout this thesis, the statistical inference is performed, unless otherwise specified, under the framework of the well-specified models. The approaches and main results are outlined as follows.

### 1.3.1 Identification of the Maximizing Set $\Theta_*$

One of our primary objectives is to provide a general method toward solving Problem 2.1 as described in Section 1.2.1, that is, showing that any parameter in the maximizing set  $\Theta_*$  of the asymptotic normalized log-likelihood yields the same distribution for the observations as for the true parameter  $\theta_*$ , under the general framework of *partially dominated* and partially observed Markov models in which many interesting models such as HMMs and ODMs are included. Here, by “partially dominated”, we refer to a situation where the distribution of the observed variable is dominated by some fixed  $\sigma$ -finite measure defined on the observation space. The proof of the main result is of probabilistic-analytic nature. The novel contribution of our present approach is that the characterization of the maximizing set  $\Theta_*$  is addressed by relying mainly on the existence and the uniqueness of the invariant distribution of the Markov chain associated to the complete data, regardless of its rate of convergence to the equilibrium. This is indeed in deep contrast with existing results in these models where the identification of  $\Theta_*$  is established under the assumption of exponential separation of measures (see Douc et al. [2011]) or geometric ergodicity (see Douc et al. [2004]). The ways of how this general result applies in the classes of HMMs and ODMs are also demonstrated. As it will be later shown, our general result can be readily applied to solve Problem 2.1 for some otherwise intractable versions of HMMs and ODMs mentioned earlier in Section 1.2.1. All of these are reported in the journal paper by Douc et al. [2014], which is going to appear soon in *The Annals of Applied Probability*.

### 1.3.2 Consistency of the MLE for Fully Dominated and Partially Observed Markov Models

In addition to solving Problem 2.1, our next objective is to investigate Problem 1 within the general context of PMMs. In this thesis, however, this task is performed separately in two directions; namely, we discuss this issue independently for the class of *fully dominated* PMMs and for the class of ODMs. A fully dominated PMM is a partially observed Markov model whose conditional marginal laws of the observation and hidden state processes both admit probability densities with respect to some  $\sigma$ -finite measures. Many HMMs, Markov switching models and Markov chains in random environment, for example, are members of this class. Solving Problem 1 for fully dominated PMM, that is, formulating the convergence of the MLE, is carried out by using the approach employed by Douc and Moulines [2012] as to derive the convergence of the MLE to the minimizing set of the relative

entropy rate in misspecified HMMs. This approach consists in establishing a key property—the exponential forgetting of the filtering distribution by using coupling method originated by Kleptsyna and Veretennikov [2008] and further refined by Douc et al. [2009]. When the forgetting of the filtering distribution is achieved, together with the assumption that the observation process is strict-sense stationary and ergodic, then the normalized log-likelihood can be approximated by an appropriately defined stationary version of it. Then applying the results in classical ergodic theory, the latter is shown to converge to a limit of functional of the parameter, say  $\ell(\theta)$ , where  $\theta$  is the parameter. It then implies by using standard argument (see for instance Pfanzagl [1969]) that the MLE of the normalized log-likelihood converges almost surely to a subset  $\Theta_\star$  of the parameter set at which this limit  $\ell(\theta)$  is maximized. Using this same technique, meanwhile, we obtain an intermediary result showing that the block-type MLE converges to a maximizing set in a similar sense as the classical MLE, regardless of whether or not the models are well-specified. These results are shown under similar conditions derived in Douc and Moulines [2012]. Moreover, with these same assumptions, Problem 2.1 is also solved, yielding the equivalence-class consistency of the MLE for the class fully dominated PMMs.

### 1.3.3 Consistency of the MLE for Observation-driven Models

In this thesis, we consider among others a typically studied class of ODMs defined in the following context. Let  $(\mathbf{X}, d_{\mathbf{X}})$  be a locally compact, complete and separable metric space endowed with the associated Borel  $\sigma$ -field  $\mathcal{X}$ ,  $(\mathbf{Y}, \mathcal{Y})$  be a Borel space and  $(\Theta, \Delta)$ , the set of parameters, be a compact metric space. The observations  $\{Y_k : k \geq 0\}$  is said to be distributed according to an observation-driven model if there exists a bivariate stochastic process  $\{(X_k, Y_k) : k \geq 0\}$  defined on  $\mathbf{X} \times \mathbf{Y}$  such that for all  $\theta \in \Theta$  and integers  $k \geq 0$ ,

$$\begin{aligned} X_{k+1} &= \psi_{Y_k}^\theta(X_k), \\ Y_{k+1} | X_{0:k+1}, Y_{0:k} &\sim G^\theta(X_{k+1}; \cdot), \end{aligned} \tag{1.1}$$

where  $(x, y) \mapsto \psi_y^\theta(x)$  is a measurable function defined from  $\mathbf{X} \times \mathbf{Y}$  to  $\mathbf{X}$  and  $G^\theta$  is a probability kernel on  $\mathbf{X} \times \mathbf{Y}$ .

Examination of the consistency of the MLE for this family is carried out by relying on the approach recently developed by Douc et al. [2013], who considered Model (1.1) in the case where the probability kernel  $G^\theta$  does not depend on the parameter  $\theta$ . A basic assumption to this approach is the unique-

ness of invariant distribution for the complete chain  $\{(X_k, Y_k) : k \geq 0\}$ . This assumption is obtained by using the theory of Markov chains without irreducibility assumption. The uniqueness of the invariant distribution of the Markov chain  $\{(X_k, Y_k) : k \geq 0\}$  allows us to extend this stationary process over the integer indices and is later shown to yield the stationarity and ergodicity for the observation process  $\{Y_k : k \geq 0\}$ . With some additional assumptions, by similar argument as in the case of fully dominated PMMs, the conditional likelihood can be approximated by an appropriately defined stationary version of it, which is then shown to converge to a limit of functional of the parameter by classical ergodic theory arguments. It then implies by Pfanzagl [1969] that the MLE of the normalized log-likelihood converges almost surely to a subset  $\Theta_*$  of parameters at which this limiting functional is maximized. In addition, Problem 2.1 is investigated and with the earlier convergence result, the equivalence-class consistency of the MLE for this class is then established. All of these results are addressed under assumptions that are quite direct and easy to check in concrete examples. As illustration, we examine three examples, namely, the NBIN-GARCH(1, 1), the NM-GARCH(1, 1) and the threshold IN-GARCH(1, 1). A numerical simulation demonstrating the convergence of the MLE for NBIN-GARCH(1, 1) is also given. All of these results are also reported in the published paper by Douc et al. [2015].

### 1.3.4 Generalizations of Observation-driven Models

Another focus of this thesis is on developing theory and inference for the class of GARCH( $p, q$ ) analogue ODMs. In this consideration, we extend (1.1) by allowing the function  $\psi^\theta$  to depend also on the past lagged variables; that is, we replace the function  $\psi_y^\theta(x)$  defined from  $\mathbf{X} \times \mathbf{Y}$  to  $\mathbf{X}$  by a function  $\psi_{y_1, \dots, y_q}^\theta(x_1, \dots, x_p)$  defined from  $\mathbf{X}^p \times \mathbf{Y}^q$  to  $\mathbf{X}$ , for some positive integers  $p$  and  $q$ . This yields the following setting: for all  $\theta \in \Theta$  and integers  $k \geq 0$ ,

$$\begin{aligned} X_{k+1} &= \psi_{Y_{k-q+1}, \dots, Y_k}^\theta(X_{k-p+1}, \dots, X_k), \\ Y_{k+1} | X_{-p+1:k+1}, Y_{-q+1:k} &\sim G^\theta(X_{k+1}; \cdot). \end{aligned} \tag{1.2}$$

We say that the observations  $\{Y_k : k \geq -q + 1\}$  taking values in  $(\mathbf{Y}, \mathcal{Y})$  is distributed according to an observation-driven model of order  $(p, q)$ , in short ODM( $p, q$ ), if there exists a stochastic process  $\{X_k : k \geq -p + 1\}$  defined on  $(\mathbf{X}, \mathcal{X})$  such that the recursive relation (1.2) holds. Despite that this is just a generalization, however, it is not always obvious that the results obtained in the classical first-order ODM can be directly or easily applied to the general-order setting. One reason may be due to the fact that the general-order ODM



lacks of a direct and useful first-order ODM representation. Nevertheless, this is not a serious problem since generally any general-order ODM can be embedded in a first-order one. To see how, a subtle way is to incorporate the trivial identity of lagged variables,

$$(X_{k-p+2}, \dots, X_k, Y_{k-q+2}, \dots, Y_k) = (X_{k-p+2}, \dots, X_k, Y_{k-q+2}, \dots, Y_k),$$

into the recursive relation (1.2) and set

$$Z_k := (X_{k-p+1}, \dots, X_k, Y_{k-q+1}, \dots, Y_{k-1}),$$

$$\Psi_{Y_k}^\theta(Z_k) := \underbrace{(X_{k-p+2}, \dots, X_k, \psi_{Y_{k-q+1}, \dots, Y_k}^\theta(X_{k-p+1}, \dots, X_k))}_p, \underbrace{(Y_{k-q+2}, \dots, Y_k)}_{q-1}$$

and

$$H^\theta(Z_k; \cdot) := G^\theta(X_k; \cdot).$$

Then (1.2) can be transformed into the following equivalent one: for all  $\theta \in \Theta$  and integers  $k \geq 0$ ,

$$\begin{aligned} Z_{k+1} &= \Psi_{Y_k}^\theta(Z_k), \\ Y_{k+1} | Z_{0:k+1}, Y_{0:k} &\sim H^\theta(Z_{k+1}; \cdot). \end{aligned} \tag{1.3}$$

The form (1.3) is indeed a first-order ODM, but on the space  $(\mathbf{X}^p \times \mathbf{Y}^{q-1}) \times \mathbf{Y}$ . With this first-order representation, the results obtained in (1.1) may be applied. By simple adaptations, the equivalence-class consistency of the MLE for the general-order ODM or for (1.3) directly follows. However, it turns out that more general assumptions may be required to establish the stationary and ergodic solutions for (1.3), compared to the usual first-order ODMs defined by (1.1). Additionally, the strong consistency and the asymptotic normality of the MLE are also investigated in this general context. We treat the asymptotic normality in a special but an important case where the function  $\Psi^\theta$  is linear with respect to the hidden state variable  $Z_k$ . For establishing the asymptotic normality result, we follow the classical approach by first approximating the score function by the stationary version of it and then developing Taylor expansion of the stationary score function around the true parameter  $\theta_*$ . By appropriate assumptions, the central limit theorem for Martingale difference applies and the stationary score function can then be shown to be asymptotically Gaussian. Then by invertibility of the asymptotic Fisher information matrix, the asymptotic normality immediately follows. All the results are reported under sufficient and easy-to-check conditions. We apply these results to the GARCH( $p, q$ ), the log-linear Poisson autoregression of order ( $p, q$ ) and the NBIN-GARCH( $p, q$ ) models. We provide an empirical study showing that in some circumstance higher-order model may fit the data better than the first-order one if the class of ODMs is used.

## 1.4 Organization of the Thesis

This thesis is organized as follows. We devote **Chapter 2** to identifying the maximizing set of the asymptotic normalized log-likelihood under the general framework of PMMs mentioned in **Section 1.3.1**. In **Chapter 3**, we present approaches to obtain the convergence and the consistency of the MLE under fully dominated versions of PMMs. **Chapter 4** is dedicated to studying ergodicity and consistency of MLE for the class of first-order ODMs. **Chapter 5** extends the results of **Chapter 4** to the class of general-order ODMs. Conclusions and future perspectives of the thesis are presented in **Chapter 6**.

## Chapter 2

# Partially Observed Markov Chains: Identification of the Maximizing Set of the Asymptotic Normalized Log-likelihood

**Summary** This chapter, mostly inspired by [Douc et al. \[2014\]](#), considers a parametrized family of partially observed bivariate Markov chains. We establish that, under very mild assumptions, the limit of the normalized log-likelihood function is maximized when the parameters belong to the equivalence class of the true parameter, which is a key feature for obtaining the consistency of the maximum likelihood estimator (MLE) in well-specified models. This result is obtained in the general framework of partially dominated models. We examine two specific cases of interest, namely, hidden Markov models (HMMs) and observation-driven models. In contrast with previous approaches, the identifiability is addressed by relying on the uniqueness of the invariant distribution of the Markov chain associated to the complete data, regardless of its rate of convergence to the equilibrium.

### 2.1 Introduction

Maximum likelihood estimation is a widespread method for identifying a parametric model of a time series from a sample of observations. Under a well-specified model assumption, it is of prime interest to show the consistency of the estimator, that is, its convergence to the true parameter, say

$\theta_*$ , as the sample size goes to infinity. The proof generally involves two important steps: 1) the maximum likelihood estimator (MLE) converges to the maximizing set  $\Theta_*$  of the asymptotic normalized log-likelihood, and 2) the maximizing set indeed reduces to the true parameter. The second step is usually referred to as solving the *identifiability* problem but it can actually be split in two sub-problems: 2.1) show that any parameter in  $\Theta_*$  yields the same distribution for the observations as for the true parameter, and 2.2) show that for a sufficiently large sample size, the set of such parameters reduces to  $\theta_*$ . Problem 2.2 can be difficult to solve; see [Allman et al. \[2009\]](#), [Gassiat et al. \[2016\]](#) and the references therein for recent advances in the case of hidden Markov models (HMMs). Nevertheless, Problem 2.1 can be solved independently, and with Step 1 above, this directly yields that the MLE is consistent in a weakened sense, namely, that the estimated parameter converges to the set of all the parameters associated to the same distribution as the one of the observed sample. This consistency result is referred to as *equivalence-class consistency*, as introduced by [Leroux \[1992\]](#). In this contribution, our goal is to provide a general approach to solve Problem 2.1 in the general framework of partially observed Markov models. These include many classes of models of interest; see, for instance, [Pieczynski \[2003\]](#) or [Ephraim and Mark \[2012\]](#). The novel aspect of this work is that the result mainly relies on the uniqueness of the invariant distribution of the Markov chain associated to the complete data, regardless of its rate of convergence to the equilibrium. We then detail how this approach applies in the context of two important subclasses of partially observed Markov models, namely, the class of HMMs and the class of observation-driven time series models.

In the context of HMMs, the consistency of the MLE is of primary importance, either as a subject of study (see [Leroux \[1992\]](#), [Douc et al. \[2004, 2011\]](#)) or as a basic assumption (see [Bickel et al. \[1998\]](#), [Jensen and Petersen \[1999\]](#)). The characterization of the maximizing set  $\Theta_*$  of the asymptotic log-likelihood (and thus the equivalence-class consistency of the MLE) remains a delicate question for HMMs. As an illustration, we consider the following example. In this example and throughout this thesis, we always denote by  $\mathbb{R} = (-\infty, \infty)$  the set of real numbers and by  $\mathbb{Z}$  the set of integers. In addition, we denote by  $\mathbb{R}_+ = [0, \infty)$ ,  $\mathbb{R}_- = (-\infty, 0]$ ,  $\mathbb{R}_+^* = (0, \infty)$  and  $\mathbb{R}_-^* = (-\infty, 0)$ , the sets of nonnegative, nonpositive, (strictly) positive and (strictly) negative real numbers, respectively. Likewise, we use the notation  $\mathbb{Z}_+$ ,  $\mathbb{Z}_-$ ,  $\mathbb{Z}_+^*$  and  $\mathbb{Z}_-^*$  for the corresponding subsets of integers. For a real number or a real function  $a$ , we denote by  $a^+ = \max(a, 0)$  and by  $a^- = \max(-a, 0)$  the nonnegative part and nonpositive part of  $a$ , respectively. Finally, for any set  $X$ , the collection  $\mathcal{B}(X)$  denotes the Borel  $\sigma$ -field on  $X$ .

**Example 2.1.1.** Set  $\mathsf{X} = \mathbb{R}_+$ ,  $\mathcal{X} = \mathcal{B}(\mathbb{R}_+)$ ,  $\mathsf{Y} = \mathbb{R}$  and  $\mathcal{Y} = \mathcal{B}(\mathbb{R})$  and define an HMM on  $\mathsf{X} \times \mathsf{Y}$  by the following recursions:

$$\begin{aligned} X_k &= (X_{k-1} + U_k - m)^+, \\ Y_k &= aX_k + V_k, \end{aligned} \tag{2.1}$$

where  $(m, a) \in \mathbb{R}_+^* \times \mathbb{R}$ , and the sequence  $\{(U_k, V_k) : k \in \mathbb{Z}_+\}$  is independent and identically distributed (i.i.d.) and is independent from  $X_0$ . This Markov model  $\{X_k : k \in \mathbb{Z}_+\}$  was proposed by [Tuominen and Tweedie \[1994\]](#) and further considered by [Järner and Roberts \[2002\]](#) as an example of polynomially ergodic Markov chain, under specific assumptions made on  $U_k$ 's. Namely, if  $U_k$ 's are centered and  $\mathbb{E}[e^{\lambda U_k^+}] = \infty$  for any  $\lambda > 0$ , it can be shown that the chain  $\{X_k : k \in \mathbb{Z}_+\}$  is not geometrically ergodic (see [Lemma 2.3.9](#) below). In such a situation, the exponential separation of measures condition introduced in [Douc et al. \[2011\]](#) seems difficult to check. We will show, nevertheless, in [Proposition 2.3.10](#), that under some mild conditions the chain  $\{X_k : k \in \mathbb{Z}_+\}$  is ergodic and the equivalence-class consistency holds.

Observation-driven time series models were introduced by [Cox \[1981\]](#) and later considered, among others, by [Streett \[2000\]](#), [Davis et al. \[2003\]](#), [Fokianos et al. \[2009\]](#), [Neumann \[2011\]](#), [Doukhan et al. \[2012\]](#) and [Douc et al. \[2013\]](#). The celebrated generalized autoregressive conditional heteroskedasticity (GARCH) model of order (1, 1) introduced by [Bollerslev \[1986\]](#) is an observation-driven model as well as most of the models derived from this one; see [Bollerslev \[2008\]](#) for a list of some of them. This class of models has the nice feature that the (conditional) likelihood function and its derivatives are easy to compute. The consistency of the MLE can however be cumbersome and is often derived using computations specific to the studied model. When the observed variable is discrete, general consistency results have been obtained only recently in [Davis and Liu \[2012\]](#) or [Douc et al. \[2013\]](#) (see also in [Henderson et al. \[2011\]](#) for the existence of stationary and ergodic solutions to some observation-driven time series models). However, in these contributions, the way of proving that the maximizing set  $\Theta_\star$  reduces to  $\{\theta_\star\}$  requires checking specific conditions in each given example and seems difficult to assert in a more general context, for instance when the distribution of the observations given the hidden variable also depends on an unknown parameter. Let us describe two such examples. The first one ([Example 2.1.2](#)) was introduced in [Zhu \[2011\]](#). To our knowledge, the consistency of the MLE has not been treated for this model.

**Example 2.1.2.** The negative binomial integer-valued GARCH (NBIN-

GARCH) (NBIN-GARCH(1, 1)) model is defined by

$$\begin{aligned} X_{k+1} &= \omega + aX_k + bY_k, \\ Y_{k+1}|X_{0:k+1}, Y_{0:k} &\sim \mathcal{NB}\left(r, \frac{X_{k+1}}{1 + X_{k+1}}\right), \end{aligned} \quad (2.2)$$

where  $X_k$  takes values in  $\mathbf{X} = \mathbb{R}_+$ ,  $Y_k$  takes values in  $\mathbb{Z}_+$  and  $\theta = (\omega, a, b, r) \in (\mathbb{R}_+^*)^4$  is an unknown parameter. In (2.2),  $\mathcal{NB}(r, p)$  denotes the negative binomial distribution with parameters  $r > 0$  and  $p \in (0, 1)$ , whose probability function is  $\frac{\Gamma(k+r)}{k!\Gamma(r)} p^r (1-p)^k$  for all  $k \in \mathbb{Z}_+$ , where  $\Gamma$  stands for the Gamma function.

The second example, [Example 2.1.3](#), proposed by [Haas et al. \[2004\]](#) and [Alexander and Lazar \[2006\]](#), is a natural extension of GARCH processes, where the usual Gaussian conditional distribution of the observations given the hidden volatility variable is replaced by a mixture of Gaussian distributions given a hidden vector volatility variable. Up to our knowledge, the usual consistency proof of the MLE for the GARCH cannot be directly adapted to this model.

**Example 2.1.3.** The normal mixture GARCH (NM( $d$ )-GARCH(1, 1)) model is defined by:

$$\begin{aligned} \mathbf{X}_{k+1} &= \boldsymbol{\omega} + \mathbf{A}\mathbf{X}_k + Y_k^2 \mathbf{b}, \\ Y_{k+1}|\mathbf{X}_{0:k+1}, Y_{0:k} &\sim G^\theta(\mathbf{X}_{k+1}; \cdot), \\ G^\theta(\mathbf{x}; dy) &= \left( \sum_{\ell=1}^d \gamma_\ell \frac{e^{-y^2/2x_\ell}}{(2\pi x_\ell)^{1/2}} \right) dy, \quad \mathbf{x} = (x_i)_{1 \leq i \leq d} \in (\mathbb{R}_+^*)^d, \quad y \in \mathbb{R}, \end{aligned} \quad (2.3)$$

where  $d$  is a positive integer;  $\mathbf{X}_k = [X_{1,k} \dots X_{d,k}]^T$  takes values in  $\mathbf{X} = \mathbb{R}_+^d$ ;  $\boldsymbol{\gamma} = [\gamma_1 \dots \gamma_d]^T$  is a  $d$ -dimensional vector of mixture coefficients belonging to the  $d$ -dimensional simplex  $\mathbf{P}_d = \{\boldsymbol{\gamma} \in \mathbb{R}_+^d : \sum_{\ell=1}^d \gamma_\ell = 1\}$ ;  $\boldsymbol{\omega}$ ,  $\mathbf{b}$  are  $d$ -dimensional vector parameters with positive and nonnegative entries, respectively;  $\mathbf{A}$  is a  $d \times d$  matrix parameter with nonnegative entries; and  $\theta = (\boldsymbol{\gamma}, \boldsymbol{\omega}, \mathbf{A}, \mathbf{b})$ .

This chapter is organized as follows. [Section 2.2](#) is dedicated to the main result ([Theorem 2.2.7](#)) which shows that the argmax of the limiting criterion reduces to the equivalence class of the true parameter, as defined in [Leroux \[1992\]](#). The general setting is introduced in [Section 2.2.1](#). The theorem is stated and proved in [Section 2.2.2](#). In [Section 2.2.3](#), we focus on the kernel involved in the assumptions, and explain how it can be obtained explicitly. Our general assumptions are then shown to hold for two important classes of partially observed Markov models:

- First, the HMMs described in [Section 2.3](#), for which the equivalence-class consistency of the MLE is derived under simplified assumptions. The polynomially ergodic HMM of [Example 2.1.1](#) is treated as an application of this result.
- Second, the observation-driven time series models described in [Section 2.4](#). The obtained results apply to the models of [Example 2.1.2](#) and [Example 2.1.3](#), where the generating process of the observations may also depend on the parameter.

The technical proofs are gathered in [Section 2.5](#).

## 2.2 A General Approach to Identifiability

### 2.2.1 General Setting and Notation: Partially Dominated and Partially Observed Markov Chains

Let  $(\mathsf{X}, \mathcal{X})$  and  $(\mathsf{Y}, \mathcal{Y})$  be two Borel spaces, that is, measurable spaces that are isomorphic to a Borel subset of  $[0, 1]$  and let  $\Theta$  be a set of parameters. Consider a statistical model determined by a class of Markov kernels  $\{K^\theta : \theta \in \Theta\}$  on  $(\mathsf{X} \times \mathsf{Y}) \times (\mathcal{X} \otimes \mathcal{Y})$ . Throughout the chapter and in the following ones, unless it is otherwise specified, we will always denote by  $\mathbb{P}_\xi^\theta$  the probability (and by  $\mathbb{E}_\xi^\theta$  the corresponding expectation) induced on  $(\mathsf{X} \times \mathsf{Y})^{\mathbb{Z}_+}$  by a Markov chain  $\{(X_k, Y_k) : k \in \mathbb{Z}_+\}$  with transition kernel  $K^\theta$  and initial distribution  $\xi$  on  $\mathsf{X} \times \mathsf{Y}$ . In the case where  $\xi$  is a Dirac mass at  $(x, y)$ , we will simply write  $\mathbb{P}_{(x,y)}^\theta$ .

For partially observed Markov chains, that is, when only a sample  $Y_{1:n} := (Y_1, \dots, Y_n) \in \mathsf{Y}^n$  of the second component is observed, it is convenient to write  $K^\theta$  as

$$K^\theta((x, y); dx' dy') = Q^\theta((x, y); dx') G^\theta((x, y, x'); dy'), \quad (2.4)$$

where  $Q^\theta$  and  $G^\theta$  are probability kernels on  $(\mathsf{X} \times \mathsf{Y}) \times \mathcal{X}$  and on  $(\mathsf{X} \times \mathsf{Y} \times \mathsf{X}) \times \mathcal{Y}$ , respectively.

We now consider the following general setting.

**Definition 2.2.1.** We say that the Markov model  $\{K^\theta : \theta \in \Theta\}$  of the form (2.4) is partially dominated if there exists a  $\sigma$ -finite measure  $\nu$  on  $\mathsf{Y}$  such that for all  $(x, y), (x', y') \in \mathsf{X} \times \mathsf{Y}$ ,

$$G^\theta((x, y, x'); dy') = g^\theta((x, y, x'); y') \nu(dy'), \quad (2.5)$$

where the conditional density function  $g^\theta$  moreover satisfies

$$g^\theta((x, y, x'); y') > 0, \quad \text{for all } (x, y), (x', y') \in \mathsf{X} \times \mathsf{Y}. \quad (2.6)$$

It follows from (2.5) that, for all  $(x, y) \in \mathsf{X} \times \mathsf{Y}$ ,  $A \in \mathcal{X}$  and  $B \in \mathcal{Y}$ ,

$$K^\theta((x, y); A \times B) = \int_B \kappa^\theta\langle y, y' \rangle(x; A) \nu(dy'),$$

where, for all  $y, y' \in \mathsf{Y}$ ,  $\kappa^\theta\langle y, y' \rangle$  is a kernel defined on  $(\mathsf{X}, \mathcal{X})$  by

$$\kappa^\theta\langle y, y' \rangle(x; dx') := Q^\theta((x, y); dx') g^\theta((x, y, x'); y'). \quad (2.7)$$

The dynamic of the partially observed bivariate Markov process can be depicted as in Figure 2.1.

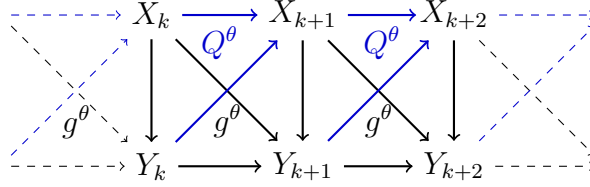


Figure 2.1 – Graphical representation of the partially observed bivariate Markov model.

**Remark 2.2.2.** Note that, in general, the kernel  $\kappa^\theta\langle y, y' \rangle$  is unnormalized since  $\kappa^\theta\langle y, y' \rangle(x; \mathsf{X})$  may be different from one. Moreover, we have for all  $(x, y, y') \in \mathsf{X} \times \mathsf{Y} \times \mathsf{Y}$ ,

$$\kappa^\theta\langle y, y' \rangle(x; \mathsf{X}) = \int_{\mathsf{X}} Q^\theta((x, y); dx') g^\theta((x, y, x'); y') > 0, \quad (2.8)$$

where the positiveness follows from the fact that  $Q^\theta((x, y); \cdot)$  is a probability on  $(\mathsf{X}, \mathcal{X})$  and Condition (2.6).

In well-specified models, it is assumed that the observations  $Y_{1:n}$  are generated from a process  $\{(X_k, Y_k) : k \in \mathbb{Z}_+\}$ , which follows the distribution  $\mathbb{P}_{\xi_\star}^{\theta_\star}$  associated to an unknown parameter  $\theta_\star \in \Theta$  and an unknown initial distribution  $\xi_\star$  (usually,  $\xi_\star$  is such that, under  $\mathbb{P}_{\xi_\star}^{\theta_\star}$ ,  $\{Y_k : k \in \mathbb{Z}_+\}$  is a stationary sequence). To form a consistent estimate of  $\theta_\star$  on the basis of the observations  $Y_{1:n}$  only, that is, without access to the hidden process  $\{X_k : k \in \mathbb{Z}_+\}$ , we define the maximum likelihood estimator (MLE)  $\hat{\theta}_{\xi, n}$  by

$$\hat{\theta}_{\xi, n} \in \operatorname{argmax}_{\theta \in \Theta} L_{\xi, n}(\theta),$$



where  $L_{\xi,n}(\theta)$  is the (conditional) log-likelihood function of the observations under parameter  $\theta$  with some arbitrary initial distribution  $\xi$  on  $\mathsf{X} \times \mathsf{Y}$ , that is,

$$\begin{aligned} L_{\xi,n}(\theta) &:= \ln \int \prod_{k=1}^n Q^\theta((x_{k-1}, y_{k-1}); dx_k) g^\theta((x_{k-1}, y_{k-1}, x_k); y_k) \xi(dx_0 dy_0) \\ &= \ln \int \kappa^\theta\langle y_0, y_1 \rangle \kappa^\theta\langle y_1, y_2 \rangle \cdots \kappa^\theta\langle y_{n-1}, y_n \rangle (x_0; \mathsf{X}) \xi(dx_0 dy_0). \end{aligned}$$

This corresponds to the log of the conditional density of  $Y_{1:n}$  given  $(X_0, Y_0)$  with the latter integrated according to  $\xi$ . In practice  $\xi$  is often taken as a Dirac mass at  $(x, y)$  with  $x$  arbitrarily chosen and  $y$  equal to the observation  $Y_0$  when it is available. In this context, a classical way (see, e.g., [Leroux \[1992\]](#)) to prove the consistency of a maximum-likelihood-type estimator  $\hat{\theta}_{\xi,n}$  may be decomposed in the following steps. The first step is to show that  $\hat{\theta}_{\xi,n}$  is, with probability tending to one, in a neighborhood of the set

$$\Theta_\star := \operatorname{argmax}_{\theta \in \Theta} \tilde{\mathbb{E}}^{\theta_\star} [\ln p^{\theta, \theta_\star}(Y_1 | Y_{-\infty:0})]. \quad (2.9)$$

This formula involves two quantities that have not yet been defined since they may require additional assumptions: first, the expectation  $\tilde{\mathbb{E}}^\theta$ , which corresponds to the distribution  $\tilde{\mathbb{P}}^\theta$  of a sequence  $\{Y_k : k \in \mathbb{Z}\}$  in accordance with the kernel  $K^\theta$ , and second, the density  $p^{\theta, \theta_\star}(\cdot | \cdot)$ , which shows up when taking the limit, under  $\tilde{\mathbb{P}}^{\theta_\star}$ , of the  $\tilde{\mathbb{P}}^\theta$ -conditional density of  $Y_1$  given its  $m$ -order past, as  $m$  goes to infinity. In many cases, such quantities appear naturally because the model is ergodic and the normalized log-likelihood  $n^{-1}L_{\xi,n}(\theta)$  can be approximated by

$$\frac{1}{n} \sum_{k=1}^n \ln p^{\theta, \theta_\star}(Y_k | Y_{-\infty:k-1}).$$

We will provide below some general assumptions, Assumptions [\(K-1\)](#) and [\(K-2\)](#), that yield precise definitions of  $\tilde{\mathbb{P}}^\theta$  and  $p^{\theta, \theta'}(\cdot | \cdot)$ .

The second step consists in proving that the set  $\Theta_\star$  in [\(2.9\)](#) is related to the true parameter  $\theta_\star$  in an exploitable way. Ideally, one could have  $\Theta_\star = \{\theta_\star\}$ , which would yield the consistency of  $\hat{\theta}_{\xi,n}$  for estimating  $\theta_\star$ . In this work, our first objective is to provide a set of general assumptions which ensures that  $\Theta_\star$  is exactly the set of parameters  $\theta$  such that  $\tilde{\mathbb{P}}^\theta = \tilde{\mathbb{P}}^{\theta_\star}$ . Then this result guarantees that the estimator converges to the set of parameters compatible with the true stationary distribution of the observations. If moreover the model  $\{\tilde{\mathbb{P}}^\theta : \theta \in \Theta\}$  is identifiable, then this set reduces to  $\{\theta_\star\}$  and consistency of  $\hat{\theta}_{\xi,n}$  directly follows.

To conclude with our general setting, we state the main assumption on the model and some subsequent notation and definitions used throughout this chapter.

(K-1) For all  $\theta \in \Theta$ , the transition kernel  $K^\theta$  admits a unique invariant probability  $\pi^\theta$ .

We now introduce some important notation used throughout this chapter.

**Definition 2.2.3.** Under Assumption (K-1), we denote by  $\pi_1^\theta$  and  $\pi_2^\theta$  the marginal distributions of  $\pi^\theta$  on  $\mathsf{X}$  and  $\mathsf{Y}$ , respectively, and by  $\mathbb{P}^\theta$  and  $\tilde{\mathbb{P}}^\theta$  the probability distributions defined respectively as follows:

- a)  $\mathbb{P}^\theta$  denotes the extension of  $\mathbb{P}_{\pi^\theta}^\theta$  on the whole line  $(\mathsf{X} \times \mathsf{Y})^{\mathbb{Z}}$ .
- b)  $\tilde{\mathbb{P}}^\theta$  is the corresponding projection on the component  $\mathsf{Y}^{\mathbb{Z}}$ .

We also use the symbols  $\mathbb{E}^\theta$  and  $\tilde{\mathbb{E}}^\theta$  to denote the expectations corresponding to  $\mathbb{P}^\theta$  and  $\tilde{\mathbb{P}}^\theta$ , respectively. Moreover, for all  $\theta, \theta' \in \Theta$ , we write  $\theta \sim \theta'$  if and only if  $\tilde{\mathbb{P}}^\theta = \tilde{\mathbb{P}}^{\theta'}$ . This defines an equivalence relation on the parameter set  $\Theta$  and the corresponding equivalence class of  $\theta$  is denoted by  $[\theta] := \{\theta' \in \Theta : \theta \sim \theta'\}$ .

The equivalence relationship  $\sim$  was introduced by [Leroux \[1992\]](#) as an alternative to the classical identifiability condition.

## 2.2.2 Main Result

Assumption (K-1) is supposed to hold all along this section and  $\mathbb{P}^\theta$ ,  $\tilde{\mathbb{P}}^\theta$  and  $\sim$  are given in [Definition 2.2.3](#). Our main result is stated under the following general assumption.

(K-2) For all  $\theta \neq \theta'$  in  $\Theta$ , there exists a probability kernel  $\Phi^{\theta, \theta'}$  on  $\mathsf{Y}^{\mathbb{Z}^-} \times \mathcal{X}$  such that for all  $A \in \mathcal{X}$ ,

$$\frac{\int_{\mathsf{X}} \Phi^{\theta, \theta'}(Y_{-\infty;0}; dx_0) \kappa^\theta \langle Y_0, Y_1 \rangle(x_0; A)}{\int_{\mathsf{X}} \Phi^{\theta, \theta'}(Y_{-\infty;0}; dx_0) \kappa^\theta \langle Y_0, Y_1 \rangle(x_0; \mathsf{X})} = \Phi^{\theta, \theta'}(Y_{-\infty;1}; A), \quad \tilde{\mathbb{P}}^{\theta'}\text{-a.s.}$$

**Remark 2.2.4.** Note that from [Remark 2.2.2](#), the denominator in the left-hand side of the last displayed equation is strictly positive, which ensures that the ratio is well-defined.

**Remark 2.2.5.** Let us give some insight about the formula appearing in (K-2) and explain why it is important to consider the cases  $\theta = \theta'$  and

$\theta \neq \theta'$  separately. Since  $\mathbf{X}$  is a Borel space, [Kallenberg, 2002, Theorem 6.3] applies and the conditional distribution of  $X_0$  given  $Y_{-\infty:0}$  under  $\mathbb{P}^\theta$  defines a probability kernel denoted by  $\Phi^\theta$ . We prove in Section 2.5.1 of Section 2.5 that this kernel satisfies, for all  $A \in \mathcal{X}$ ,

$$\frac{\int_{\mathbf{X}} \Phi^\theta(Y_{-\infty:0}; dx_0) \kappa^\theta \langle Y_0, Y_1 \rangle (x_0; A)}{\int_{\mathbf{X}} \Phi^\theta(Y_{-\infty:0}; dx_0) \kappa^\theta \langle Y_0, Y_1 \rangle (x_0; \mathbf{X})} = \Phi^\theta(Y_{-\infty:1}; A), \quad \tilde{\mathbb{P}}^\theta\text{-a.s.} \quad (2.10)$$

Assumption (K-2) asserts that the kernel  $\Phi^{\theta, \theta'}$  satisfies a similar identity  $\tilde{\mathbb{P}}^{\theta'}$ -a.s. for  $\theta' \neq \theta$ . It is not necessary at this stage to precise how  $\Phi^{\theta, \theta'}$  shows up. This is done in Section 2.2.3.

**Remark 2.2.6.** The denominator in the ratio displayed in (K-2) can be written as  $p^{\theta, \theta'}(Y_1 | Y_{-\infty:0})$ , where, for all  $y \in \mathbf{Y}$  and  $y_{-\infty:0} \in \mathbf{Y}^{\mathbb{Z}^-}$ ,

$$p^{\theta, \theta'}(y | y_{-\infty:0}) := \int_{\mathbf{X}} \Phi^{\theta, \theta'}(y_{-\infty:0}; dx_0) \kappa^\theta \langle y_0, y \rangle (x_0; \mathbf{X}) \quad (2.11)$$

is a conditional density with respect to the measure  $\nu$ , since for all  $(x, y) \in \mathbf{X} \times \mathbf{Y}$ ,  $\int \kappa^\theta \langle y, y' \rangle (x; \mathbf{X}) \nu(dy') = 1$ .

Since  $\mathbf{Y}$  is a Borel space, [Kallenberg, 2002, Theorem 6.3] applies and the conditional distribution of  $Y_{1:n}$  given  $Y_{-\infty:0}$  defines a probability kernel. Since  $\tilde{\mathbb{P}}^\theta(Y_{1:n} \in \cdot)$  is dominated by  $\nu^{\otimes n}$ , this in turns defines a conditional density with respect to  $\nu^{\otimes n}$ , which we denote by  $p_n^\theta(\cdot | \cdot)$ , so that for all  $B \in \mathcal{Y}^{\otimes n}$ ,

$$\tilde{\mathbb{P}}^\theta(Y_{1:n} \in B | Y_{-\infty:0}) = \int_B p_n^\theta(y_{1:n} | Y_{-\infty:0}) \nu(dy_1) \cdots \nu(dy_n), \quad \tilde{\mathbb{P}}^\theta\text{-a.s.} \quad (2.12)$$

Let us now state the main result.

**Theorem 2.2.7.** *Assume that (K-1) holds and define  $\mathbb{P}^\theta$ ,  $\tilde{\mathbb{P}}^\theta$  and  $[\theta]$  as in Definition 2.2.3. Suppose that Assumption (K-2) holds. For all  $\theta, \theta' \in \Theta$ , define  $p^{\theta, \theta'}(Y_1 | Y_{-\infty:0})$  by (2.11) if  $\theta \neq \theta'$  and by  $p^{\theta, \theta}(Y_1 | Y_{-\infty:0}) = p_1^\theta(Y_1 | Y_{-\infty:0})$  as in (2.12) otherwise. Then for all  $\theta_\star \in \Theta$ , we have*

$$\operatorname{argmax}_{\theta \in \Theta} \tilde{\mathbb{E}}^{\theta_\star} [\ln p^{\theta, \theta_\star}(Y_1 | Y_{-\infty:0})] = [\theta_\star]. \quad (2.13)$$

Before proving Theorem 2.2.7, we first extend the definition of the conditional density on  $\mathbf{Y}$  in (2.11) to a conditional density on  $\mathbf{Y}^n$ .

**Definition 2.2.8.** For every positive integer  $n$  and  $\theta \neq \theta' \in \Theta$ , define the function  $p_n^{\theta, \theta'}(\cdot | \cdot)$  on  $Y^n \times Y^{\mathbb{Z}^-}$  by

$$p_n^{\theta, \theta'}(y_{1:n} | y_{-\infty:0}) := \int_{\mathcal{X}^n} \Phi^{\theta, \theta'}(y_{-\infty:0}; dx_0) \prod_{k=0}^{n-1} \kappa^{\theta} \langle y_k, y_{k+1} \rangle(x_k; dx_{k+1}). \quad (2.14)$$

Again, it is easy to check that each  $p_n^{\theta, \theta'}(\cdot | y_{-\infty:0})$  is indeed a density on  $Y^n$ . Assumption (K-2) ensures that these density functions moreover satisfy the successive conditional formula, as for conditional densities, provided that we restrict ourselves to sequences in a set of  $\tilde{\mathbb{P}}^{\theta'}$ -probability one, as stated in the following lemma.

**Lemma 2.2.9.** Suppose that Assumption (K-2) holds and let  $p_n^{\theta, \theta'}(\cdot | \cdot)$  be as defined in Definition 2.2.8. Then for all  $\theta, \theta' \in \Theta$  and  $n \geq 2$ , we have

$$p_n^{\theta, \theta'}(Y_{1:n} | Y_{-\infty:0}) = p_1^{\theta, \theta'}(Y_n | Y_{-\infty:n-1}) p_{n-1}^{\theta, \theta'}(Y_{1:n-1} | Y_{-\infty:0}), \quad \tilde{\mathbb{P}}^{\theta'}\text{-a.s.} \quad (2.15)$$

The proof of this lemma is postponed to Section 2.5.2 in Section 2.5. We now have all the tools for proving the main result.

*Proof of Theorem 2.2.7.* Within this proof section, we will drop the subscript  $n$  and respectively write  $p^{\theta, \theta'}(y_{1:n} | y_{-\infty:0})$  and  $p^\theta(y_{1:n} | y_{-\infty:0})$  instead of  $p_n^{\theta, \theta'}(y_{1:n} | y_{-\infty:0})$  and  $p_n^\theta(y_{1:n} | y_{-\infty:0})$  when no ambiguity occurs.

For all  $\theta \in \Theta$ , we have by conditioning on  $Y_{-\infty:0}$  and by using (2.12),

$$\begin{aligned} & \tilde{\mathbb{E}}^{\theta_*} [\ln p^{\theta_*}(Y_1 | Y_{-\infty:0})] - \tilde{\mathbb{E}}^{\theta_*} [\ln p^{\theta, \theta_*}(Y_1 | Y_{-\infty:0})] \\ &= \tilde{\mathbb{E}}^{\theta_*} \left[ \tilde{\mathbb{E}}^{\theta_*} \left[ \ln \frac{p^{\theta_*}(Y_1 | Y_{-\infty:0})}{p^{\theta, \theta_*}(Y_1 | Y_{-\infty:0})} \middle| Y_{-\infty:0} \right] \right] \\ &= \tilde{\mathbb{E}}^{\theta_*} \left[ \text{KL} \left( p_1^{\theta_*}(\cdot | Y_{-\infty:0}) \middle\| p_1^{\theta, \theta_*}(\cdot | Y_{-\infty:0}) \right) \right], \end{aligned} \quad (2.16)$$

where  $\text{KL}(p||q)$  denotes the Kullback–Leibler divergence between the densities  $p$  and  $q$ . The nonnegativity of the Kullback–Leibler divergence shows that  $\theta_*$  belongs to the maximizing set on the left-hand side of (2.13). This implies

$$\operatorname{argmax}_{\theta \in \Theta} \tilde{\mathbb{E}}^{\theta_*} [\ln p^{\theta, \theta_*}(Y_1 | Y_{-\infty:0})] \supseteq [\theta_*], \quad (2.17)$$

where we have used the following lemma.

**Lemma 2.2.10.** Assume that (K-1) holds and define  $\tilde{\mathbb{E}}^\theta$  and  $[\theta]$  as in Definition 2.2.3. Suppose that for all  $\theta \in \Theta$ ,  $G(\theta)$  is a  $\sigma(Y_{-\infty:\infty})$ -measurable random variable such that, for all  $\theta_* \in \Theta$ ,

$$\sup_{\theta \in \Theta} \tilde{\mathbb{E}}^{\theta_*} [G(\theta)] = \tilde{\mathbb{E}}^{\theta_*} [G(\theta_*)].$$

Then for all  $\theta_\star \in \Theta$  and  $\theta' \in [\theta_\star]$ , we have

$$\tilde{\mathbb{E}}^{\theta_\star} [G(\theta')] = \sup_{\theta \in \Theta} \tilde{\mathbb{E}}^{\theta_\star} [G(\theta)].$$

*Proof.* Take  $\theta_\star \in \Theta$  and  $\theta' \in [\theta_\star]$ . Then we have, for all  $\theta \in \Theta$ ,  $\tilde{\mathbb{E}}^{\theta_\star} [G(\theta)] = \tilde{\mathbb{E}}^{\theta'} [G(\theta)]$ , and it follows that

$$\tilde{\mathbb{E}}^{\theta_\star} [G(\theta')] = \tilde{\mathbb{E}}^{\theta'} [G(\theta')] = \sup_{\theta \in \Theta} \tilde{\mathbb{E}}^{\theta'} [G(\theta)] = \sup_{\theta \in \Theta} \tilde{\mathbb{E}}^{\theta_\star} [G(\theta)],$$

which completes the proof.  $\square$

The proof of the reverse inclusion of (2.17) is more tricky. Let us take  $\theta \in \Theta_\star$  such that  $\theta \neq \theta_\star$  and show that it implies  $\theta \sim \theta_\star$ . By (2.16), we have

$$\tilde{\mathbb{E}}^{\theta_\star} \left[ \text{KL} \left( p_1^{\theta_\star}(\cdot | Y_{-\infty:0}) \middle| \middle| p_1^{\theta}(\cdot | Y_{-\infty:0}) \right) \right] = 0.$$

Consequently,

$$p^{\theta_\star}(Y_1 | Y_{-\infty:0}) = p^{\theta, \theta_\star}(Y_1 | Y_{-\infty:0}), \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

Applying Lemma 2.2.9 and using that  $\tilde{\mathbb{P}}^{\theta_\star}$  is shift-invariant, this relation propagates to all  $n \geq 2$ , so that

$$p^{\theta_\star}(Y_{1:n} | Y_{-\infty:0}) = p^{\theta, \theta_\star}(Y_{1:n} | Y_{-\infty:0}), \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.} \quad (2.18)$$

For any measurable function  $H : \mathcal{Y}^n \rightarrow \mathbb{R}_+$ , we get

$$\begin{aligned} \tilde{\mathbb{E}}^{\theta_\star} [H(Y_{1:n})] &= \tilde{\mathbb{E}}^{\theta_\star} \left\{ \tilde{\mathbb{E}}^{\theta_\star} \left[ H(Y_{1:n}) \frac{p^{\theta, \theta_\star}(Y_{1:n} | Y_{-\infty:0})}{p^{\theta_\star}(Y_{1:n} | Y_{-\infty:0})} \middle| Y_{-\infty:0} \right] \right\} \\ &= \tilde{\mathbb{E}}^{\theta_\star} \left[ \int H(y_{1:n}) p^{\theta, \theta_\star}(y_{1:n} | Y_{-\infty:0}) \nu^{\otimes n}(dy_{1:n}) \right], \end{aligned}$$

where the last equality follows from (2.12). Using Definition 2.2.8 and Tonelli's theorem, we obtain

$$\begin{aligned} \tilde{\mathbb{E}}^{\theta_\star} [H(Y_{1:n})] &= \tilde{\mathbb{E}}^{\theta_\star} \int H(y_{1:n}) \int \Phi^{\theta, \theta_\star}(Y_{-\infty:0}; dx_0) \kappa^\theta \langle Y_0, y_1 \rangle (x_0; dx_1) \times \\ &\quad \prod_{k=1}^{n-1} \kappa^\theta \langle y_k, y_{k+1} \rangle (x_k; dx_{k+1}) \nu^{\otimes n}(dy_{1:n}), \\ &= \tilde{\mathbb{E}}^{\theta_\star} \int \Phi^{\theta, \theta_\star}(Y_{-\infty:0}; dx_0) \int H(y_{1:n}) \kappa^\theta \langle Y_0, y_1 \rangle (x_0; dx_1) \times \\ &\quad \prod_{k=1}^{n-1} \kappa^\theta \langle y_k, y_{k+1} \rangle (x_k; dx_{k+1}) \nu^{\otimes n}(dy_{1:n}), \\ &= \tilde{\mathbb{E}}^{\theta_\star} \int \Phi^{\theta, \theta_\star}(Y_{-\infty:0}; dx_0) \mathbb{E}_{(x_0, Y_0)}^\theta [H(Y_{1:n})], \\ &= \mathbb{E}_{\pi^{\theta, \theta_\star}}^\theta [H(Y_{1:n})], \end{aligned}$$

where  $\pi^{\theta, \theta_\star}$  is a probability on  $\mathsf{X} \times \mathsf{Y}$  defined by

$$\pi^{\theta, \theta_\star}(A \times B) := \tilde{\mathbb{E}}^{\theta_\star} [\Phi^{\theta, \theta_\star}(Y_{-\infty:0}; A) \mathbb{1}_B(Y_0)],$$

for all  $(A, B) \in \mathcal{X} \times \mathcal{Y}$ . Consequently, for all  $B \in \mathcal{Y}^{\otimes \mathbb{Z}_+^*}$ ,

$$\tilde{\mathbb{P}}^{\theta_\star}(\mathsf{Y}^{\mathbb{Z}^-} \times B) = \mathbb{P}_{\pi^{\theta, \theta_\star}}^\theta(\mathsf{X}^{\mathbb{Z}^+} \times (\mathsf{Y} \times B)). \quad (2.19)$$

If we had  $\pi^\theta = \pi^{\theta, \theta_\star}$ , then we could conclude that the two shift-invariant distributions  $\tilde{\mathbb{P}}^{\theta_\star}$  and  $\tilde{\mathbb{P}}^\theta$  are the same and thus  $\theta \sim \theta_\star$ . Therefore, to complete the proof, it only remains to show that  $\pi^\theta = \pi^{\theta, \theta_\star}$ , which by (K-1) is equivalent to showing that  $\pi^{\theta, \theta_\star}$  is an invariant distribution for  $K^\theta$ .

Let us now prove this latter fact. Using that  $\tilde{\mathbb{P}}^{\theta_\star}$  is shift-invariant and then conditioning on  $Y_{-\infty:0}$ , we have, for any  $(A, B) \in \mathcal{X} \times \mathcal{Y}$ ,

$$\begin{aligned} \pi^{\theta, \theta_\star}(A \times B) &= \tilde{\mathbb{E}}^{\theta_\star} [\Phi^{\theta, \theta_\star}(Y_{-\infty:1}; A) \mathbb{1}_B(Y_1)], \\ &= \tilde{\mathbb{E}}^{\theta_\star} \int \Phi^{\theta, \theta_\star}(Y_{-\infty:0}, y_1; A) \mathbb{1}_B(y_1) p^{\theta_\star}(y_1 | Y_{-\infty:0}) \nu(dy_1), \\ &= \tilde{\mathbb{E}}^{\theta_\star} \int \Phi^{\theta, \theta_\star}(Y_{-\infty:0}, y_1; A) \mathbb{1}_B(y_1) p^{\theta, \theta_\star}(y_1 | Y_{-\infty:0}) \nu(dy_1), \end{aligned}$$

where in the last equality we have used (2.18). Using (K-2) we then get

$$\begin{aligned} \pi^{\theta, \theta_\star}(A \times B) &= \tilde{\mathbb{E}}^{\theta_\star} \int \Phi^{\theta, \theta_\star}(Y_{-\infty:0}; dx_0) \kappa^\theta \langle Y_0, y_1 \rangle (x_0; dx_1) \mathbb{1}_A(x_1) \mathbb{1}_B(y_1) \nu(dy_1), \\ &= \tilde{\mathbb{E}}^{\theta_\star} \int \Phi^{\theta, \theta_\star}(Y_{-\infty:0}; dx_0) K^\theta((x_0, Y_0); A \times B), \\ &= \pi^{\theta, \theta_\star} K^\theta(A \times B). \end{aligned}$$

Thus,  $\pi^{\theta, \theta_\star}$  is an invariant distribution for  $K^\theta$ , which completes the proof.  $\square$

### 2.2.3 Construction of the Kernel $\Phi^{\theta, \theta'}$ as a Backward Limit

Again, all along this section, Assumption (K-1) is supposed to hold and the symbols  $\mathbb{P}^\theta$  and  $\tilde{\mathbb{P}}^\theta$  refer to the probabilities introduced in Definition 2.2.3. In addition to Assumption (K-1), Theorem 2.2.7 fundamentally relies on Assumption (K-2). These assumptions ensure the existence of the probability kernel  $\Phi^{\theta, \theta'}$  that yields the definition of  $p_1^{\theta, \theta'}(\cdot | \cdot)$ . We now explain how the

kernel  $\Phi^{\theta, \theta'}$  may arise as a limit under  $\mathbb{P}^{\theta'}$  of explicit kernels derived from  $K^\theta$ . It will generally apply to observation-driven models, treated in [Section 2.4](#), but also in the more classical case of HMMs, as explained in [Section 2.3](#). A natural approach is to define the kernel  $\Phi^{\theta, \theta'}$  as the weak limit of the following ones.

**Definition 2.2.11.** Let  $n$  be a positive integer. For all  $\theta \in \Theta$  and  $x \in \mathsf{X}$ , we define the probability kernel  $\Phi_{x,n}^\theta$  on  $\mathsf{Y}^{n+1} \times \mathcal{X}$  by, for all  $y_{0:n} \in \mathsf{Y}^{n+1}$  and  $A \in \mathcal{X}$ ,

$$\Phi_{x,n}^\theta(y_{0:n}; A) := \frac{\int_{\mathsf{X}^{n-1} \times A} \prod_{k=0}^{n-1} \kappa^\theta \langle y_k, y_{k+1} \rangle (x_k; dx_{k+1})}{\int_{\mathsf{X}^n} \prod_{k=0}^{n-1} \kappa^\theta \langle y_k, y_{k+1} \rangle (x_k; dx_{k+1})} \quad \text{with } x_0 = x.$$

We will drop the subscript  $n$  when no ambiguity occurs.

It is worth noting that  $\Phi_{x,n}^\theta(Y_{0:n}; \cdot)$  is the conditional distribution of  $X_n$  given  $Y_{1:n}$  under  $\mathbb{P}_{(x, Y_0)}^\theta$ . To derive the desired  $\Phi^{\theta, \theta'}$  we take, for a well-chosen  $x$ , the limit of  $\Phi_{x,n}^\theta(y_{0:n}; \cdot)$  as  $n \rightarrow \infty$  for a sequence  $y_{0:n}$  corresponding to a path under  $\tilde{\mathbb{P}}^{\theta'}$ . The precise statement is provided in Assumption [\(K-3\)](#) below, which requires the following definition. For all  $\theta \in \Theta$  and for all nonnegative measurable functions  $f$  defined on  $\mathsf{X}$ , we set

$$\mathcal{F}_f^\theta := \{x \mapsto \kappa^\theta \langle y, y' \rangle (x; f) : (y, y') \in \mathsf{Y}^2\}.$$

We can now state the assumption as follows.

[\(K-3\)](#) For all  $\theta \neq \theta' \in \Theta$ , there exist  $x \in \mathsf{X}$ , a probability kernel  $\Phi^{\theta, \theta'}$  on  $\mathsf{Y}^{\mathbb{Z}^-} \times \mathcal{X}$  and a countable class  $\mathcal{F}$  of  $\mathsf{X} \rightarrow \mathbb{R}_+$  measurable functions such that for all  $f \in \mathcal{F}$ ,

$$\tilde{\mathbb{P}}^{\theta'} \left( \forall f' \in \mathcal{F}_f^\theta \cup \{f\}, \lim_{m \rightarrow \infty} \Phi_{x,m}^\theta(Y_{-m:0}; f') = \Phi^{\theta, \theta'}(Y_{-\infty:0}; f') < \infty \right) = 1.$$

The next lemma shows that, provided that  $\mathcal{F}$  is rich enough, Assumption [\(K-3\)](#) can be directly used to obtain Assumption [\(K-2\)](#). In what follows, we say that a class of  $\mathsf{X} \rightarrow \mathbb{R}$  functions is separating if, for any two probability measures  $\mu_1$  and  $\mu_2$  on  $(\mathsf{X}, \mathcal{X})$ , the equality of  $\mu_1(f)$  and  $\mu_2(f)$  over  $f$  in the class implies the equality of the two measures.

**Lemma 2.2.12.** *Suppose that Assumption [\(K-3\)](#) holds and that  $\mathcal{F}$  is a separating class of functions containing  $\mathbb{1}_{\mathsf{X}}$ . Then, the kernel  $\Phi^{\theta, \theta'}$  satisfies Assumption [\(K-2\)](#).*

*Proof.* Let  $x \in \mathsf{X}$  be given in Assumption (K-3). From Definition 2.2.11, we may write, for all  $f \in \mathcal{F}$ , setting  $x_{-m} = x$ ,

$$\Phi_{x,m}^\theta(Y_{-m:0}; f) = \frac{\int f(x_0) \prod_{k=-m}^{-1} \kappa^\theta \langle Y_k, Y_{k+1} \rangle (x_k; dx_{k+1})}{\int \prod_{k=-m}^{-1} \kappa^\theta \langle Y_k, Y_{k+1} \rangle (x_k; dx_{k+1})}$$

and, similarly,

$$\Phi_{x,m+1}^\theta(Y_{-m:1}; f) = \frac{\int f(x_1) \prod_{k=-m}^0 \kappa^\theta \langle Y_k, Y_{k+1} \rangle (x_k; dx_{k+1})}{\int \prod_{k=-m}^0 \kappa^\theta \langle Y_k, Y_{k+1} \rangle (x_k; dx_{k+1})}. \quad (2.20)$$

Dividing both numerator and denominator of (2.20) by

$$\int \prod_{k=-m}^{-1} \kappa^\theta \langle Y_k, Y_{k+1} \rangle (x_k; dx_{k+1}),$$

which is strictly positive by Remark 2.2.2, then (2.20) can be rewritten as

$$\Phi_{x,m+1}^\theta(Y_{-m:1}; f) = \frac{\Phi_{x,m}^\theta(Y_{-m:0}; \kappa^\theta \langle Y_0, Y_1 \rangle (\cdot; f))}{\Phi_{x,m}^\theta(Y_{-m:0}; \kappa^\theta \langle Y_0, Y_1 \rangle (\cdot; \mathbb{1}_{\mathsf{X}}))}. \quad (2.21)$$

Letting  $m \rightarrow \infty$  and applying Assumption (K-3), then  $\tilde{\mathbb{P}}^{\theta'}$ -a.s.,

$$\begin{aligned} \Phi^{\theta,\theta'}(Y_{-\infty:1}; f) &= \frac{\Phi^{\theta,\theta'}(Y_{-\infty:0}; \kappa^\theta \langle Y_0, Y_1 \rangle (\cdot; f))}{\Phi^{\theta,\theta'}(Y_{-\infty:0}; \kappa^\theta \langle Y_0, Y_1 \rangle (\cdot; \mathbb{1}_{\mathsf{X}}))}, \\ &= \frac{\int \Phi^{\theta,\theta'}(Y_{-\infty:0}; dx_0) \kappa^\theta \langle Y_0, Y_1 \rangle (x_0; f)}{\int \Phi^{\theta,\theta'}(Y_{-\infty:0}; dx_0) \kappa^\theta \langle Y_0, Y_1 \rangle (x_0; \mathbb{1}_{\mathsf{X}})}. \end{aligned}$$

Since  $\mathcal{F}$  is a separating class, the proof is complete. □



## 2.3 Application to Hidden Markov Models

### 2.3.1 Definitions and Assumptions

Hidden Markov models belong to a subclass of partially observed Markov models defined as follows.

**Definition 2.3.1.** Consider a partially observed and partially dominated Markov model given in [Definition 2.2.1](#) with Markov kernels  $\{K^\theta : \theta \in \Theta\}$ . We will say that this model is a hidden Markov model if the kernel  $K^\theta$  satisfies

$$K^\theta((x, y); dx'dy') = Q^\theta(x; dx')G^\theta(x'; dy'). \quad (2.22)$$

Moreover, in this context, we always assume that  $(\mathbf{X}, d_X)$  is a complete separable metric space and  $\mathcal{X}$  denotes the associated Borel  $\sigma$ -field.

In [\(2.22\)](#),  $Q^\theta$  and  $G^\theta$  are transition kernels on  $\mathbf{X} \times \mathcal{X}$  and  $\mathbf{X} \times \mathcal{Y}$ , respectively. Since the model is partially dominated, we denote by  $g^\theta$  the corresponding Radon–Nikodym derivative of  $G^\theta(x; \cdot)$  with respect to the dominating measure  $\nu$ : for all  $(x, y) \in \mathbf{X} \times \mathcal{Y}$ ,

$$\frac{dG^\theta(x; \cdot)}{d\nu}(y) = g^\theta(x; y).$$

One can directly observe that the unnormalized kernel  $\kappa^\theta\langle y, y' \rangle$  defined in [\(2.7\)](#) does no longer depend on  $y$ , and in this case, one can write

$$\kappa^\theta\langle y, y' \rangle(x; dx') = \kappa^\theta\langle y' \rangle(x; dx') = Q^\theta(x; dx')g^\theta(x'; y'). \quad (2.23)$$

The dynamic of hidden Markov process can be depicted as in [Figure 2.2](#).

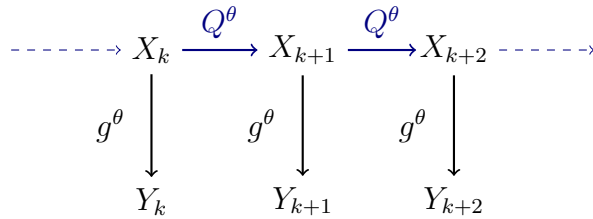


Figure 2.2 – Graphical representation of the hidden Markov Model by directed arrows.

For any integer  $n \geq 1$ ,  $\theta \in \Theta$  and sequence  $y_{0:n-1} \in \mathcal{Y}^n$ , consider the unnormalized kernel  $\mathbf{L}^\theta\langle y_{0:n-1} \rangle$  on  $\mathbf{X} \times \mathcal{X}$  defined by, for all  $x_0 \in \mathbf{X}$  and

$A \in \mathcal{X}$ ,

$$\mathbf{L}^\theta \langle y_{0:n-1} \rangle (x_0; A) = \int \cdots \int \left[ \prod_{k=0}^{n-1} g^\theta(x_k; y_k) Q^\theta(x_k; dx_{k+1}) \right] \mathbb{1}_A(x_n), \quad (2.24)$$

so that the MLE  $\hat{\theta}_{\xi,n}$ , associated to the observations  $Y_{0:n-1}$  with an arbitrary initial distribution  $\xi$  on  $\mathbf{X}$  is defined by

$$\hat{\theta}_{\xi,n} \in \operatorname{argmax}_{\theta \in \Theta} \xi \mathbf{L}^\theta \langle Y_{0:n-1} \rangle \mathbb{1}_{\mathbf{X}}.$$

We now follow the approach taken by [Douc and Moulines \[2012\]](#) in misspecified models and show that in the context of well-specified models, the maximizing set of the asymptotic normalized log-likelihood can be identified by relying neither on the exponential separation of measures, nor on the rates of convergence to the equilibrium, but only on the uniqueness of the invariant probability. We note the following fact which can be used to check [\(K-1\)](#).

**Remark 2.3.2.** In the HMM context,  $\pi^\theta$  is an invariant distribution of  $K^\theta$  if and only if  $\pi_1^\theta$  is an invariant distribution of  $Q^\theta$  and  $\pi^\theta(dx dy) = \pi_1^\theta(dx) G^\theta(x; dy)$ .

We illustrate the application of the main result ([Theorem 2.2.7](#)) in the context of HMMs by considering the assumptions of [Douc and Moulines \[2012\]](#) in the particular case of blocks of size 1 ( $r = 1$ ). Of course, general assumptions with arbitrary sizes of blocks could also be used but this complicates significantly the expressions and may confine the attention of the reader to unnecessary technicalities. To keep the discussion simple, we only consider blocks of size 1, which already covers many cases of interest.

Before listing the main assumptions, we recall the definition of a so-called *local Doeblin set* (in the particular case where  $r = 1$ ) as introduced in [[Douc and Moulines, 2012](#), Definition 1].

**Definition 2.3.3.** A set  $C$  is local Doeblin with respect to the family of kernels  $\{Q^\theta : \theta \in \Theta\}$  if there exist positive constants  $\epsilon_C^-$ ,  $\epsilon_C^+$  and a family of probability measures  $\{\lambda_C^\theta : \theta \in \Theta\}$  such that, for any  $\theta \in \Theta$ ,  $\lambda_C^\theta(C) = 1$ , and, for any  $A \in \mathcal{X}$  and  $x \in C$ ,

$$\epsilon_C^- \lambda_C^\theta(A) \leq Q^\theta(x; A \cap C) \leq \epsilon_C^+ \lambda_C^\theta(A).$$

Consider now the following set of assumptions.

(D-1) There exists a  $\sigma$ -finite measure  $\mu$  on  $(\mathbf{X}, \mathcal{X})$  that dominates  $Q^\theta(x; \cdot)$  for all  $(x, \theta) \in \mathbf{X} \times \Theta$ . Moreover, denoting  $q^\theta(x; x') := \frac{dQ^\theta(x; \cdot)}{d\mu}(x')$ , we have

$$q^\theta(x; x') > 0, \quad \text{for all } (x, x', \theta) \in \mathbf{X} \times \mathbf{X} \times \Theta.$$

(D-2) For all  $y \in \mathsf{Y}$ , we have  $\sup_{\theta \in \Theta} \sup_{x \in \mathsf{X}} g^\theta(x; y) < \infty$ .

(D-3) (a) For all  $\theta_* \in \Theta$ , there exists a set  $K \in \mathcal{Y}$  with  $\tilde{\mathbb{P}}^{\theta_*}(Y_0 \in K) > 2/3$  such that for all  $\eta > 0$ , there exists a local Doeblin set  $C \in \mathcal{X}$  with respect to  $\{Q^\theta : \theta \in \Theta\}$  satisfying, for all  $\theta \in \Theta$  and all  $y \in K$ ,

$$\sup_{x \in C^c} g^\theta(x; y) \leq \eta \sup_{x \in \mathsf{X}} g^\theta(x; y) < \infty. \quad (2.25)$$

(b) For all  $\theta_* \in \Theta$ , there exists a set  $D \in \mathcal{X}$  satisfying

$$\inf_{\theta \in \Theta} \inf_{x \in D} Q^\theta(x; D) > 0 \quad \text{and} \quad \tilde{\mathbb{E}}^{\theta_*} \left[ \ln^- \inf_{\theta \in \Theta} \inf_{x \in D} g^\theta(x; Y_0) \right] < \infty.$$

(D-4) For all  $\theta_* \in \Theta$ ,  $\tilde{\mathbb{E}}^{\theta_*} \left[ \ln^+ \sup_{\theta \in \Theta} \sup_{x \in \mathsf{X}} g^\theta(x; Y_0) \right] < \infty$ .

(D-5) There exists  $p \in \mathbb{Z}_+$  such that for any  $x \in \mathsf{X}$  and  $n \geq p$ , the function  $\theta \mapsto \mathbf{L}^\theta \langle Y_{0:n} \rangle(x; \mathsf{X})$  is  $\tilde{\mathbb{P}}^{\theta_*}$ -a.s. continuous on  $\Theta$ .

**Remark 2.3.4.** Under (D-1), for all  $\theta \in \Theta$ , the Markov kernel  $Q^\theta$  is  $\mu$ -irreducible, so that, using Remark 2.3.2, (K-1) reduces to the existence of a stationary distribution for  $Q^\theta$ .

**Remark 2.3.5.** Assumptions (D-3), (D-4) and (D-5) and (2.6) in Definition 2.2.1 correspond to (A1), (A2) and (A3) in Douc and Moulines [2012], where the blocks are of size  $r = 1$ .

**Remark 2.3.6.** Assumption (D-4) implies (D-2) up to a modification of  $g^\theta(x; y)$  on  $\nu$ -negligible set of  $y \in \mathsf{Y}$  for all  $x \in \mathsf{X}$ . Indeed, (D-4) implies that  $\sup_{\theta} \sup_x g^\theta(x; Y_0) < \infty$ ,  $\tilde{\mathbb{P}}^{\theta_*}$ -a.s., and it can be shown that under (D-1),  $\pi_2^{\theta_*} = \pi^{\theta_*}(\mathsf{X} \times \cdot)$  is equivalent to  $\nu$  for all  $\theta \in \Theta$ .

In these models, the kernel  $\Phi_{x,n}^\theta$  introduced in Definition 2.2.11 writes

$$\Phi_{x,n}^\theta(y_{0:n}; A) = \frac{\int_{\mathsf{X}^{n-1} \times A} \prod_{k=0}^{n-1} Q^\theta(x_k; dx_{k+1}) g^\theta(x_{k+1}; y_{k+1})}{\int_{\mathsf{X}^n} \prod_{k=0}^{n-1} Q^\theta(x_k; dx_{k+1}) g^\theta(x_{k+1}; y_{k+1})} \quad \text{with } x_0 = x.$$

The distribution  $\Phi_{x,n}^\theta(Y_{0:n}; \cdot)$  is usually referred to as the *filter distribution*. Proposition 2.3.7 (below) can be derived from [Douc and Moulines, 2012, Proposition 1]. For blocks of size 1, the initial distributions in Douc and

Moulines [2012] are constrained to belong to the set  $\mathcal{M}^{\theta_*}(D)$  of all probability distributions  $\xi$  defined on  $(\mathbf{X}, \mathcal{X})$  such that

$$\tilde{\mathbb{E}}^{\theta_*} \left[ \ln^- \inf_{\theta \in \Theta} \int \xi(\mathrm{d}x) g^\theta(x; Y_0) Q^\theta(x; D) \right] < \infty, \quad (2.26)$$

where  $D \in \mathcal{X}$  is the set appearing in (D-3). It turns out that under (D-3)-(b), all probability distributions  $\xi$  on  $(\mathbf{X}, \mathcal{X})$  satisfy (2.26), so the constraint on the initial distribution vanishes in our case.

**Proposition 2.3.7.** *Assume (D-3) and (D-4). Then the following assertions hold:*

- (i) *For any  $\theta, \theta_* \in \Theta$ , there exists a probability kernel  $\Phi^{\theta, \theta_*}$  on  $\mathbf{Y}^{\mathbb{Z}^-} \times \mathcal{X}$  such that for any  $x \in \mathbf{X}$ ,*

$$\tilde{\mathbb{P}}^{\theta_*} \left( \text{for all bounded } f, \lim_{m \rightarrow \infty} \Phi_{x,m}^\theta(Y_{-m:0}; f) = \Phi^{\theta, \theta_*}(Y_{-\infty:0}; f) \right) = 1.$$

- (ii) *For any  $\theta, \theta_* \in \Theta$  and probability measure  $\xi$  on  $(\mathbf{X}, \mathcal{X})$ ,*

$$\lim_{n \rightarrow \infty} n^{-1} \ln \xi \mathbf{L}^\theta \langle Y_{0:n-1} \rangle \mathbb{1}_{\mathbf{X}} = \ell(\theta, \theta_*), \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.},$$

where

$$\ell(\theta, \theta_*) := \tilde{\mathbb{E}}^{\theta_*} \left[ \ln \int \Phi^{\theta, \theta_*}(Y_{-\infty:0}; \mathrm{d}x_0) \kappa^\theta \langle Y_1 \rangle (x_0; \mathbf{X}) \right]. \quad (2.27)$$

### 2.3.2 Equivalence-class Consistency

We can now state the main result on the consistency of the MLE for HMMs.

**Theorem 2.3.8.** *Assume that (K-1) holds and define  $\mathbb{P}^\theta, \tilde{\mathbb{P}}^\theta$  and the equivalence class  $[\theta]$  as in Definition 2.2.3. Moreover, suppose that  $(\Theta, \Delta)$  is a compact metric space and that Assumptions (D-1)–(D-5) hold. Then, for any probability measure  $\xi$  on  $(\mathbf{X}, \mathcal{X})$ ,*

$$\lim_{n \rightarrow \infty} \Delta(\hat{\theta}_{\xi,n}, [\theta_*]) = 0, \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.}$$

*Proof.* According to [Douc and Moulines, 2012, Theorem 2],  $\theta \mapsto \ell(\theta, \theta_*)$  defined by (2.27) is upper semi-continuous (so that  $\Theta_* := \operatorname{argmax}_{\theta \in \Theta} \ell(\theta, \theta_*)$  is non-empty) and moreover

$$\lim_{n \rightarrow \infty} \Delta(\hat{\theta}_{\xi,n}, \Theta_*) = 0, \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.}$$

The proof then follows from [Theorem 2.2.7](#), provided that  $\ell(\theta, \theta_*)$  can be expressed as in the statement of [Theorem 2.2.7](#) and that [\(K-2\)](#) is satisfied. First note that, for  $\theta \neq \theta_*$ , the integral appearing within the logarithm in [\(2.27\)](#) corresponds to  $p^{\theta, \theta_*}(Y_1|Y_{-\infty:0})$  with  $p^{\theta, \theta_*}$  as defined in [\(2.11\)](#).

Let  $\mathcal{F}$  be a countable separating class of nonnegative bounded functions containing  $\mathbb{1}_X$ , see [[Parthasarathy, 2005](#), Theorem 6.6, Chapter 6] for the existence of such a class. By [Lemma 2.2.12](#), we check [\(K-2\)](#) by showing that [\(K-3\)](#) is satisfied. Condition [\(D-2\)](#) and [\(2.23\)](#) imply that for all bounded functions  $f$ ,  $\mathcal{F}_f^\theta$  is a class of bounded functions, and this in turn implies [\(K-3\)](#) by applying [Proposition 2.3.7-\(i\)](#) to all  $x$ . Thus, [\(K-2\)](#) is satisfied, and for  $\theta \neq \theta_*$ ,  $\ell(\theta, \theta_*)$  can be expressed as in the statement of [Theorem 2.2.7](#). To complete the proof, it only remains to consider the case where  $\theta = \theta_*$  and to show that  $\ell(\theta_*, \theta_*)$  can be written as

$$\ell(\theta_*, \theta_*) = \tilde{\mathbb{E}}^{\theta_*} [\ln p_1^{\theta_*}(Y_1|Y_{-\infty:0})], \quad (2.28)$$

where  $p_1^{\theta_*}(\cdot|\cdot)$  is the conditional density given in [\(2.12\)](#). According to [[Barron, 1985](#), Theorem 1], we have

$$\tilde{\mathbb{E}}^{\theta_*} [\ln p_1^{\theta_*}(Y_1|Y_{-\infty:0})] = \lim_{n \rightarrow \infty} n^{-1} \ln \pi_1^{\theta_*} \mathbf{L}^{\theta_*} \langle Y_{0:n-1} \rangle \mathbb{1}_X, \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.} \quad (2.29)$$

On the other hand, it follows from [Proposition 2.3.7-\(ii\)](#) that for any  $\theta \in \Theta$  and any initial probability distribution  $\xi$  on  $(X, \mathcal{X})$ ,

$$\lim_{n \rightarrow \infty} n^{-1} \ln \xi \mathbf{L}^\theta \langle Y_{0:n-1} \rangle \mathbb{1}_X = \ell(\theta, \theta_*), \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.},$$

where  $\ell(\theta, \theta_*)$  is deterministic and is defined by [\(2.27\)](#). In particular, for  $\xi = \pi_1^{\theta_*}$  and for  $\theta = \theta_*$ , it holds that

$$\lim_{n \rightarrow \infty} n^{-1} \ln \pi_1^{\theta_*} \mathbf{L}^{\theta_*} \langle Y_{0:n-1} \rangle \mathbb{1}_X = \ell(\theta_*, \theta_*), \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.} \quad (2.30)$$

From [\(2.29\)](#) and [\(2.30\)](#) and by the uniqueness of the limit, [\(2.28\)](#) follows and the proof is complete.  $\square$

### 2.3.3 A Polynomially Ergodic Example

As an application of [Theorem 2.3.8](#), we consider the HMM model described in [Example 2.1.1](#). In addition to the assumptions introduced in [Example 2.1.1](#), we assume that  $U_0$  and  $V_0$  are independent and centered and they both admit densities with respect to the Lebesgue measure  $\lambda$  over  $\mathbb{R}$ , denoted by  $r$  and  $h$ , respectively, and

(E-1) the density  $r$  satisfies:

- (a)  $r$  is continuous and positive over  $\mathbb{R}$ ,
- (b) there exists  $\alpha > 2$  such that  $r(u)|u|^{\alpha+1}$  is bounded away from  $\infty$  as  $|u| \rightarrow \infty$  and from 0 as  $u \rightarrow \infty$ ,

(E-2) the density  $h$  satisfies:

- (a)  $h$  is continuous and positive over  $\mathbb{R}$ , and  $\lim_{|v| \rightarrow \infty} h(v) = 0$ ,
- (b) there exist  $\beta \in [1, \alpha - 1)$  (where  $\alpha$  is given in (E-1)) and  $b, c > 0$  such that  $\mathbb{E}(|V_0|^\beta) < \infty$  and  $h(v) \geq be^{-c|v|^\beta}$  for all  $v \in \mathbb{R}$ .

For example, a symmetric Pareto distribution with a parameter strictly larger than 2 satisfies (E-1) and provided that  $\alpha > 3$ , (E-2) holds with a centered Gaussian distribution. The model is parameterized by  $\theta = (m, a) \in \Theta := [\underline{m}, \bar{m}] \times [\underline{a}, \bar{a}]$  where  $0 < \underline{m} < \bar{m}$  and  $\underline{a} < \bar{a}$ . In this model, the Markov transition  $Q^\theta$  of  $\{X_k : k \in \mathbb{Z}_+\}$  has a transition density  $q^\theta$  with respect to the dominating measure  $\mu(dx) = \lambda(dx) + \delta_0(dx)$ , which can be written as follows: for all  $(x, x') \in \mathbb{R}_+^2$ ,

$$q^\theta(x; x') = r(x' - x + m)\mathbb{1}\{x' > 0\} + \left( \int_{-\infty}^{m-x} r(u)du \right) \mathbb{1}\{x' = 0\}. \quad (2.31)$$

Moreover, (2.1) implies

$$g^\theta(x; y) = h(y - ax). \quad (2.32)$$

Following [Jarner and Roberts \[2002\]](#), we have the following lemma.

**Lemma 2.3.9.** *Assume (E-1) and (E-2). For all  $\theta \in \Theta$ , the Markov kernel  $Q^\theta$  is not geometrically ergodic. Moreover,  $Q^\theta$  is polynomially ergodic and its (unique) stationary distribution  $\pi_1^\theta$ , defined on  $\mathbf{X} = \mathbb{R}_+$ , satisfies  $\int \pi_1^\theta(dx)x^\beta < \infty$ , for all  $\beta \in [1, \alpha - 1)$ .*

*Proof.* The proof of this lemma is postponed to [Section 2.5.3](#) in [Section 2.5](#).  $\square$

**Proposition 2.3.10.** *Consider the HMM of [Example 2.1.1](#) under Assumptions (E-1) and (E-2). Then (K-1) holds and we define  $\mathbb{P}^\theta$ ,  $\tilde{\mathbb{P}}^\theta$  and the equivalence class  $[\theta]$  as in [Definition 2.2.3](#). Moreover, for any probability measure  $\xi$ , the MLE  $\hat{\theta}_{\xi, n}$  is equivalence-class consistent, that is, for any  $\theta_\star \in \Theta$ ,*

$$\lim_{n \rightarrow \infty} \Delta(\hat{\theta}_{\xi, n}, [\theta_\star]) = 0, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

*Proof.* To apply [Theorem 2.3.8](#), we need to check (K-1) and (D-1)–(D-5). First observe that Assumption (K-1) immediately follows from [Remark 2.3.2](#)

and [Lemma 2.3.9](#), and Assumptions [\(D-1\)](#) and [\(D-2\)](#) directly follow from the positiveness of the density  $r$  and the boundedness of the density  $h$ , respectively. Now, using [\(E-1\)-\(a\)](#), it can be easily shown that all compact sets are local Doeblin sets and this in turn implies, via  $\lim_{|x| \rightarrow \infty} h(x) = 0$ , that Assumption [\(D-3\)-\(a\)](#) is satisfied. We now check [\(D-3\)-\(b\)](#). By [\(E-1\)-\(a\)](#), we have for all compact sets  $D$ ,  $\inf \{r(x' - x + m) : (x, x', m) \in D^2 \times [\underline{m}, \bar{m}]\} > 0$ , which by [\(2.31\)](#) implies

$$\inf_{\theta \in \Theta} \inf_{x \in D} Q^\theta(x; D) > 0.$$

To obtain [\(D-3\)-\(b\)](#), it thus remains to show

$$\tilde{\mathbb{E}}^{\theta_\star} \left[ \ln^- \inf_{\theta \in \Theta} \inf_{x \in D} g^\theta(x; Y_0) \right] < \infty.$$

By [\(E-2\)-\(b\)](#), there exist positive constants  $b$  and  $c$  such that  $h(v) \geq be^{-c|v|^\beta}$ . Plugging this into [\(2.32\)](#) yields

$$\begin{aligned} \tilde{\mathbb{E}}^{\theta_\star} \left[ \ln^- \inf_{\theta \in \Theta} \inf_{x \in D} g^\theta(x; Y_0) \right] &\leq \tilde{\mathbb{E}}^{\theta_\star} \left[ |\ln b| + c \left( |Y_0| + \bar{a} \sup_{x \in D} |x| \right)^\beta \right] \\ &= \mathbb{E}^{\theta_\star} \left[ |\ln b| + c \left( |aX_0 + V_0| + \bar{a} \sup_{x \in D} |x| \right)^\beta \right] < \infty, \end{aligned}$$

where the finiteness follows from [\(E-2\)-\(b\)](#) and [Lemma 2.3.9](#). Finally, [\(D-3\)](#) is satisfied. [\(D-4\)](#) is checked by writing

$$\tilde{\mathbb{E}}^{\theta_\star} \left[ \ln^+ \sup_{\theta \in \Theta} \sup_{x \in \mathbf{X}} g^\theta(x; Y_0) \right] \leq \ln^+ \sup_{x \in \mathbb{R}} h(x) < \infty.$$

To obtain [\(D-5\)](#), we show by induction on  $n$  that for all  $n \geq 1$ ,  $y_{0:n-1} \in \mathbb{R}^n$  and  $x_0 \in \mathbb{R}_+$ , the function  $\theta \mapsto \mathbf{L}^\theta \langle y_{0:n-1} \rangle (x_0; \mathbf{X})$  is continuous on  $\Theta$ . The case where  $n = 1$  is obvious since  $\mathbf{L}^\theta \langle y_0 \rangle (x_0; \mathbf{X}) = g^\theta(x_0; y_0) = h(y_0 - ax_0)$ . We next assume the induction hypothesis on  $n$  and note that

$$\mathbf{L}^\theta \langle y_{0:n} \rangle (x_0; \mathbf{X}) = g^\theta(x_0; y_0) \int \mu(dx_1) q^\theta(x_0; x_1) \mathbf{L}^\theta \langle y_{1:n} \rangle (x_1; \mathbf{X}).$$

The continuity of  $\theta \mapsto g^\theta(x_0; y_0)$  follows from [\(2.32\)](#) and the continuity of  $h$ . Similarly, the continuity of  $\theta \mapsto q^\theta(x_0; x_1)$  follows from [\(2.31\)](#) and the continuity of  $r$ . Moreover,  $\theta \mapsto \mathbf{L}^\theta \langle y_{1:n} \rangle (x_1; \mathbf{X})$  is continuous by the induction

assumption. The continuity of  $\theta \mapsto \int \mu(dx_1)q^\theta(x_0; x_1)\mathbf{L}^\theta\langle y_{1:n}\rangle(x_1; \mathbf{X})$  then follows from the Lebesgue convergence theorem provided that

$$\int \mu(dx_1) \sup_{\theta \in \Theta} q^\theta(x_0; x_1)\mathbf{L}^\theta\langle y_{1:n}\rangle(x_1; \mathbf{X}) < \infty \quad (2.33)$$

holds. Note further that by the expression of  $q^\theta(x_0; x_1)$  given in (2.31) and the tail assumption (E-1)-(b), we obtain for all  $x_0 \in \mathbf{X}$ ,

$$\int \mu(dx_1) \sup_{\theta \in \Theta} q^\theta(x_0; x_1) < \infty.$$

Combining with that  $\mathbf{L}^\theta\langle y_{1:n}\rangle(x_1; \mathbf{X}) \leq (\sup_{x \in \mathbb{R}} h(x))^n$  yields (2.33). Finally, we have (D-5), and thus Theorem 2.3.8 holds under (E-1) and (E-2).  $\square$

## 2.4 Application to Observation-driven Models

Observation-driven models are a subclass of partially dominated and partially observed Markov models.

We split our study of the observation-driven model into several parts. Specific definitions and notation are introduced in Section 2.4.1. Then we provide sufficient conditions that allow to apply our general result Theorem 2.2.7, that is,  $\Theta_\star = [\theta_\star]$ . This is done in Section 2.4.2.

### 2.4.1 Definitions and Notation

Observation-driven models are formally defined as follows.

**Definition 2.4.1.** Consider a partially observed and partially dominated Markov model given in Definition 2.2.1 with Markov kernels  $\{K^\theta : \theta \in \Theta\}$ . We say that this model is an observation-driven model if the kernel  $K^\theta$  satisfies

$$K^\theta((x, y); dx'dy') = \delta_{\psi_y^\theta(x)}(dx')G^\theta(x'; dy'), \quad (2.34)$$

where  $\delta_a$  denotes the Dirac mass at point  $a$ ,  $G^\theta$  is a probability kernel on  $\mathbf{X} \times \mathcal{Y}$  and  $\{(x, y) \mapsto \psi_y^\theta(x) : \theta \in \Theta\}$  is a family of measurable functions from  $(\mathbf{X} \times \mathcal{Y}, \mathcal{X} \otimes \mathcal{Y})$  to  $(\mathbf{X}, \mathcal{X})$ . Moreover, in this context, we always assume that  $(\mathbf{X}, d_X)$  is a complete separable metric space and  $\mathcal{X}$  denotes the associated Borel  $\sigma$ -field.

Note that a Markov chain  $\{(X_k, Y_k) : k \in \mathbb{Z}_+\}$  with probability kernel given by (2.34) can be equivalently defined by the following recursions:

$$\begin{aligned} X_{k+1} &= \psi_{Y_k}^\theta(X_k), \\ Y_{k+1}|X_{0:k+1}, Y_{0:k} &\sim G^\theta(X_{k+1}; \cdot) \end{aligned} \quad (2.35)$$



and can be graphically represented as in [Figure 2.3](#).

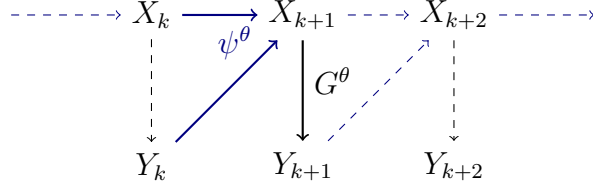


Figure 2.3 – Graphical representation of the observation-driven model.

The most celebrated example is the GARCH(1, 1) process, where  $G^\theta(x; \cdot)$  is a centered (say Gaussian) distribution with variance  $x$  and  $\psi_y^\theta(x)$  is an affine function of  $x$  and  $y^2$ .

As a special case of [Definition 2.2.1](#), for all  $x \in \mathbf{X}$ ,  $G^\theta(x; \cdot)$  is dominated by some  $\sigma$ -finite measure  $\nu$  on  $(\mathbf{Y}, \mathcal{Y})$  and we denote by  $g^\theta(x; \cdot)$  its Radon–Nikodym derivative,  $g^\theta(x; y) = \frac{dG^\theta(x; \cdot)}{d\nu}(y)$ . A dominated parametric observation-driven model is thus defined by the collection  $\{(g^\theta, \psi^\theta) : \theta \in \Theta\}$ . Moreover, [\(2.6\)](#) may be rewritten in this case: for all  $(x, y) \in \mathbf{X} \times \mathbf{Y}$  and for all  $\theta \in \Theta$ ,

$$g^\theta(x; y) > 0.$$

Under [\(K-1\)](#), we assume that the model is well-specified, that is, the observation sample  $(Y_1, \dots, Y_n)$  is distributed according to  $\tilde{\mathbb{P}}^{\theta_\star}$  for some unknown parameter  $\theta_\star$ . The inference of  $\theta_\star$  is based on the conditional likelihood of  $(Y_1, \dots, Y_n)$  given  $X_1 = x$  for an arbitrary  $x \in \mathbf{X}$ . The corresponding density function with respect to  $\nu^{\otimes n}$  is, under parameter  $\theta$ ,

$$y_{1:n} \mapsto \prod_{k=1}^n g^\theta(\psi^\theta\langle y_{1:k-1} \rangle(x); y_k), \quad (2.36)$$

where, for any vector  $y_{1:p} = (y_1, \dots, y_p) \in \mathbf{Y}^p$ ,  $\psi^\theta\langle y_{1:p} \rangle$  is the  $\mathbf{X} \rightarrow \mathbf{X}$  function defined as the successive composition of  $\psi_{y_1}^\theta$ ,  $\psi_{y_2}^\theta$ , ..., and  $\psi_{y_p}^\theta$ ,

$$\psi^\theta\langle y_{1:p} \rangle = \psi_{y_p}^\theta \circ \psi_{y_{p-1}}^\theta \circ \dots \circ \psi_{y_1}^\theta, \quad (2.37)$$

with the convention  $\psi^\theta\langle y_{s:t} \rangle(x) = x$  for  $s > t$ . Then the corresponding (conditional) MLE  $\hat{\theta}_{x,n}$  of the parameter  $\theta$  is defined by

$$\hat{\theta}_{x,n} \in \operatorname{argmax}_{\theta \in \Theta} \mathbf{L}_{x,n}^\theta\langle Y_{1:n} \rangle, \quad (2.38)$$

where

$$\mathbb{L}_{x,n}^\theta \langle y_{1:n} \rangle := n^{-1} \ln \left( \prod_{k=1}^n g^\theta (\psi^\theta \langle y_{1:k-1} \rangle (x); y_k) \right). \quad (2.39)$$

We will provide simple conditions for the consistency of  $\hat{\theta}_{x,n}$  in the sense that, with probability tending to one, for a well-chosen  $x$ ,  $\hat{\theta}_{x,n}$  belongs to a neighborhood of the equivalence class  $[\theta_\star]$  of  $\theta_\star$ , as given by [Definition 2.2.3](#).

### 2.4.2 Identifiability

Let us consider the following assumptions.

(C-1) For all  $\theta \neq \theta_\star \in \Theta$ , there exist  $x \in \mathsf{X}$  and a measurable function  $\psi^{\theta, \theta_\star} \langle \cdot \rangle$  defined on  $\mathsf{Y}^{\mathbb{Z}^-}$  such that

$$\lim_{m \rightarrow \infty} \psi^\theta \langle Y_{-m:0} \rangle (x) = \psi^{\theta, \theta_\star} \langle Y_{-\infty:0} \rangle, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.} \quad (2.40)$$

(C-2) For all  $\theta \in \Theta$  and  $y \in \mathsf{Y}$ , the function  $x \mapsto g^\theta(x; y)$  is continuous on  $\mathsf{X}$ .

(C-3) For all  $\theta \in \Theta$  and  $y \in \mathsf{Y}$ , the function  $x \mapsto \psi_y^\theta(x)$  is continuous on  $\mathsf{X}$ .

In observation-driven models, the kernel  $\kappa^\theta$  defined in (2.7) reads

$$\begin{aligned} \kappa^\theta \langle y, y' \rangle (x; dx') &= g^\theta(x'; y') \delta_{\psi_y^\theta(x)}(dx') \\ &= g^\theta(\psi_y^\theta(x); y') \delta_{\psi_y^\theta(x)}(dx'), \end{aligned} \quad (2.41)$$

and the probability kernel  $\Phi_{x,n}^\theta$  in [Definition 2.2.11](#) reads, for all  $x \in \mathsf{X}$  and  $y_{0:n} \in \mathsf{Y}^{n+1}$ ,

$$\Phi_{x,n}^\theta(y_{0:n}; \cdot) = \delta_{\psi^{\theta, \theta_\star} \langle y_{0:n-1} \rangle (x)} \quad (2.42)$$

(the Dirac point mass at  $\psi^{\theta, \theta_\star} \langle y_{0:n-1} \rangle (x)$ ). Using these expressions, we get the following result which is a special case of [Theorem 2.2.7](#).

**Theorem 2.4.2.** *Assume that (K-1) holds in the observation-driven model setting and define  $\mathbb{P}^\theta$ ,  $\tilde{\mathbb{P}}^\theta$  and  $[\theta]$  as in [Definition 2.2.3](#). Suppose that Assumptions (C-1), (C-2) and (C-3) hold and define  $p^{\theta, \theta_\star}(\cdot | \cdot)$  by setting, for  $\tilde{\mathbb{P}}^{\theta_\star}$ -a.e.  $y_{-\infty:0} \in \mathsf{Y}^{\mathbb{Z}^-}$ ,*

$$p^{\theta, \theta_\star}(y_1 | y_{-\infty:0}) = \begin{cases} g^\theta(\psi^{\theta, \theta_\star} \langle y_{-\infty:0} \rangle; y_1) & \text{if } \theta \neq \theta_\star, \\ p_1^\theta(y_1 | y_{-\infty:0}) \text{ as defined by (2.12)} & \text{otherwise.} \end{cases} \quad (2.43)$$

Then, for all  $\theta_\star \in \Theta$ , we have

$$\operatorname{argmax}_{\theta \in \Theta} \tilde{\mathbb{E}}^{\theta_\star} [\ln p^{\theta, \theta_\star}(Y_1 | Y_{-\infty:0})] = [\theta_\star]. \quad (2.44)$$

*Proof.* We apply [Theorem 2.2.7](#). It is thus sufficient to show that [\(C-1\)](#), [\(C-2\)](#) and [\(C-3\)](#) implies [\(K-2\)](#) with

$$\Phi^{\theta, \theta_*}(y_{-\infty:0}; \cdot) = \delta_{\psi^{\theta, \theta_*} \langle y_{-\infty:-1} \rangle}, \quad \text{for all } y_{-\infty:0} \in \mathbf{Y}^{\mathbb{Z}^-}, \quad (2.45)$$

and that for  $\theta \neq \theta_*$ , the conditional density  $p^{\theta, \theta_*}$  defined by [\(2.11\)](#) satisfies

$$p^{\theta, \theta_*}(y|Y_{-\infty:0}) = g^\theta(\psi^{\theta, \theta_*} \langle Y_{-\infty:0} \rangle; y), \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.} \quad (2.46)$$

By [Lemma 2.2.12](#), it is sufficient to prove that Assumption [\(K-3\)](#) holds for the kernel  $\Phi^{\theta, \theta_*}$  defined above. Denote by  $\mathcal{C}(\mathbf{X})$  the set of continuous functions on  $\mathbf{X}$ , and by  $\mathcal{C}_b(\mathbf{X})$  the set of bounded functions in  $\mathcal{C}(\mathbf{X})$ . By [[Parthasarathy, 2005](#), Theorem 6.6, Chapter 6], there is a countable and separating subclass  $\mathcal{F}$  of nonnegative functions in  $\mathcal{C}_b(\mathbf{X})$  such that  $\mathbb{1}_{\mathbf{X}} \in \mathcal{F}$ . Now, let us take  $\theta, \theta_* \in \Theta$  and  $f \in \mathcal{F}$ . Then, by [\(C-2\)](#), [\(C-3\)](#) and [\(2.41\)](#), we have

$$\mathcal{F}_f^\theta = \{x \mapsto \kappa^\theta \langle y, y' \rangle(x; f) : (y, y') \in \mathbf{Y}^2\} \subset \mathcal{C}(\mathbf{X}).$$

By [\(2.42\)](#), [\(C-1\)](#) and [\(2.45\)](#), we obtain [\(K-3\)](#) with  $x$  chosen as in [\(C-1\)](#).

To complete, we need to show [\(2.46\)](#). Note that [\(2.45\)](#) together with [\(2.41\)](#) and the usual definition [\(2.11\)](#) of  $p^{\theta, \theta_*}$  yields

$$p^{\theta, \theta_*}(y|y_{-\infty:0}) = g^\theta(\psi_{y_0}^\theta(\psi^{\theta, \theta_*} \langle y_{-\infty:-1} \rangle); y).$$

By Assumption [\(C-3\)](#) and the definition of  $\psi^{\theta, \theta_*} \langle \cdot \rangle$  in [\(C-1\)](#), we get [\(2.46\)](#).  $\square$

### 2.4.3 Examples

In the context of observation-driven time series, easy-to-check conditions are derived in [Douc et al. \[2015\]](#) in order to establish the convergence of the MLE  $\hat{\theta}_{x,n}$  defined by [\(2.38\)](#) to the maximizing set of the asymptotic normalized log-likelihood. It turns out that the conditions of [[Douc et al., 2015](#), Theorem 1] also imply the conditions of [Theorem 2.4.2](#). More precisely, the assumptions (B-2) and (B-3) of [[Douc et al., 2015](#), Theorem 1] are stronger than [\(C-2\)](#) and [\(C-3\)](#) used in [Theorem 2.4.2](#) above, and it is shown that the assumptions of [[Douc et al., 2015](#), Theorem 1] imply [\(C-1\)](#) (see the proof of Lemma 2 in Section 6.3 of [Douc et al. \[2015\]](#)). Moreover, the conditions of Theorem 1 are shown to be satisfied in the context of Examples [2.1.2](#) and [2.1.3](#) (see [[Douc et al., 2015](#), Theorem 3 and Theorem 4]), provided that  $\Theta$  in [\(2.38\)](#) is a compact metric space such that:

1. in the case of [Example 2.1.2](#), all  $\theta = (\omega, a, b, r) \in \Theta$  satisfy  $rb + a < 1$ ;
2. in the case of [Example 2.1.3](#), all  $\theta = (\gamma, \omega, \mathbf{A}, \mathbf{b}) \in \Theta$  are such that the spectral radius of  $\mathbf{A} + \mathbf{b}\gamma^T$  is strictly less than 1.

Under these assumptions, we conclude that the MLE is equivalence-class consistent for both examples, which up to our best knowledge had not been proven so far.

## 2.5 Postponed Proofs

### 2.5.1 Proof of Eq. (2.10)

Let  $\theta \in \Theta$ . Recall that in [Remark 2.2.5](#),  $\Phi^\theta$  is defined as the probability kernel of the conditional distribution of  $X_0$  given  $Y_{-\infty:0}$  under  $\mathbb{P}^\theta$ , that is, for all  $A \in \mathcal{X}$ ,

$$\Phi^\theta(Y_{-\infty:0}; A) = \mathbb{P}^\theta(X_0 \in A \mid Y_{-\infty:0}), \quad \tilde{\mathbb{P}}^\theta\text{-a.s.}$$

Conditioning on  $X_0, Y_0$  and using the definition of  $\kappa^\theta$  in [\(2.7\)](#), we get that, for all  $A \in \mathcal{X}, B \in \mathcal{Y}$  and  $C \in \mathcal{Y}^{\otimes \mathbb{Z}^-}$ ,

$$\begin{aligned} & \mathbb{P}^\theta(X_1 \in A, Y_1 \in B, Y_{-\infty:0} \in C) \\ &= \mathbb{E}^\theta \left[ \int_B \kappa^\theta \langle Y_0, y_1 \rangle (X_0; A) \mathbb{1}_C(Y_{-\infty:0}) \nu(dy_1) \right] \\ &= \tilde{\mathbb{E}}^\theta \left[ \int_{\mathbf{X} \times B} \Phi^\theta(Y_{-\infty:0}; dx_0) \kappa^\theta \langle Y_0, y_1 \rangle (x_0; A) \mathbb{1}_C(Y_{-\infty:0}) \nu(dy_1) \right]. \end{aligned} \quad (2.47)$$

Let us denote

$$\hat{\Phi}^\theta(Y_{-\infty:0}, y_1; A) = \frac{\int \Phi^\theta(Y_{-\infty:0}; dx_0) \kappa^\theta \langle Y_0, y_1 \rangle (x_0; A)}{\int \Phi^\theta(Y_{-\infty:0}; dx_0) \kappa^\theta \langle Y_0, y_1 \rangle (x_0; \mathbf{X})},$$

which is always defined since the denominator does not vanish by [Remark 2.2.2](#). With this notation, we deduce from [\(2.47\)](#) that

$$\begin{aligned} & \mathbb{P}^\theta(X_1 \in A, Y_1 \in B, Y_{-\infty:0} \in C) \\ &= \tilde{\mathbb{E}}^\theta \left[ \int_B \hat{\Phi}^\theta(Y_{-\infty:0}, y_1; A) \left( \int \Phi^\theta(Y_{-\infty:0}; dx_0) \kappa^\theta \langle Y_0, y_1 \rangle (x_0; \mathbf{X}) \right) \right. \\ & \quad \left. \mathbb{1}_C(Y_{-\infty:0}) \nu(dy_1) \right]. \end{aligned}$$

This can be more compactly written as

$$\begin{aligned} \mathbb{P}^\theta(X_1 \in A, Y_1 \in B, Y_{-\infty:0} \in C) &= \tilde{\mathbb{E}}^\theta \left[ \int \hat{\Phi}^\theta(Y_{-\infty:0}, y_1; A) \right. \\ & \quad \left. \mathbb{1}_B(y_1) \mathbb{1}_C(Y_{-\infty:0}) \Phi^\theta(Y_{-\infty:0}; dx_0) \kappa^\theta \langle Y_0, y_1 \rangle (x_0; \mathbf{X}) \nu(dy_1) \right]. \end{aligned} \quad (2.48)$$

Observe that (2.47) with  $A = \mathbf{X}$  provides a way to write  $\tilde{\mathbb{E}}^\theta [g(Y_{-\infty:0}, Y_1)]$  for  $g = \mathbb{1}_{C \times B}$  that can be extended to any nonnegative measurable function  $g$  defined on  $\mathbf{Y}^{\mathbb{Z}^-} \times \mathbf{Y}$  as

$$\begin{aligned} & \tilde{\mathbb{E}}^\theta [g(Y_{-\infty:0}, Y_1)] \\ &= \tilde{\mathbb{E}}^\theta \left[ \int g(Y_{-\infty:0}, y_1) \Phi^\theta(Y_{-\infty:0}; dx_0) \kappa^\theta \langle Y_0, y_1 \rangle (x_0; \mathbf{X}) \nu(dy_1) \right]. \end{aligned}$$

Now, we observe that the right-hand side of (2.48) can be interpreted as the right-hand side of the previous display with  $g(Y_{-\infty:0}, y_1) = \hat{\Phi}^\theta(Y_{-\infty:0}, y_1; A) \mathbb{1}_B(y_1) \mathbb{1}_C(Y_{-\infty:0})$ . Hence, we conclude that, for all  $A \in \mathcal{X}$  and  $C \in \mathcal{Y}^{\otimes \mathbb{Z}^-}$ ,

$$\mathbb{P}^\theta (X_1 \in A, Y_1 \in B, Y_{-\infty:0} \in C) = \mathbb{E}^\theta \left[ \hat{\Phi}^\theta(Y_{-\infty:0}, Y_1; A) \mathbb{1}_B(Y_1) \mathbb{1}_C(Y_{-\infty:0}) \right].$$

Notice that  $\hat{\Phi}^\theta(Y_{-\infty:0}, Y_1; A)$  precisely is the probability kernel on  $(\mathbf{Y}^{\mathbb{Z}^-} \times \mathbf{Y}) \times \mathcal{X}$  appearing on the left-hand side of (2.10). The last display implies that this probability kernel is the conditional distribution of  $X_1$  given  $Y_{-\infty:1}$  under  $\mathbb{P}^\theta$ , which completes the proof of (2.10).

### 2.5.2 Proof of Lemma 2.2.9

First observe that, by induction on  $n$ , having (2.15) for all  $n \geq 2$  is equivalent to having, for all  $n \geq 2$ ,

$$\begin{aligned} & p^{\theta, \theta^*}(Y_{1:n} | Y_{-\infty:0}) \\ &= p^{\theta, \theta^*}(Y_n | Y_{-\infty:n-1}) p^{\theta, \theta^*}(Y_{n-1} | Y_{-\infty:n-2}) \cdots p^{\theta, \theta^*}(Y_1 | Y_{-\infty:0}), \quad \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.}, \end{aligned}$$

which, using that  $\tilde{\mathbb{P}}^{\theta^*}$  is shift-invariant, is in turn equivalent to having that, for all  $n \geq 2$ ,

$$p^{\theta, \theta^*}(Y_{1:n} | Y_{-\infty:0}) = p^{\theta, \theta^*}(Y_{2:n} | Y_{-\infty:1}) p^{\theta, \theta^*}(Y_1 | Y_{-\infty:0}), \quad \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.} \quad (2.49)$$

Thus to conclude the proof, we only need to show that (2.49) holds for all  $n \geq 2$ . By Definition 2.2.8, we have, for all  $n \geq 2$  and  $y_{-\infty:n} \in \mathbf{Y}^{\mathbb{Z}^-}$ ,

$$\begin{aligned} & p^{\theta, \theta^*}(y_{2:n} | y_{-\infty:1}) p^{\theta, \theta^*}(y_1 | y_{-\infty:0}) \\ &= \int \Phi^{\theta, \theta^*}(y_{-\infty:1}; dx_1) p^{\theta, \theta^*}(y_1 | y_{-\infty:0}) \prod_{k=1}^{n-1} \kappa^\theta \langle y_k, y_{k+1} \rangle (x_k; dx_{k+1}). \end{aligned}$$

Using (K-2) we now get, for all  $n \geq 2$ ,

$$\begin{aligned} p^{\theta, \theta^*}(Y_{2:n}|Y_{-\infty:1})p^{\theta, \theta^*}(Y_1|Y_{-\infty:0}) \\ = \int \Phi^{\theta, \theta^*}(Y_{-\infty:0}; dx_0) \prod_{k=0}^{n-1} \kappa^\theta \langle Y_k, Y_{k+1} \rangle(x_k; dx_{k+1}), \quad \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.} \end{aligned}$$

We conclude (2.49) by observing that, according to Definition 2.2.8, the second line of the last display is  $p^{\theta, \theta^*}(Y_{1:n}|Y_{-\infty:0})$ .

### 2.5.3 Proof of Lemma 2.3.9

Let  $\beta \in [1, \alpha - 1)$ . Since  $1 + \beta < \alpha$  and by (E-1)-(b), we obtain  $\mathbb{E}[(U_0^+)^{1+\beta}] < \infty$ . Combining this with  $\mathbb{E}[U_0 - m] = -m < 0$ , we may apply [Jarner and Roberts, 2002, Proposition 5.1] so that the Markov kernel  $Q^\theta$  is polynomially ergodic and thus admits a unique stationary distribution  $\pi_1^\theta$ , which is well-defined on  $\mathbf{X} = \mathbb{R}_+$ . Moreover, [Jarner and Roberts, 2002, Proposition 5.1] also shows that there exist a finite interval  $C = [0, x_0]$  and some constants  $\varrho, \varrho' \in (0, \infty)$  such that

$$Q^\theta V \leq V - \varrho W + \varrho' \mathbb{1}_C,$$

where  $V(x) = (1+x)^{1+\beta}$  and  $W(x) = (1+x)^\beta$ . Applying [Meyn and Tweedie, 1993, Theorem 14.0.1] yields

$$\int \pi_1^\theta(dx) x^\beta \leq \pi_1^\theta W < \infty.$$

It remains to show that the kernel  $Q^\theta$  is not geometrically ergodic for all  $\theta \in \Theta$  and this will be done by contradiction.

Now suppose on the contrary that the kernel  $Q^\theta$  is geometrically ergodic for some  $\theta \in \Theta$ . Since the singleton  $\{0\}$  is an accessible atom (for  $Q^\theta$ ), then there exists some  $\rho > 1$  such that

$$\sum_{k=0}^{\infty} \rho^k |(Q^\theta)^k(0, \{0\}) - \pi_1^\theta(\{0\})| < \infty.$$

Hence, the atom  $\{0\}$  is geometrically ergodic as defined in [Meyn and Tweedie, 1993, Section 15.1.3]. Applying [Meyn and Tweedie, 1993, Theorem 15.1.5], then there exists some  $\kappa > 1$  such that  $\mathbb{E}_0[\kappa^{\tau_0}] < \infty$ , where  $\tau_0 = \inf\{n \geq 1 : X_n = 0\}$  is the first return time to  $\{0\}$ .

Recall that the i.i.d. sequence  $\{U_k : k \in \mathbb{Z}_+\}$  is linked to  $\{X_k : k \in \mathbb{Z}_+\}$  through (2.1), and note that  $\mathbb{E}_0[\kappa^{\tau_0}] = \mathbb{E}[\kappa^{\tau(0)}]$ , where we have set for all  $u \in \mathbb{R}$ ,

$$\tau(u) := \inf \left\{ n \geq 1 : \sum_{k=1}^n (U_k - m) < u \right\}.$$

Now, denote

$$\tilde{\tau}(u) := \inf \left\{ n \geq 1 : \sum_{k=1}^n (U_{k+1} - m) < u \right\}.$$

To arrive at the contradiction, it is finally sufficient to show that for all  $\kappa > 1$ ,  $\mathbb{E}[\kappa^{\tau(0)}] = \infty$ . Actually, we will show that there exists a constant  $\gamma > 0$  such that

$$\liminf_{u \rightarrow \infty} \kappa^{-\gamma u} \mathbb{E}[\kappa^{\tau(-u+m)}] > 0. \quad (2.50)$$

This will indeed imply  $\mathbb{E}[\kappa^{\tau(0)}] = \infty$  by writing

$$\begin{aligned} \mathbb{E}[\kappa^{\tau(0)}] &\geq \mathbb{E}[\kappa^{\tau(0)} \mathbb{1}\{U_1 \geq m\}] = \mathbb{E}[\kappa^{1+\tilde{\tau}(-U_1+m)} \mathbb{1}\{U_1 \geq m\}] \\ &= \mathbb{E} \left[ \int_m^\infty \kappa^{1+\tilde{\tau}(-u+m)} r(u) du \right] = \kappa \int_m^\infty \mathbb{E}[\kappa^{\tau(-u+m)}] r(u) du, \end{aligned} \quad (2.51)$$

where the last equality follows from  $\tau \stackrel{d}{=} \tilde{\tau}$ . Provided that (2.50) holds, the right-hand side of (2.51) is infinite since  $r(u) \gtrsim u^{-\alpha-1}$  as  $u \rightarrow \infty$  by (E-1)-(b).

We now turn to the proof of (2.50). By Markov's inequality, we have for any  $\gamma > 0$ ,

$$\kappa^{-\gamma u} \mathbb{E}[\kappa^{\tau(-u+m)}] \geq \mathbb{P}(\tau(-u+m) > \gamma u). \quad (2.52)$$

Now, let  $M_n = \sum_{k=1}^n U_i$ ,  $n \geq 1$ , and note that for all nonnegative  $u$ ,

$$\begin{aligned} \left\{ \left( \inf_{1 \leq k \leq \gamma u} M_k \right) - \gamma u m \geq -u + m \right\} &\subset \left\{ \inf_{1 \leq k \leq \gamma u} (M_k - km) \geq -u + m \right\} \\ &= \{\tau(-u+m) > \gamma u\}. \end{aligned} \quad (2.53)$$

Moreover, since  $\{U_k : k \in \mathbb{Z}_+\}$  is i.i.d. and centered, Doob's maximal inequality implies, for all  $\tilde{\gamma} > 0$ ,

$$\begin{aligned} \mathbb{P} \left( \inf_{1 \leq k \leq \gamma u} M_k < -\tilde{\gamma} \right) &\leq \mathbb{P} \left( \sup_{1 \leq k \leq \gamma u} |M_k| > \tilde{\gamma} \right) \\ &\leq \frac{\mathbb{E}[|M_{\lfloor \gamma u \rfloor}|]}{\tilde{\gamma}} \leq \frac{\lfloor \gamma u \rfloor \mathbb{E}[|U_1|]}{\tilde{\gamma}}. \end{aligned} \quad (2.54)$$

Now, pick  $\gamma > 0$  sufficiently small so that  $\gamma \mathbb{E}[|U_1|] / (1 - \gamma m) < 1$ . Observe that for this  $\gamma$ ,  $\tilde{\gamma} = (1 - \gamma m)u - m$  is positive for  $u$  sufficiently large, so that combining (2.54) with (2.53) and (2.52) yields

$$\liminf_{u \rightarrow \infty} \kappa^{-\gamma u} \mathbb{E}[\kappa^{\tau(-u+m)}] \geq 1 - \limsup_{u \rightarrow \infty} \frac{\lfloor \gamma u \rfloor \mathbb{E}[|U_1|]}{(1 - \gamma m)u - m} = 1 - \frac{\gamma \mathbb{E}[|U_1|]}{1 - \gamma m} > 0.$$

This shows (2.50) and the proof is complete.



## Chapter 3

# Consistency of the MLE for the Class of Fully Dominated Markov Chains

**Summary.** In this chapter, we investigate the consistency of the maximum likelihood estimator (MLE) for the class of fully dominated partially observed Markov models (PMMs). Following the approach by [Douc and Moulines \[2012\]](#) as to derive the convergence of the MLE to the minimizing set of the relative entropy rate in misspecified hidden Markov models (HMMs), under mild sufficient conditions, the convergence of the MLE for this class is established. This result is obtained, regardless of whether or not the model is well specified, as a byproduct of the convergence of the block-type MLE of some likelihood kernel. Moreover, when the model is well specified, following [Douc et al. \[2014\]](#), under these same conditions the identifiability is also obtained, establishing the equivalence-class consistency of the MLE.

### 3.1 Introduction

Maximum likelihood method is widely used for identifying a parametric model of a time series from a sample of observations. When the studied model is well-specified, it is of prime interest to obtain the consistency of the estimator, namely, its convergence to the true parameter, say  $\theta_*$ , as the sample size of the observations goes to infinity. The proof toward this property usually involves two important steps: 1) the maximum likelihood estimator (MLE) converges to the maximizing set, say  $\Theta_*$ , of the asymptotic normalized log-likelihood, and 2) this maximizing set  $\Theta_*$  is indeed the singleton set

of the true parameter  $\theta_*$ . Identifying the maximizing set  $\Theta_*$  in the second step is usually referred to as solving the *identifiability* problem. This task can actually be split in two sub-problems, that is, 2.1) showing that any parameter in  $\Theta_*$  yields the same distribution for the observations as for the true parameter, and 2.2) showing that for a sufficiently large sample size, the set of such parameters reduces to  $\theta_*$ . Problem 2.2 can be hard to solve, for example, for the models that belong to the class of hidden Markov models (HMMs), see [Allman et al. \[2009\]](#), [Gassiat et al. \[2016\]](#) and the references therein for recent advances in the case of HMMs. Nevertheless, Problem 2.1 can be handled independently, and with Step 1 above, this directly yields that the MLE is consistent in a weakened sense that the estimator converges to the set of all the parameters associated to the same distribution as the one of the observed sample. The consistency of this type, according [Leroux \[1992\]](#), is referred to as *equivalence-class consistency*.

In chapter, we investigate problem in Step 1 and Problem 2.1 together, that is, deriving the equivalence-class consistency of the MLE under a general framework of time series models, namely, the class of fully dominated PMMs. Many HMMs developed by Baum and coworkers (see [Cappé et al. \[2005\]](#) and the inferences therein for a list of examples and a comprehensive account on HMMs), Markov switching models ([Hamilton \[1988, 1989\]](#)) and many of Markov chains in random environment introduced by [Cogburn \[1980\]](#) and [Bourgin and Cogburn \[1981\]](#), for example, are members of this class (see also [Pieczynski \[2003\]](#) or [Ephraim and Mark \[2012\]](#) for other variants).

Formulating the convergence of the MLE in Step 1 is carried out by relying on the approach employed by [Douc and Moulines \[2012\]](#) as to derive the convergence of the MLE to the minimizing set of the relative entropy rate in misspecified HMMs. This approach consists in establishing a key property, namely, the exponential forgetting of the filtering distribution, by using coupling method originated by [Kleptsyna and Veretennikov \[2008\]](#) and further refined by [Douc et al. \[2009\]](#). When the forgetting of the filtering distribution is achieved, together with the assumption that the observation process is stationary and ergodic, then the normalized log-likelihood can be approximated by an appropriately defined stationary version of it. Then applying the results in classical ergodic theory, the latter is shown to converge to a limit of functional of the parameter, say  $\ell(\theta)$ , where  $\theta$  is the parameter. It then implies by using standard argument (see, for instance, [Pfanzagl \[1969\]](#)) that the MLE of the normalized log-likelihood converges almost surely to a subset  $\Theta_*$  of parameters at which the limit  $\ell(\theta)$  is maximized. Using this same technique, meanwhile, we obtain an intermediary result showing that the block-type MLE converges to a maximizing set in a similar sense as the classical MLE, regardless of whether or not the models are well-specified.

These results are shown under conditions similar to the conditions derived in [Douc and Moulines \[2012\]](#).

To complete, we solve Problem 2.1 by applying the general result derived in [Douc et al. \[2014\]](#) under the general framework of PMMs. Under these same assumptions, the equivalence-class consistency of the MLE for the class fully dominated PMMs is established.

The organization of this chapter is as follows. In [Section 3.2](#), we establish the convergence of the MLE for the block-type likelihood function. We then apply this result in [Section 3.3](#) to derive the consistency of the MLE for the class of fully dominated PMMs. Finally, the postponed proofs and some useful lemmas are gathered in [Section 3.4](#).

## 3.2 Consistency of the Block-type Maximum Likelihood Estimators

### 3.2.1 Notation and Definitions

Let us first recall some notation pertaining to transition kernels in measure theory. Let  $L$  be a (possibly unnormalized) transition kernel defined on a measurable space  $(\mathsf{S}, \mathcal{S})$ , that is,

- for all  $s \in \mathsf{S}$ ,  $L(s; \cdot)$  is a nonnegative measure on  $(\mathsf{S}, \mathcal{S})$  and
- for all  $A \in \mathcal{S}$ ,  $s \mapsto L(s; A)$  is a measurable function.

Let  $f$  be a (nonnegative) measurable function from  $(\mathsf{S}, \mathcal{X})$  to  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , where  $\mathcal{B}(\mathbb{R})$  denotes the Borel  $\sigma$ -field on  $\mathbb{R}$ , and let  $\mu$  be a  $\sigma$ -finite positive measure on  $(\mathsf{S}, \mathcal{S})$ . The transition kernel  $L$  acts on the function  $f$  and the measure  $\mu$  via

$$Lf(s) = L(s; f) := \int L(s; ds')f(s'),$$

$$\mu L(A) = \mu L\mathbb{1}_A := \int \mu(dx)L(s; A).$$

Moreover, if  $L_1$  and  $L_2$  are two transition kernels on  $(\mathsf{S}, \mathcal{S})$ , then the kernel application  $L_1L_2$  also defines a transition kernel on  $(\mathsf{S}, \mathcal{S})$  and is given by, for any  $s \in \mathsf{S}$  and  $A \in \mathcal{S}$ ,

$$L_1L_2(s; A) := \int L_1(s; ds')L_2(s'; A).$$

Now consider the following setting. Let  $(\mathsf{X}, \mathcal{X})$  and  $(\mathsf{Z}, \mathcal{Z})$  be two Borel spaces and  $\Theta$  denote a parameter space, which throughout this chapter is

moreover assumed to be a compact metric space equipped with the metric  $\Delta$ . Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space, and let  $\{Z_k : k \in \mathbb{Z}\}$  be a stochastic process taking values on  $(Z, \mathcal{Z})$ , which is assumed to be observed. Within this section, we assume the following assumption.

(L-1) The stochastic process  $\{Z_k : k \in \mathbb{Z}\}$  is stationary and ergodic on  $(Z, \mathcal{Z})$ .

We denote by  $\mathbb{P}_Z$  the image probability of  $\mathbb{P}$  induced by the stationary process  $\{Z_k : k \in \mathbb{Z}\}$  on the product space  $(Z^{\mathbb{Z}}, \mathcal{Z}^{\otimes \mathbb{Z}})$  and denote by  $\mathbb{E}_Z$  the associated expectation. Let  $\{\mathbf{L}^\theta \langle z \rangle : (\theta, z) \in \Theta \times Z\}$  be a family of transition kernels on  $(X, \mathcal{X})$  such that for all  $\theta \in \Theta$ , probability measure  $\chi$  on  $(X, \mathcal{X})$  and nonnegative measurable function  $f$  on  $X$ , the function  $z \mapsto \chi \mathbf{L}^\theta \langle z \rangle f$  is measurable on  $(Z, \mathcal{Z})$ . Here we should stress that this family of transition kernels may or may not link to the observed stochastic process  $\{Z_k : k \in \mathbb{Z}\}$ . Let us now define the following.

**Definition 3.2.1** (Likelihood kernel). Let  $\theta \in \Theta$ ,  $n \in \mathbb{Z}_+^*$  and  $z_{0:n-1} \in Z^n$ . The (*parametric*) *likelihood kernel* with respect to the family  $\{\mathbf{L}^\theta \langle z \rangle : (\theta, z) \in \Theta \times Z\}$  of the sample  $z_{0:n-1}$ , denoted by  $\mathbf{L}_n^\theta \langle z_{0:n-1} \rangle$ , is defined as the successive kernel application of the transition kernels  $\mathbf{L}^\theta \langle z_0 \rangle, \dots, \mathbf{L}^\theta \langle z_{n-1} \rangle$  from the family  $\{\mathbf{L}^\theta \langle z \rangle : (\theta, z) \in \Theta \times Z\}$ . Formally, for all  $(x_0, A) \in X \times \mathcal{X}$ ,

$$\mathbf{L}_n^\theta \langle z_{0:n-1} \rangle(x_0; A) := \int \mathbf{L}^\theta \langle z_0 \rangle(x_0; dx_1) \cdots \mathbf{L}^\theta \langle z_{n-1} \rangle(x_{n-1}; dx_n) \mathbb{1}_A(x_n). \quad (3.1)$$

Note that by definition,  $\mathbf{L}_n^\theta \langle z_{0:n-1} \rangle$  also defines a transition kernel on  $(X, \mathcal{X})$ . Moreover, for all  $\theta \in \Theta$ ,  $n, p \in \mathbb{Z}_+^*$  and  $z_{0:n+p-1} \in Z^{n+p}$ , the identity:

$$\mathbf{L}_{n+p}^\theta \langle z_{0:n+p-1} \rangle = \mathbf{L}_n^\theta \langle z_{0:n-1} \rangle \mathbf{L}_p^\theta \langle z_{n:n+p-1} \rangle$$

holds true. In what follows,  $\mathcal{P}(X, \mathcal{X})$  will always denote the set of probability measures on  $(X, \mathcal{X})$ . For notational convenience, when there is no possibility of confusion, the subscript  $n$  will be dropped from  $\mathbf{L}_n^\theta$  which is defined earlier in (3.1).

Next, we always assume that for all  $\theta \in \Theta$ ,  $n \in \mathbb{Z}_+^*$ ,  $z_{0:n-1} \in Z^n$  and probability measure  $\chi \in \mathcal{P}(X, \mathcal{X})$ ,

$$0 < \chi \mathbf{L}^\theta \langle z_{0:n-1} \rangle \mathbb{1}_X < \infty. \quad (3.2)$$

The following definition defines the conditional version of the likelihood kernel.

**Definition 3.2.2.** Let  $\theta \in \Theta$ . For all  $m, n \in \mathbb{Z}_+$  with  $m < n$ ,  $z_{m:n} \in Z^{n-m+1}$ ,  $\chi \in \mathcal{P}(X, \mathcal{X})$  and measurable function  $f$  from  $(X, \mathcal{X})$  to  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , define

$$\chi \mathbf{L}^\theta \langle z_{p:n} | z_{m:p-1} \rangle f := \frac{\chi \mathbf{L}^\theta \langle z_{m:n} \rangle f}{\chi \mathbf{L}^\theta \langle z_{m:p-1} \rangle \mathbb{1}_X}, \quad m < p \leq n. \quad (3.3)$$

**Definition 3.2.3.** Let  $n \in \mathbb{Z}_+^*$ , the sequence  $Z_{0:n-1}$  be observations realized from the stochastic process  $\{Z_k : k \in \mathbb{Z}\}$  and let  $x_0 \in \mathbf{X}$ . The function  $\theta \mapsto \mathbf{L}_n^\theta \langle Z_{0:n-1} \rangle (x_0; \mathbf{X})$  is called the (*conditional*) *L-likelihood function* with respect to the family  $\{\mathbf{L}^\theta \langle z \rangle : (\theta, z) \in \Theta \times \mathbf{Z}\}$  of the observations  $Z_{0:n-1}$  given  $X_0 = x_0$ . We define the *maximum L-likelihood estimator* with respect to the family  $\{\mathbf{L}^\theta \langle z \rangle : (\theta, z) \in \Theta \times \mathbf{Z}\}$  associated to an arbitrary initial probability distribution  $\chi \in \mathcal{P}(\mathbf{X}, \mathcal{X})$  by

$$\hat{\theta}_{\chi,n}^L \in \operatorname{argmax}_{\theta \in \Theta} \chi \mathbf{L}_n^\theta \langle Z_{0:n-1} \rangle \mathbb{1}_{\mathbf{X}}. \quad (3.4)$$

When  $\chi$  is  $\delta_x$ , the Dirac mass at point  $x \in \mathbf{X}$ , the maximum *L-likelihood estimator* given in (3.4) will be simply denoted by  $\hat{\theta}_{x,n}^L$ .

So far, we have not mentioned any specific statistical model, except that we have introduced the family of transition kernels  $\{\mathbf{L}^\theta \langle z \rangle : (\theta, z) \in \Theta \times \mathbf{Z}\}$ . To have some insights on how these kernels may arise in specific situations, consider the following examples.

**Example 3.2.4.** Consider a family of HMMs  $\{(X_k, Y_k) : k \in \mathbb{Z}_+\}$  on  $(\mathbf{X} \times \mathbf{Y}, \mathcal{X} \otimes \mathcal{Y})$ , which is parametrized by  $\theta \in \Theta$ . For each  $\theta \in \Theta$ , the distribution of the HMM is specified by the transition kernel  $Q^\theta$  of the Markov chain  $\{X_k : k \in \mathbb{Z}_+\}$  and by the conditional probability density  $g^\theta$  of the observation  $Y_k$  given the hidden state  $X_k$ , which is also known as the likelihood of the observation (see [Definition 2.3.1](#) in [Section 2.3](#) of [Chapter 2](#)). The dynamic of hidden Markov process can be depicted as in [Figure 3.1](#).

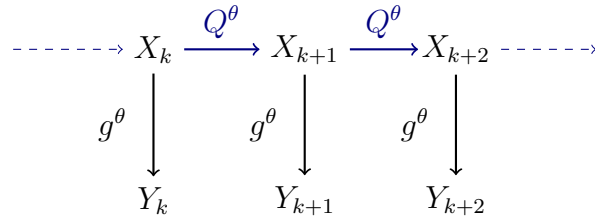


Figure 3.1 – Graphical representation of the hidden Markov Model by directed arrows.

For any probability measure  $\chi$  on  $(\mathbf{X}, \mathcal{X})$  and positive integer  $r$ , the likelihood of the observations  $Y_{0:r-1} \in \mathbf{Y}^r$  associated to  $\chi$ , denoted by  $p_\chi^\theta(Y_{0:r-1})$ , can be expressed by

$$p_\chi^\theta(Y_{0:r-1}) = \int \chi(dx_0) g^\theta(x_0; Y_0) Q^\theta(x_0; dx_1) \cdots g^\theta(x_{r-1}; Y_{r-1}) Q^\theta(x_{r-1}; dx_r).$$

Now, if we set  $Z := Y^r$ ,  $\mathcal{Z} := \mathcal{Y}^{\otimes r}$ ,  $Z_0 := Y_{0:r-1}$ , and for all  $\theta \in \Theta$ ,  $x_0 \in \mathsf{X}$  and  $A \in \mathcal{X}$ ,

$$\mathbf{L}^\theta \langle Z_0 \rangle (x_0; A) := \int \left( \prod_{k=0}^{r-1} g^\theta(x_k; Y_k) Q^\theta(x_k; dx_{k+1}) \right) \mathbb{1}_A(x_r),$$

we then have that  $\chi \mathbf{L}^\theta \langle Z_0 \rangle \mathbb{1}_\mathsf{X}$  is indeed the likelihood of the observations  $Y_{0:r-1}$  associated to the initial probability measure  $\chi \in \mathcal{P}(\mathsf{X}, \mathcal{X})$ . Moreover, for all  $n \in \mathbb{Z}_+^*$ , if we set  $Z_i := Y_{ir:(i+1)r-1}$ ,  $i \in \{0, \dots, n-1\}$ , the conditional likelihood kernel defined in (3.1) is given by

$$\mathbf{L}^\theta \langle Z_{0:n-1} \rangle (x_0; A) = \int \left( \prod_{k=0}^{nr-1} g^\theta(x_k; Y_k) Q^\theta(x_k; dx_{k+1}) \right) \mathbb{1}_A(x_{nr}).$$

This implies  $\chi \mathbf{L}^\theta \langle Z_{0:n-1} \rangle \mathbb{1}_\mathsf{X} = p_\chi^\theta(Y_{0:nr-1})$ , which is referred to as the block likelihood function with block size equal to  $r$  in Douc and Moulines [2012]. In this case, the maximum  $L$ -likelihood estimator  $\hat{\theta}_{\chi,n}^L$  coincides with the block MLE  $\hat{\theta}_{\chi,nr}$ , a maximizer of  $p_\chi^\theta(Y_{0:nr-1})$  on  $\Theta$ . It turns out that, when  $r = 1$ , the class of HMMs which is specified by  $\{(Q^\theta, g^\theta) : \theta \in \Theta\}$ , can also be specified by the family  $\{\mathbf{L}^\theta \langle z \rangle : (\theta, z) \in \Theta \times \mathsf{Z}\}$ .

**Example 3.2.5.** Let  $\{(X_k, Y_k) : k \in \mathbb{Z}_+\}$  be a parametric family of bivariate Markov chains on  $(\mathsf{X} \times \mathsf{Y}, \mathcal{X} \otimes \mathcal{Y})$ , whose parameters are  $\theta \in \Theta$ . As in HMM case, we assume that only the process  $\{Y_k : k \in \mathbb{Z}_+\}$  is observed. To formulate the model, we moreover suppose that for each  $\theta \in \Theta$ , the process  $\{X_k : k \in \mathbb{Z}_+\}$  is a Markov chain with transition kernel  $Q^\theta$ ; and conditionally on the past variables up to time  $k-1$  with  $k \geq 1$ , the law of the observation  $Y_k$  has density  $g^\theta((X_{k-1}, Y_{k-1}); \cdot)$  with respect to a fixed  $\sigma$ -finite positive measure  $\nu$  on  $(\mathsf{Y}, \mathcal{Y})$ . The whole chain can then be fully specified by the pair  $(Q^\theta, g^\theta)$ . By this definition, it is straightforward that for each  $\theta \in \Theta$  and for all  $k \geq 0$  and  $B \in \mathcal{Y}$ , we have

$$\mathbb{P}^\theta(Y_{k+1} \in B \mid Y_{0:k}, X_{0:\infty}) = \int_B g^\theta((X_k, Y_k); y) \nu(dy). \quad (3.5)$$

Following the terminology of Cogburn [1980] and Bourgin and Cogburn [1981], (3.5) suggests that the bivariate chain  $\{(X_k, Y_k) : k \in \mathbb{Z}_+\}$  is a Markov chain in random environment with the Markov environment process  $\{X_k : k \in \mathbb{Z}_+\}$ . Markov chains in random environment have been studied extensively by Cogburn [1980], Bourgin and Cogburn [1981], Nawrotzki [1982], Cogburn [1984, 1990, 1991], Orey [1991] in the case of countable spaces and

by Van Handel [2009] in the case of general state spaces. Random walk in random environment models introduced by Chernov [1967] are also embedded in this class (see Comets et al. [2014a] and Comets et al. [2014b] for parametric versions of it). The dynamic of this bivariate process can be depicted as in Figure 3.2. However, according to Nawrotzki [1982], we should be aware that not all Markov chains in random environment are themselves Markov chains; and thus Figure 3.2 depicts the case where the bivariate processes are Markov only.

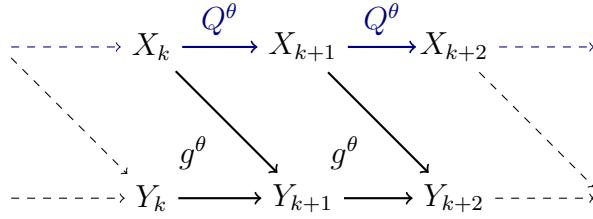


Figure 3.2 – Graphical representation of the Markov chain in random environment

For any  $\chi \in \mathcal{P}(\mathsf{X}, \mathcal{X})$  and  $r \in \mathbb{Z}_+^*$ , the likelihood of the observations  $Y_{0:r} \in \mathsf{Y}^r$  associated to  $\chi$ , denoted by  $p_{\chi \otimes \delta_{Y_0}}^\theta(Y_{1:r})$ , is defined by

$$p_{\chi \otimes \delta_{Y_0}}^\theta(Y_{1:r}) = \int \chi(dx_0) \left( \prod_{k=0}^{r-1} g^\theta((x_k, Y_k); Y_{k+1}) Q^\theta(x_k; dx_{k+1}) \right).$$

By setting  $\mathsf{Z} := \mathsf{Y}^{r+1}$ ,  $\mathcal{Z} := \mathcal{Y}^{\otimes(r+1)}$ ,  $Z_0 := Y_{0:r}$ , and for all  $\theta \in \Theta$ ,  $x_0 \in \mathsf{X}$  and  $A \in \mathcal{X}$ ,

$$\mathbf{L}^\theta \langle Z_0 \rangle(x_0; A) := \int \left( \prod_{k=0}^{r-1} g^\theta((x_k, Y_k); Y_{k+1}) Q^\theta(x_k; dx_{k+1}) \right) \mathbb{1}_A(x_r),$$

we then see that  $\chi \mathbf{L}^\theta \langle Z_0 \rangle \mathbb{1}_{\mathsf{X}}$  is indeed the likelihood of the observations  $Y_{0:r}$  associated to the probability measure  $\chi$ . Moreover, for any integer  $n > 0$ , by setting  $Z_i := Y_{ir:(i+1)r}$ ,  $i \in \{0, \dots, n-1\}$ , the conditional likelihood kernel defined in (3.1) writes

$$\mathbf{L}^\theta \langle Z_{0:n-1} \rangle(x_0; A) = \int \left( \prod_{k=0}^{nr-1} g^\theta((x_k, Y_k); Y_{k+1}) Q^\theta(x_k; dx_{k+1}) \right) \mathbb{1}_A(x_{nr}).$$

This yields that  $\chi \mathbf{L}^\theta \langle Z_{0:n-1} \rangle \mathbb{1}_{\mathsf{X}} = p_{\chi \otimes \delta_{Y_0}}^\theta(Y_{1:nr})$ , which, as in Example 3.2.4, may be referred to as the block likelihood function with block size equal to

$r$ . In this case, the maximum  $L$ -likelihood estimator  $\hat{\theta}_{\chi,n}^L$  coincides with the block MLE  $\hat{\theta}_{\chi,nr} \in \operatorname{argmax}_{\theta \in \Theta} p_{\chi \otimes \delta_{Y_0}}^\theta(Y_{1:nr})$ .

**Example 3.2.6.** Consider a parametrized family of Markov switching models  $\{(X_k, Y_k) : k \in \mathbb{Z}_+\}$  on  $(\mathcal{X} \times \mathcal{Y}, \mathcal{X} \otimes \mathcal{Y})$  with parameters  $\theta \in \Theta$ . As in HMM case, we assume that only the process  $\{Y_k : k \in \mathbb{Z}_+\}$  is observed. For each  $\theta \in \Theta$ , the distribution of the Markov switching model is specified by the transition kernel  $Q^\theta$  of the Markov chain  $\{X_k : k \in \mathbb{Z}_+\}$  and by the conditional probability density  $g^\theta$  of the observation  $Y_k$  given the hidden state  $X_k$  and the most recent past observation  $Y_{k-1}$  ( $g^\theta$  is also known as the likelihood of the observation). The Markov switching model can be specified by the pair  $(Q^\theta, g^\theta)$  and can be graphically represented as in [Figure 3.3](#). More formally, the bivariate stochastic process  $\{(X_k, Y_k) : k \in \mathbb{Z}_+\}$  is a parametric Markov switching model if for all  $\theta \in \Theta$ ,  $k \in \mathbb{Z}_+$ ,  $A \in \mathcal{X}$  and  $B \in \mathcal{Y}$ , we have

$$\begin{aligned} \mathbb{P}^\theta((X_{k+1}, Y_{k+1}) \in A \times B \mid Y_{0:k}, X_{0:k}) \\ = \int_{A \times B} Q^\theta(X_k; dx) g^\theta((Y_k, x); y) \nu(dy). \end{aligned} \quad (3.6)$$

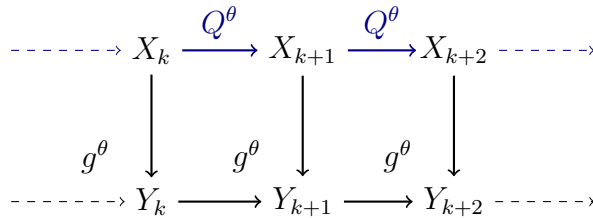


Figure 3.3 – Graphical representation of the Markov switching model.

Markov switching models have been introduced by [Hamilton \[1988, 1989\]](#) and further considered by many authors including [Hamilton \[1996\]](#), [Timmermann \[2000\]](#), [Fong and See \[2002\]](#), [Yoshida et al. \[2005\]](#), [Hamilton \[2008\]](#), [Bauwens et al. \[2010\]](#), [Ailliot and Monbet \[2012\]](#); see also [Frühwirth-Schnatter \[2006\]](#) for a comprehensive account. Note that if the transition kernel  $Q^\theta$  in (3.6) is also allowed to depend on  $Y_k$ , that is  $Q^\theta$  is no longer a Markov kernel, the resulting model is called *non-homogeneous Markov switching*, see [Ailliot and Pene \[2013\]](#) and [Ailliot et al. \[2015\]](#).

For any  $\chi \in \mathcal{P}(\mathcal{X}, \mathcal{X})$  and  $r \in \mathbb{Z}_+^*$ , the likelihood of the observations



$Y_{0:r} \in \mathcal{Y}^r$  associated to  $\chi$ , denoted by  $p_{\chi \otimes \delta_{Y_0}}^\theta(Y_{1:r})$ , is defined by

$$p_{\chi \otimes \delta_{Y_0}}^\theta(Y_{1:r}) := \int \chi(dx_0) \left( \prod_{k=0}^{r-1} Q^\theta(x_k; dx_{k+1}) g^\theta((Y_k, x_{k+1}); Y_{k+1}) \right).$$

By setting  $\mathbf{Z} := \mathcal{Y}^{r+1}$ ,  $\mathcal{Z} := \mathcal{Y}^{\otimes r+1}$ ,  $Z_0 := Y_{0:r}$ , and for all  $\theta \in \Theta$ ,  $x_0 \in \mathbf{X}$  and  $A \in \mathcal{X}$ ,

$$\mathbf{L}^\theta \langle Z_0 \rangle (x_0; A) := \int \left( \prod_{k=0}^{r-1} Q^\theta(x_k; dx_{k+1}) g^\theta((Y_k, x_{k+1}); Y_{k+1}) \right) \mathbb{1}_A(x_r),$$

we then see that  $\chi \mathbf{L}^\theta \langle Z_0 \rangle \mathbb{1}_{\mathbf{X}}$  is indeed the likelihood of the observations  $Y_{0:r}$  associated to  $\chi$ . Moreover, for any integer  $n > 0$ , if setting  $Z_i := Y_{ir:(i+1)r}$ ,  $i \in \{0, \dots, n-1\}$ , the conditional likelihood kernel defined in (3.1) writes

$$\mathbf{L}^\theta \langle Z_{0:n-1} \rangle (x_0; A) = \int \left( \prod_{k=0}^{nr-1} Q^\theta(x_k; dx_{k+1}) g^\theta((Y_k, x_{k+1}); Y_{k+1}) \right) \mathbb{1}_A(x_{nr}).$$

This yields that  $\chi \mathbf{L}^\theta \langle Z_{0:n-1} \rangle \mathbb{1}_{\mathbf{X}} = p_{\chi \otimes \delta_{Y_0}}^\theta(Y_{1:nr})$ , which, as in [Example 3.2.4](#), may be referred to as the block likelihood function with block size equal to  $r$ . In this case, the maximum  $L$ -likelihood estimator  $\hat{\theta}_{\chi, n}^L$  coincides with the block MLE  $\hat{\theta}_{\chi, nr} \in \operatorname{argmax}_{\theta \in \Theta} p_{\chi \otimes \delta_{Y_0}}^\theta(Y_{1:nr})$ . By these examples, the maximum  $L$ -likelihood estimator  $\hat{\theta}_{\chi, n}^L$ , in certain circumstances, can be referred to as the block MLE.

The purpose of this present section is to investigate the asymptotic properties of the maximum  $L$ -likelihood estimator  $\hat{\theta}_{\chi, n}^L$  of the  $L$ -likelihood  $\chi \mathbf{L}^\theta \langle Z_{0:n-1} \rangle \mathbb{1}_{\mathbf{X}}$ . More precisely, we intend to establish that under some certain conditions there exists a non-empty subset  $\Theta_\star^L \subseteq \Theta$  such that  $\Delta(\hat{\theta}_{\chi, n}^L, \Theta_\star^L)$ , the distance between  $\hat{\theta}_{\chi, n}^L$  and  $\Theta_\star^L$ , converges  $\mathbb{P}_Z$ -a.s. to zero as  $n \rightarrow \infty$ . As illustrated by the above examples, the maximum  $L$ -likelihood estimator is closely related to the classical MLE and thus studying the asymptotic properties of the maximum  $L$ -likelihood estimator can be beneficial for the case of the MLE. Our proof that will proceed follows similar lines of [Douc and Moulines \[2012\]](#) as to obtain the same result when the family  $\{\mathbf{L}^\theta \langle z \rangle : (\theta, z) \in \Theta \times \mathbf{Z}\}$  corresponds to the class of HMMs (however, we should keep in mind that this family is far more general than that). As for the case of classical likelihood, the proof commonly involves two main steps. The first step consists in establishing the limit of the  $L$ -likelihood, that is to show that there exists a functional  $\mathbf{L}(\theta)$  in  $\theta \in \Theta$  such that for all  $\theta \in \Theta$ ,

$$\lim_{n \rightarrow \infty} n^{-1} \ln \chi \mathbf{L}^\theta \langle Z_{0:n-1} \rangle \mathbb{1}_{\mathbf{X}} = \mathbf{L}(\theta), \quad \mathbb{P}_Z\text{-a.s.} \quad (3.7)$$

The second step consists in proving that, when the first step is accomplished, as  $n \rightarrow \infty$ , the maximizer of  $\theta \mapsto \chi \mathbf{L}^\theta \langle Z_{0:n-1} \rangle \mathbb{1}_X$  converges  $\mathbb{P}_Z$ -a.s. to the maximizing set  $\Theta_*^L = \operatorname{argmax}_{\theta \in \Theta} \mathbf{L}(\theta)$ . Let us now briefly describe how these two essential steps are established herein. In formulating the first step, we first note that from (3.3) of Definition 3.2.2, we can write

$$n^{-1} \ln \chi \mathbf{L}^\theta \langle Z_{0:n-1} \rangle \mathbb{1}_X = n^{-1} \sum_{k=0}^{n-1} \ln \chi \mathbf{L}^\theta \langle Z_k | Z_{0:k-1} \rangle \mathbb{1}_X.$$

Next, we need to show that there exists a  $\mathbb{P}_Z$ -integrable function  $\pi_Z^\theta : Z^{\mathbb{Z}^-} \rightarrow \mathbb{R}$  such that  $|\chi \mathbf{L}^\theta \langle Z_k | Z_{0:k-1} \rangle \mathbb{1}_X - \pi_Z^\theta(Z_{-\infty:k-1})|$  converges  $\mathbb{P}_Z$ -a.s. to zero as  $k \rightarrow \infty$ . This result can be obtained by establishing the exponential forgetting of the initial distribution of the conditional  $L$ -likelihood, and to this end, we employ the coupling technique originally introduced in Kleptsyna and Veretennikov [2008] and enriched by Douc et al. [2009]. Then since the process  $\{Z_k : k \in \mathbb{Z}\}$  is stationary and ergodic, the Birkhoff ergodic theorem applies, and (3.7) follows with  $\mathbf{L}(\theta) = \mathbb{E}_Z [\ln \pi_Z^\theta(Z_{-\infty:0})]$ , provided that the latter quantity is finite (the approach used in this stage was originated by Baum and Petrie [1966] for finite state-space HMMs, and later extended by Douc et al. [2004] for general state-space HMMs, but under stringent technical conditions). When the first step is achieved, we apply Pfanzagl [1969] to obtain the second step, provided that we can show that  $\theta \mapsto \mathbf{L}(\theta)$  is semi-continuous.

The following assumptions can be used to yield the convergence of the maximum  $L$ -likelihood estimator  $\hat{\theta}_{X,n}^L$  of the  $L$ -likelihood  $\chi \mathbf{L}^\theta \langle Z_{0:n-1} \rangle \mathbb{1}_X$ . Before listing these assumptions, let us invoke the following useful definition.

**Definition 3.2.7** ( $L$ -Local Doeblin set). Let  $C \in \mathcal{X}$  be a nonempty subset of  $X$ . The set  $C$  is called a  $L$ -local Doeblin set with respect to the family  $\{\mathbf{L}^\theta \langle z \rangle : (\theta, z) \in \Theta \times Z\}$  if there exist positive functions  $\epsilon_C^- : Z \rightarrow \mathbb{R}_+$ ,  $\epsilon_C^+ : Z \rightarrow \mathbb{R}_+$ , a family of probability measures  $\{\mu_C^\theta \langle z \rangle : (\theta, z) \in \Theta \times Z\}$  on  $(X, \mathcal{X})$  and a family of positive functions  $\{\varphi_C^\theta \langle z \rangle : (\theta, z) \in \Theta \times Z\}$  from  $X$  to  $\mathbb{R}$  such that for any  $(\theta, z) \in \Theta \times Z$ ,  $\mu_C^\theta \langle z \rangle(C) = 1$ , and for any nonnegative and measurable function  $f : (X, \mathcal{X}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  and  $x \in C$ ,

$$\epsilon_C^- \langle z \rangle \varphi_C^\theta \langle z \rangle(x) \mu_C^\theta \langle z \rangle(\mathbb{1}_C f) \leq \delta_x \mathbf{L}^\theta \langle z \rangle(\mathbb{1}_C f) \leq \epsilon_C^+ \langle z \rangle \varphi_C^\theta \langle z \rangle(x) \mu_C^\theta \langle z \rangle(\mathbb{1}_C f). \quad (3.8)$$

**Remark 3.2.8.** The  $L$ -Local Doeblin set in Definition 3.2.7 is similar to the so-called  $r$ -local Doeblin set defined in [Douc and Moulines, 2012, Definition 1], except that the former set is defined with respect to the general family of kernels  $\{\mathbf{L}^\theta \langle z \rangle : (\theta, z) \in \Theta \times Z\}$ , whereas the latter set is defined with respect to the specific family of transition kernels that specify the HMMs.

(L-2) There exists a set  $K \in \mathcal{Z}$  such that the following hold.

- (i)  $\mathbb{P}_Z(Z_0 \in K) > 2/3$ .
- (ii) For all  $\eta > 0$  there exists a  $L$ -local Doeblin set  $C \in \mathcal{X}$  such that for all  $\theta \in \Theta$  and  $z \in K$ ,

$$\sup_{x \in C^c} \delta_x \mathbf{L}^\theta \langle z \rangle \mathbb{1}_X \leq \eta \sup_{x \in X} \delta_x \mathbf{L}^\theta \langle z \rangle \mathbb{1}_X < \infty \quad (3.9)$$

and

$$\inf_{z \in K} \frac{\epsilon_C^- \langle z \rangle}{\epsilon_C^+ \langle z \rangle} > 0, \quad (3.10)$$

where the functions  $\epsilon_C^+$  and  $\epsilon_C^-$  are given in [Definition 3.2.7](#).

(L-3) There exists a set  $D \in \mathcal{X}$  such that

$$\mathbb{E}_Z \left[ \ln^- \inf_{\theta \in \Theta} \inf_{x \in D} \mathbf{L}^\theta \langle Z_0 \rangle (x; D) \right] < \infty. \quad (3.11)$$

(L-4) We have

$$\mathbb{E}_Z \left[ \ln^+ \sup_{\theta \in \Theta} \sup_{x \in X} \mathbf{L}^\theta \langle Z_0 \rangle (x; X) \right] < \infty. \quad (3.12)$$

(L-5) There exists a positive integer  $p$  such that for all  $n \geq p$  and  $x \in X$ , the function  $\theta \mapsto \mathbf{L}^\theta \langle Z_{0:n-1} \rangle (x; X)$  is  $\mathbb{P}_Z$ -a.s. continuous on  $\Theta$ .

**Definition 3.2.9.** Let  $A \in \mathcal{X}$  be a measurable subset of  $X$ . The collection  $\mathcal{M}^+(A)$  denotes the set of all probability distributions on  $(X, \mathcal{X})$  that do not vanish on  $A$ , that is, if  $\chi \in \mathcal{P}(X, \mathcal{X})$  such that  $\chi(A) > 0$ , then  $\chi \in \mathcal{M}^+(A)$ .

### 3.2.2 Forgetting of Initial Distribution for the Conditional $L$ -likelihood

For all integer  $n > 0$ ,  $z_{0:n-1} \in Z^n$ ,  $\chi, \chi' \in \mathcal{P}(X, \mathcal{X})$  and nonnegative measurable functions  $f$  and  $h$  on  $(X, \mathcal{X})$ , define

$$\begin{aligned} & \Delta_{\chi, \chi'}^\theta \langle z_{0:n-1} \rangle (f, h) \\ & := (\chi \mathbf{L}^\theta \langle z_{0:n-1} \rangle f)(\chi' \mathbf{L}^\theta \langle z_{0:n-1} \rangle h) - (\chi \mathbf{L}^\theta \langle z_{0:n-1} \rangle h)(\chi' \mathbf{L}^\theta \langle z_{0:n-1} \rangle f). \end{aligned} \quad (3.13)$$

Let  $\bar{X} := X \times X$ ,  $\bar{\mathcal{X}} := \mathcal{X} \otimes \mathcal{X}$  and denote  $\bar{L}$  kernel on  $(\bar{X}, \bar{\mathcal{X}})$  such that for all  $(x, x') \in \bar{X}$ ,  $A \times A' \in \bar{\mathcal{X}}$ ,

$$\bar{L}((x, x'); A \times A') = L(x; A)L(x'; A'),$$

where  $L$  is a kernel on  $\mathbf{X} \times \mathcal{X}$ . For  $\chi$  and  $\chi'$  two probability measures on  $(\mathbf{X}, \mathcal{X})$  and  $f, h$  real-valued measurable functions on  $(\mathbf{X}, \mathcal{X})$ , define for  $\bar{A} \in \bar{\mathbf{X}}$  and  $\bar{x} = (x, x') \in \bar{\mathbf{X}}$ ,

$$(\chi \otimes \chi')(\bar{A}) := \int \chi(dx) \chi'(dx') \mathbb{1}_{\bar{A}}(x, x') \quad \text{and} \quad (f \otimes h)(\bar{x}) := f(x)h(x').$$

With above notation we can rewrite (3.13) as:

$$\begin{aligned} \Delta_{\chi, \chi'}^{\theta} \langle z_{0:n-1} \rangle (f, h) \\ = \int (\chi \otimes \chi')(d\bar{x}_0) \left( \prod_{i=0}^{n-1} \bar{\mathbf{L}}^{\theta} \langle z_i \rangle (\bar{x}_i; d\bar{x}_{i+1}) \right) (f \otimes h - h \otimes f)(\bar{x}_n). \end{aligned}$$

For  $a \in \mathbb{R}$ , let  $[a]$  denote the integer part of  $a$  and for any real-valued function  $f$  on  $\mathbf{X}$ , let  $|f|_{\infty} := \sup\{|f(x)| : x \in \mathbf{X}\}$ . Following Douc and Moulines [2012], we have the following results.

**Proposition 3.2.10.** *Assume that (L-1) and (L-2) hold. Let  $0 \leq \gamma^- < \gamma^+ \leq 1$ . Then for any  $\eta > 0$ , there exists  $\rho \in (0, 1)$  such that for all positive integer  $n$  and  $z_{0:n-1} \in \mathbf{Z}^n$  satisfying*

$$n^{-1} \sum_{i=0}^{n-1} \mathbb{1}_{\mathbf{K}}(z_i) \geq \max(1 - \gamma^-, (1 + \gamma^+)/2), \quad (3.14)$$

for any  $\beta \in (\gamma^-, \gamma^+)$ , nonnegative bounded functions  $f$  and  $h$ , probability measures  $\chi$  and  $\chi'$  on  $(\mathbf{X}, \mathcal{X})$  and  $\theta \in \Theta$ ,

$$\begin{aligned} & |\Delta_{\chi, \chi'}^{\theta} \langle z_{0:n-1} \rangle (f, h)| \\ & \leq \rho^{\lfloor n(\beta - \gamma^-) \rfloor} [(\chi \mathbf{L}^{\theta} \langle z_{0:n-1} \rangle f)(\chi' \mathbf{L}^{\theta} \langle z_{0:n-1} \rangle h) + (\chi \mathbf{L}^{\theta} \langle z_{0:n-1} \rangle h)(\chi' \mathbf{L}^{\theta} \langle z_{0:n-1} \rangle f)] \\ & \quad + 2\eta^{\lfloor n(\gamma^+ - \beta) \rfloor / 2} \left[ \prod_{i=0}^{n-1} |\mathbf{L}^{\theta} \langle z_i \rangle (\cdot; \mathbf{X})|_{\infty}^2 \right] |f|_{\infty} |h|_{\infty}. \end{aligned} \quad (3.15)$$

*Proof.* The proof is identical to [Douc and Moulines, 2012, Proposition 5] and is thus omitted.  $\square$

**Proposition 3.2.11.** *Assume that (L-1)–(L-4) hold. Then there exist a constant  $\tau \in (0, 1)$  and a  $\mathbb{P}_{\mathbf{Z}}$ -a.s. finite integer-valued random variable  $T$  such that for all  $\chi, \chi' \in \mathcal{M}^+(\mathbf{D})$  and bounded measurable function  $f \geq 1$  on  $(\mathbf{X}, \mathcal{X})$ ,  $\mathbb{P}_{\mathbf{Z}}$ -a.s.,*

$$\sup_{\theta \in \Theta} \sup_{k \geq T} \sup_{m \geq 0} \tau^{-(m+k)} |\ln \chi \mathbf{L}^{\theta} \langle Z_k | Z_{-m:k-1} \rangle f - \ln \chi' \mathbf{L}^{\theta} \langle Z_k | Z_{-m:k-1} \rangle f| < \infty, \quad (3.16)$$

$$\sup_{\theta \in \Theta} \sup_{k \geq T} \sup_{m \geq 0} \tau^{-(m+k)} \left| \ln \chi \mathbf{L}^\theta \langle Z_k | Z_{-m:k-1} \rangle f - \ln \chi \mathbf{L}^\theta \langle Z_k | Z_{-m-1:k-1} \rangle f \right| < \infty, \quad (3.17)$$

$$\sup_{\theta \in \Theta} \sup_{m \geq 0} \tau^{-m} \left| \ln \chi \mathbf{L}^\theta \langle Z_0 | Z_{-m:-1} \rangle f - \ln \chi \mathbf{L}^\theta \langle Z_0 | Z_{-m-1:-1} \rangle f \right| < \infty. \quad (3.18)$$

*Proof.* See [Section 3.4.1](#). □

As a consequence of [Proposition 3.2.11](#), we have the following result.

**Corollary 3.2.12.** *Assume that (L-1)–(L-4) hold. Then for all  $\theta \in \Theta$ , there exists a nonnegative measure kernel  $\lambda_Z^\theta$  on  $\mathbb{Z}^{\mathbb{Z}^-} \times \mathcal{X}$  such that for any probability measure  $\chi \in \mathcal{M}^+(D)$  and nonnegative and bounded measurable function  $f$  on  $(\mathcal{X}, \mathcal{X})$ ,*

$$\lim_{m \rightarrow \infty} \chi \mathbf{L}^\theta \langle Z_0 | Z_{-m:-1} \rangle f = \lambda_Z^\theta \langle Z_{-\infty:0} \rangle f, \quad \mathbb{P}_Z\text{-a.s.} \quad (3.19)$$

*Proof.* Now let  $\theta \in \Theta$ . Results (3.16) and (3.18) of [Proposition 3.2.11](#) ensure that for each bounded measurable function  $f \geq 1$  and for all  $\chi \in \mathcal{M}^+(D)$ , there exists a nonnegative measurable function  $z_{-\infty:0} \mapsto \lambda_Z^\theta \langle z_{-\infty:0} \rangle f$  on  $(\mathbb{Z}^{\mathbb{Z}^-}, \mathcal{Z}^{\otimes \mathbb{Z}^-})$  satisfying (3.19). This result propagates to any nonnegative and bounded measurable function  $h$  on  $(\mathcal{X}, \mathcal{X})$ , by setting  $f = h + 1$  and noting that (3.19) also holds when  $f = \mathbb{1}_{\mathcal{X}}$ . Observe that each limit  $\lambda_Z^\theta$  does not depend on the initial probability distribution  $\chi \in \mathcal{M}^+(D)$ . Furthermore, since  $\{\chi \mathbf{L}^\theta \langle Z_0 | Z_{-m:-1} \rangle : m \in \mathbb{Z}_+\}$  is a sequence of finite measures on  $(\mathcal{X}, \mathcal{X})$  and (3.19) holds for all nonnegative and bounded measurable function  $f$ , then by Vitali-Hahn-Saks theorem (see [Brooks \[1969\]](#)),  $\lambda_Z^\theta \langle Z_{-\infty:0} \rangle$  defines a (nonnegative) measure on  $(\mathcal{X}, \mathcal{X})$ . □

In what follows, for notational convenience, the measure  $\lambda_Z^\theta \langle Z_{-\infty:0} \rangle$  will be denoted by  $\mathbf{L}^\theta \langle Z_0 | Z_{-\infty:-1} \rangle$ , and for all integer  $n > 0$ , the successive product  $\prod_{k=0}^{n-1} \lambda_Z^\theta \langle Z_{-\infty:k} \rangle \mathbb{1}_{\mathcal{X}}$  will be denoted by  $\mathbf{L}^\theta \langle Z_{0:n-1} | Z_{-\infty:-1} \rangle \mathbb{1}_{\mathcal{X}}$ .

### 3.2.3 Convergence of the Maximum $L$ -likelihood Estimator

We now state the convergence result of the maximum  $L$ -likelihood estimator as follows.

**Proposition 3.2.13.** *Assume that (L-1)–(L-4) hold. Then*

(i) *For all  $\theta \in \Theta$ ,*

$$\mathbb{E}_Z \left[ \left| \ln \mathbf{L}^\theta \langle Z_0 | Z_{-\infty:-1} \rangle \mathbb{1}_{\mathcal{X}} \right| \right] < \infty. \quad (3.20)$$

(ii) For all  $\chi \in \mathcal{M}^+(D)$ ,  $\mathbb{P}_Z$ -a.s.,

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in \Theta} n^{-1} \left| \ln \chi \mathbf{L}^\theta \langle Z_{0:n-1} \rangle \mathbb{1}_X - \ln \mathbf{L}^\theta \langle Z_{0:n-1} | Z_{-\infty:-1} \rangle \mathbb{1}_X \right| = 0.$$

(iii) Assume moreover that (L-5) holds. Then the function

$$\theta \mapsto \mathbb{E}_Z \left[ \ln \mathbf{L}^\theta \langle Z_0 | Z_{-\infty:-1} \rangle \mathbb{1}_X \right]$$

is  $\mathbb{P}_Z$ -a.s. upper semi-continuous.

*Proof.* For convenience, the proof is postponed to [Section 3.4.2](#).  $\square$

**Theorem 3.2.14.** Assume that (L-1)–(L-5) hold. Assume in addition that for all  $\chi \in \mathcal{M}^+(D)$ ,  $\mathbb{P}_Z$ -a.s.,

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in \Theta} n^{-1} \left| \ln \chi \check{\mathbf{L}}^\theta \langle Z_{0:n-1} \rangle \mathbb{1}_X - \ln \chi \mathbf{L}^\theta \langle Z_{0:n-1} \rangle \mathbb{1}_X \right| = 0,$$

where for all  $n \in \mathbb{Z}_+^*$ ,  $z_{0:n-1} \in \mathcal{Z}^n$  and  $\theta \in \Theta$ ,  $\check{\mathbf{L}}^\theta \langle z_{0:n-1} \rangle$  is a nonnegative kernel on  $\mathcal{X} \times \mathcal{X}$  and such that  $\theta \mapsto \chi \check{\mathbf{L}}^\theta \langle z_{0:n-1} \rangle \mathbb{1}_X$  is continuous. Then for all  $\chi \in \mathcal{M}^+(D)$ ,

$$\lim_{n \rightarrow \infty} \Delta(\check{\theta}_{\chi,n}^L, \Theta_\star^L) = 0, \quad \mathbb{P}_Z\text{-a.s.}, \quad (3.21)$$

where

$$\check{\theta}_{\chi,n}^L \in \operatorname{argmax}_{\theta \in \Theta} n^{-1} \ln \chi \check{\mathbf{L}}^\theta \langle Z_{0:n-1} \rangle \mathbb{1}_X \quad (3.22)$$

and

$$\Theta_\star^L := \operatorname{argmax}_{\theta \in \Theta} \mathbb{E}_Z \left[ \ln \mathbf{L}^\theta \langle Z_0 | Z_{-\infty:-1} \rangle \mathbb{1}_X \right]. \quad (3.23)$$

*Proof.* First note that from the compactness of  $\Theta$  and the semi-continuity on  $\Theta$  of the objective functions in (3.22) and (3.23), the corresponding argmax's are well defined. The rest of the proof directly follows from triangular inequality, from [Proposition 3.2.13](#) and from [Pfanzagl \[1969\]](#).  $\square$

Result below is a direct implication of [Theorem 3.2.14](#).

**Corollary 3.2.15.** Assume that (L-1)–(L-5) hold. Then for all  $\chi \in \mathcal{M}^+(D)$ ,  $\mathbb{P}_Z$ -a.s.,

$$\lim_{n \rightarrow \infty} \Delta(\hat{\theta}_{\chi,n}^L, \Theta_\star^L) = 0, \quad \mathbb{P}_Z\text{-a.s.},$$

where  $\hat{\theta}_{\chi,n}^L$  is the maximum  $L$ -likelihood estimator defined in (3.4).

### 3.3 Consistency of the MLE for Fully Dominated Markov Models

In this section, we treat the consistency property of the MLE for a large subclass of PMMs discussed in [Chapter 2](#), namely the class of fully-dominated PMMs, under the framework of the well-specified models. This class contains many models of interest, for instance HMMs, Markov switching models and many Markov chain in random environment models, whose distributional laws are dominated by some fixed  $\sigma$ -finite measures. The main result of this present section is obtained as a combination of convergence result of the maximum  $L$ -likelihood estimator derived in [Section 3.2](#) above and the identifiability result obtained in [Chapter 2](#).

For self-contained representation, we should recall some notation and settings already introduced in [Chapter 2](#). Now, let  $(\mathbf{X}, \mathcal{X})$  and  $(\mathbf{Y}, \mathcal{Y})$  be two Borel spaces, that is, measurable spaces that are isomorphic to a Borel subset of  $[0, 1]$ . Let  $\Theta$  be a set of parameter, and as in [Section 3.2](#), this set is assumed to be a compact metric space equipped with a metric  $\Delta$ . Consider a statistical model determined by a class of Markov kernels  $\{K^\theta : \theta \in \Theta\}$  on  $(\mathbf{X} \times \mathbf{Y}) \times (\mathcal{X} \otimes \mathcal{Y})$ . We denote by  $\mathbb{P}_\xi^\theta$  the probability (and by  $\mathbb{E}_\xi^\theta$  the corresponding expectation) induced on  $(\mathbf{X} \times \mathbf{Y})^{\mathbb{Z}_+}$  by a Markov chain  $\{(X_k, Y_k) : k \in \mathbb{Z}_+\}$  with transition kernel  $K^\theta$  and initial distribution  $\xi$  on  $\mathbf{X} \times \mathbf{Y}$ . In the case where  $\xi$  is a Dirac mass at  $(x, y)$  we will simply write  $\mathbb{P}_{(x,y)}^\theta$ . For partially-observed Markov chains, that is, when only a sample  $Y_{1:n} = (Y_1, \dots, Y_n) \in \mathbf{Y}^n$  of the second component is realized, the transition kernel  $K^\theta$  can be conveniently written as

$$K^\theta((x, y); dx'dy') = Q^\theta((x, y); dx')G^\theta((x, y, x'); dy'), \quad (3.24)$$

where  $Q^\theta$  and  $G^\theta$  are probability kernels on  $(\mathbf{X} \times \mathbf{Y}) \times \mathcal{X}$  and  $(\mathbf{X} \times \mathbf{Y} \times \mathbf{X}) \times \mathcal{Y}$ , respectively. In this section, we investigate a particular case where both  $Q^\theta$  and  $G^\theta$  are absolutely continuous with respect to some fixed nonnegative  $\sigma$ -finite measures, namely the case where the model is fully-dominated.

**Definition 3.3.1.** We say that the Markov model  $\{K^\theta : \theta \in \Theta\}$  of the form [\(3.24\)](#) is a fully-dominated and partially observed Markov model if there exist some nonnegative  $\sigma$ -finite measures  $\mu$  on  $\mathbf{X}$  and  $\nu$  on  $\mathbf{Y}$  such that for all  $(x, y), (x', y') \in \mathbf{X} \times \mathbf{Y}$ ,

$$Q^\theta((x, y); dx') = q^\theta((x, y); x') \mu(dx') \quad (3.25)$$

and

$$G^\theta((x, y, x'); dy') = g^\theta((x, y, x'); y') \nu(dy'), \quad (3.26)$$

respectively, where the conditional density functions  $q^\theta$  and  $g^\theta$  in addition satisfy

$$q^\theta((x, y); x') > 0, \quad (3.27)$$

$$g^\theta((x, y, x'); y') > 0, \quad (3.28)$$

for all  $(x, y), (x', y') \in \mathbf{X} \times \mathbf{Y}$ .

It follows from (3.25) and (3.26) that for all  $(x, y) \in \mathbf{X} \times \mathbf{Y}$ ,  $A \in \mathcal{X}$  and  $B \in \mathcal{Y}$ ,

$$K^\theta((x, y); A \times B) = \int_B \kappa^\theta\langle y, y' \rangle(x; A) \nu(dy'),$$

where, for all  $y, y' \in \mathbf{Y}$ ,  $\kappa^\theta\langle y, y' \rangle$  is a kernel defined on  $(\mathbf{X}, \mathcal{X})$  by

$$\kappa^\theta\langle y, y' \rangle(x; dx') := q^\theta((x, y); x') g^\theta((x, y, x'); y') \mu(dx'). \quad (3.29)$$

Note that in the partially dominated setting we do not require (3.25) and (3.27). The fully dominated PMM can be graphically represented as in Figure 3.4.

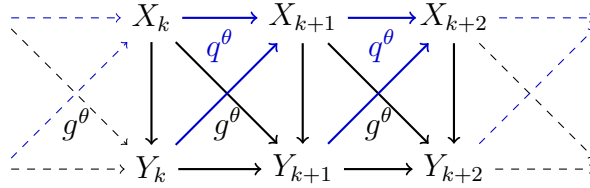


Figure 3.4 – Graphical representation of the fully dominated and partially observed bivariate Markov chain.

In what follows, we will restrict our consideration on the case where  $\mathbf{X}$  is a complete and separable metric space equipped with a metric  $d_{\mathbf{X}}$  and  $\mathcal{X}$  is the associated Borel  $\sigma$ -field.

**Remark 3.3.2.** In this setting, for all  $\theta \in \Theta$  and positive integer  $s \geq 1$ , the kernel  $(K^\theta)^s$  is  $\varphi$ -irreducible with  $\varphi = \mu \otimes \nu$ .

In well-specified models, it is assumed that the observations  $Y_{1:n}$  are generated from a process  $\{(X_k, Y_k) : k \in \mathbb{Z}_+\}$ , which follows the distribution  $\mathbb{P}_{\xi_\star}^{\theta_\star}$  associated to an unknown parameter  $\theta_\star \in \Theta$  and an unknown initial distribution  $\xi_\star$  (usually,  $\xi_\star$  is such that, under  $\mathbb{P}_{\xi_\star}^{\theta_\star}$ ,  $\{Y_k : k \in \mathbb{Z}_+\}$  is a stationary



sequence). To form a consistent estimate of  $\theta_*$  on the basis of the observations  $Y_{1:n}$  only, i.e., without access to the hidden process  $\{X_k : k \in \mathbb{Z}_+\}$ , we define the maximum likelihood estimator (MLE)  $\hat{\theta}_{\xi,n}$  by

$$\hat{\theta}_{\xi,n} \in \operatorname{argmax}_{\theta \in \Theta} n^{-1} \ln p_{\xi}^{\theta}(Y_{1:n}), \quad (3.30)$$

where  $p_{\xi}^{\theta}(y_{1:n})$  is the (conditional) likelihood function of the observations  $y_{1:n}$  under parameter  $\theta \in \Theta$  with some arbitrary initial distribution  $\xi$  on  $\mathbf{X} \times \mathbf{Y}$ , that is,

$$\begin{aligned} p_{\xi}^{\theta}(y_{1:n}) &:= \int \prod_{i=0}^{n-1} Q^{\theta}((x_i, y_i); dx_{i+1}) g^{\theta}((x_i, y_i, x_{i+1}); y_{i+1}) \xi(dx_0 dy_0) \\ &= \int \kappa^{\theta}\langle y_0, y_1 \rangle \kappa^{\theta}\langle y_1, y_2 \rangle \cdots \kappa^{\theta}\langle y_{n-1}, y_n \rangle(x_0; \mathbf{X}) \xi(dx_0 dy_0). \end{aligned} \quad (3.31)$$

This corresponds to the conditional density of  $Y_{1:n}$  given  $(X_0, Y_0)$  with the latter integrated according to initial probability distribution  $\xi$ . In practice  $\xi$  is often taken to be a Dirac mass at  $(x, y)$  with  $x$  arbitrarily chosen and  $y$  equal to the observation  $Y_0$  when it is available. Yet for sake of generality, in our study, the initial probability  $\xi$  will be the one of the form  $\chi \otimes \delta_{Y_0}$ , where  $\chi$  is some probability distribution on  $(\mathbf{X}, \mathcal{X})$  and  $\delta_{Y_0}$  is the Dirac mass at the observation  $Y_0$ . We then refer to  $p_{\chi \otimes \delta_{Y_0}}^{\theta}(Y_{1:n})$  as the conditional likelihood of the observations  $Y_{0:n}$  associated to the probability measure  $\chi$ .

Condition (3.28) guarantees that for all  $\theta \in \Theta$ ,  $n \in \mathbb{Z}_+^*$ ,  $y_{0:n} \in \mathbf{Y}^{n+1}$  and  $\chi \in \mathcal{P}(\mathbf{X}, \mathcal{X})$ ,  $p_{\chi \otimes \delta_{y_0}}^{\theta}(y_{1:n})$  given in (3.31) is positive. Now, define

$$p_{\chi \otimes \delta_{y_0}}^{\theta}(y_{m:n} | y_{1:m-1}) := \frac{p_{\chi \otimes \delta_{y_0}}^{\theta}(y_{1:n})}{p_{\chi \otimes \delta_{y_0}}^{\theta}(y_{1:m-1})}, \quad 1 < m \leq n. \quad (3.32)$$

The term in left-hand side of (3.32) can be interpreted as the conditional density of the the observations  $y_{m:n}$  given  $Y_{0:m-1} = y_{0:m-1}$  and  $X_0 \sim \chi$ . For all integer  $s \geq 1$ ,  $y_{0:s} \in \mathbf{Y}^{s+1}$  and  $\theta \in \Theta$ , we define the (possibly unnormalized) kernel  $\ell^{\theta}\langle y_{0:s} \rangle$  on  $\mathbf{X} \times \mathcal{X}$  by, for all  $(x_0, A) \in \mathbf{X} \times \mathcal{X}$ ,

$$\ell^{\theta}\langle y_{0:s} \rangle(x_0; A) := \int_{\mathbf{X}^s} \mathbb{1}_A(x_s) \prod_{i=0}^{s-1} Q^{\theta}((x_i, y_i); dx_{i+1}) g^{\theta}((x_i, y_i, x_{i+1}); y_{i+1}). \quad (3.33)$$

Note that for all  $\theta \in \Theta$ , the kernel  $\ell^{\theta}$  is nonnegative and from (3.31), for all integer  $s \geq 1$ ,  $y_{0:s} \in \mathbf{Y}^{s+1}$  and  $\chi \in \mathcal{P}(\mathbf{X}, \mathcal{X})$ , we have

$$\chi \ell^{\theta}\langle y_{0:s} \rangle \mathbb{1}_{\mathbf{X}} = p_{\chi \otimes \delta_{y_0}}^{\theta}(y_{1:s}) > 0. \quad (3.34)$$

Analogously to (3.32), we moreover define, for all  $\theta \in \Theta$ ,  $n \in \mathbb{Z}_+^*$ ,  $y_{0:n} \in \mathbf{Y}^{n+1}$ ,  $\chi \in \mathcal{P}(\mathbf{X}, \mathcal{X})$  and real-valued measurable function  $f$  on  $(\mathbf{X}, \mathcal{X})$ ,

$$\chi^{\ell^\theta} \langle y_{m:n} | y_{0:m-1} \rangle f := \frac{\chi^{\ell^\theta} \langle y_{0:n} \rangle f}{\chi^{\ell^\theta} \langle y_{0:m-1} \rangle \mathbb{1}_{\mathbf{X}}}, \quad 1 < m \leq n. \quad (3.35)$$

To complete our setting, the following assumption is needed and is supposed to hold throughout.

(F-1) For all  $\theta \in \Theta$ , the transition kernel  $K^\theta$  admits a unique invariant probability distribution  $\pi^\theta$ .

**Remark 3.3.3.** Assumption (F-1) can be equivalently stated as: for all positive integer  $r$  and  $\theta \in \Theta$ , the transition kernel  $(K^\theta)^r$  admits a unique invariant probability distribution. Moreover, the invariant probability distributions for  $K^\theta$  and  $(K^\theta)^r$  are the same. To see this, first suppose that (F-1) holds. It then implies that for any positive integer  $r$ ,  $\pi^\theta$  is an invariant distribution for  $(K^\theta)^r$ . By Remark 3.3.2, the skeleton  $(K^\theta)^r$  is also  $\varphi$ -irreducible and by [Meyn and Tweedie, 2009, Theorem 10.0.1 and Proposition 10.1.1], the skeleton  $(K^\theta)^r$  admits a unique invariant  $\pi^\theta$ . To see the converse, note that for each  $r$ , if  $\tilde{\pi}^\theta$  is the invariant probability distribution for  $(K^\theta)^r$ , then  $\pi^\theta = r^{-1} \sum_{s=0}^{r-1} \tilde{\pi}^\theta (K^\theta)^s$  defines an invariant for  $K^\theta$ . The uniqueness of  $\pi^\theta$  for  $K^\theta$  follows from the uniqueness of  $\tilde{\pi}^\theta$  for  $(K^\theta)^r$ .

Finally, let us recall (from Chapter 2) the following notation and definition regarding stationary distributions.

**Definition 3.3.4.** Under Assumption (F-1), we denote by  $\pi_1^\theta$  and  $\pi_2^\theta$  the marginal distributions of  $\pi^\theta$  on  $\mathbf{X}$  and  $\mathbf{Y}$ , respectively, and by  $\mathbb{P}^\theta$  and  $\tilde{\mathbb{P}}^\theta$  the probability distributions defined respectively as follows.

- a)  $\mathbb{P}^\theta$  denotes the extension of  $\mathbb{P}_{\pi^\theta}^\theta$  on the whole line  $(\mathbf{X} \times \mathbf{Y})^{\mathbb{Z}}$ .
- b)  $\tilde{\mathbb{P}}^\theta$  is the corresponding projection on the component  $\mathbf{Y}^{\mathbb{Z}}$ .

We also use the symbols  $\mathbb{E}^\theta$  and  $\tilde{\mathbb{E}}^\theta$  to denote the expectations corresponding to  $\mathbb{P}^\theta$  and  $\tilde{\mathbb{P}}^\theta$ , respectively. Moreover, for all  $\theta, \theta' \in \Theta$ , we write  $\theta \sim \theta'$  if and only if  $\tilde{\mathbb{P}}^\theta = \tilde{\mathbb{P}}^{\theta'}$ . This defines an equivalence relation on the parameter set  $\Theta$  and the corresponding equivalence class of  $\theta$  is denoted by  $[\theta] := \{\theta' \in \Theta : \theta \sim \theta'\}$ .

The equivalence relationship  $\sim$  was introduced by Leroux [1992] as an alternative to the classical identifiability condition.

Assumption (F-1) directly implies that the process  $\{Y_k : k \in \mathbb{Z}\}$  is stationary and ergodic; however, whether it also implies that  $\{Y_{k:k+r} : k \in \mathbb{Z}\}$ ,

$r \in \mathbb{Z}_+^*$ , is a stationary and ergodic process is not obvious though this may be the case. Indeed, this result can be induced from the following construction.

Let for all integer  $k$ ,  $W_k := (X_k, Y_k)$  and let  $\mathbf{W} := \mathbf{X} \times \mathbf{Y}$  and  $\mathcal{W} := \mathcal{X} \otimes \mathcal{Y}$ . Now consider a stochastic process  $\{\bar{W}_k : k \in \mathbb{Z}_+\}$  on  $(\mathbf{W}^{r+1}, \mathcal{W}^{\otimes(r+1)})$  with  $\bar{W}_k = W_{kr:(k+1)r}$ . Obviously,  $\{\bar{W}_k : k \in \mathbb{Z}_+\}$  is a Markov chain on  $(\mathbf{W}^{r+1}, \mathcal{W}^{\otimes(r+1)})$  with transition kernel denoted by  $\bar{K}^\theta$ . The Markov kernel  $\bar{K}^\theta$  can be explicitly expressed as, for all  $\bar{w} = w_{0:r} \in \mathbf{W}^{r+1}$  and  $A_0, \dots, A_r \in \mathcal{W}$ ,

$$\begin{aligned} \bar{K}^\theta(\bar{w}; A_0 \times \dots \times A_r) &= \int_{A_0 \times \dots \times A_r} \delta_{w_r}(dw'_0) K^\theta(w'_0; dw'_1) \dots K^\theta(w'_{r-1}; dw'_r) \\ &= \int_{A_1 \times \dots \times A_r} \mathbb{1}_{A_0}(w_r) T^\theta(w_r; dw'_{1:r}) \\ &= \mathbb{1}_{A_0}(w_r) T^\theta(w_r; A_1 \times \dots \times A_r), \end{aligned} \quad (3.36)$$

where  $T^\theta$  is a kernel on  $\mathbf{W} \times \mathcal{W}^{\otimes r}$  defined by, for all  $w_0 \in \mathbf{W}$  and  $A_1, \dots, A_r \in \mathcal{W}$ ,

$$T^\theta(w_0; A_1 \times \dots \times A_r) = \int_{A_1 \times \dots \times A_r} K^\theta(w_0; dw_1) \dots K^\theta(w_{r-1}; dw_r). \quad (3.37)$$

**Lemma 3.3.5** stated below can be instrumental for obtaining stationary and ergodic solutions for the sequence  $\{Y_{k:k+r} : k \in \mathbb{Z}\}$ . Remarkably, this result does not rely on the  $\varphi$ -irreducibility assumption.

**Lemma 3.3.5.** *The kernel  $\bar{K}^\theta$  admits a (unique) invariant distribution on  $(\mathbf{W}^{r+1}, \mathcal{W}^{\otimes(r+1)})$  if and only if the kernel  $(K^\theta)^r$  admits a (unique) invariant distribution on  $(\mathbf{W}, \mathcal{W})$ .*

*Proof.* The proof is postponed to [Section 3.4.3](#) for convenience.  $\square$

Let us now state the additional assumptions yielding the consistency of the MLE defined by [\(3.30\)](#).

(F-2) Assumption [\(F-1\)](#) holds. Moreover, for all  $\theta_\star \in \Theta$ , there exist a nonnegative integer  $r \geq 1$  and a set  $\mathbf{K} \in \mathcal{Y}^{\otimes(r+1)}$  such that, denoting  $\mathbf{Z} := \mathbf{Y}^{r+1}$ , and for  $k \in \mathbb{Z}$ ,

$$z_i := y_{kr:(k+1)r} \in \mathbf{Z}, \quad Z_k := Y_{kr:(k+1)r} \in \mathbf{Z},$$

each of the following holds.

- (i)  $\tilde{\mathbb{P}}^{\theta_\star}(Z_0 \in \mathbf{K}) > 2/3$ .

- (ii) For all  $\eta > 0$  there exists a  $L$ -local Doeblin set  $C \in \mathcal{X}$  with respect to the family  $\{\ell^\theta \langle z \rangle\}_{(\theta, z) \in \Theta \times Z}$  (defined in [Definition 3.2.7](#) with  $\ell^\theta$  defined by [\(3.33\)](#)) such that for all  $\theta \in \Theta$  and  $z \in K$ ,

$$\sup_{x \in C^c} \ell^\theta \langle z \rangle(x; \mathbf{X}) \leq \eta \sup_{x \in X} \ell^\theta \langle z \rangle(x; \mathbf{X}) < \infty \quad (3.38)$$

and

$$\inf_{z \in K} \frac{\epsilon_C^- \langle z \rangle}{\epsilon_C^+ \langle z \rangle} > 0, \quad (3.39)$$

where the functions  $\epsilon_C^+$  and  $\epsilon_C^-$  are given in [Definition 3.2.7](#).

- (iii) There exists a set  $D \in \mathcal{X}$  such that

$$\tilde{\mathbb{E}}^{\theta_*} \left[ \ln^- \inf_{\theta \in \Theta} \inf_{x \in D} \ell^\theta \langle Z_0 \rangle(x; D) \right] < \infty. \quad (3.40)$$

(F-3) We have,

$$\tilde{\mathbb{E}}^{\theta_*} \left[ \ln^+ \sup_{\theta \in \Theta} \sup_{(x, x') \in X^2} g^\theta((x, Y_0, x'); Y_1) \right] < \infty. \quad (3.41)$$

(F-4) For all  $\theta_* \in \Theta$ , there exists a positive integer  $p$  such that for all  $n \geq p$  and  $x \in X$ , the function  $\theta \mapsto \ell^\theta \langle Y_{0:n} \rangle(x; \mathbf{X})$  is  $\tilde{\mathbb{P}}^{\theta_*}$ -a.s. continuous on  $\Theta$ .

**Remark 3.3.6.** Note that these assumptions can be used to derive the convergence of the MLE under the misspecified settings, except that in [\(F-2\)](#), rather than assuming that Assumption [\(F-1\)](#) holds, we should instead assume, with the same  $r$ , that the observed process  $\{Y_{k:k+r} : k \in \mathbb{Z}\}$  is stationary and ergodic. This in turn extends the result of [Douc and Moulines \[2012\]](#) beyond the HMMs.

**Remark 3.3.7.** To obtain [\(F-2\)\(iii\)](#), one may need to check that

- (i)  $\tilde{\mathbb{E}}^{\theta_*} \left[ \ln^- \inf_{\theta \in \Theta} \inf_{x \in D} Q^\theta((x, Y_0); D) \right] < \infty$  and
- (ii)  $\tilde{\mathbb{E}}^{\theta_*} \left[ \ln^- \inf_{\theta \in \Theta} \inf_{(x, x') \in D^2} g^\theta((x, Y_0, x'); Y_1) \right] < \infty$ .

**Remark 3.3.8.** Assumption [\(F-4\)](#) is usually obtained as a consequence of the continuity of  $\theta \mapsto q^\theta$  and  $\theta \mapsto g^\theta$  via classical techniques to deal with integrals involving some parameters.

For the integer  $r$  and the set  $D \in \mathcal{X}$  defined in [\(F-2\)](#), let  $\mathcal{M}^{\theta_*}(D, r)$  be the subset of  $\mathcal{P}(X, \mathcal{X})$ , the set of all probability measures on  $(X, \mathcal{X})$ , satisfying

$$\mathcal{M}^{\theta_*}(D, r) = \left\{ \chi \in \mathcal{P}(X, \mathcal{X}) : \tilde{\mathbb{E}}^{\theta_*} \left[ \ln^- \inf_{\theta \in \Theta} \chi \ell^\theta \langle Y_{0:s} \rangle \mathbb{1}_D \right] < \infty, \text{ for all } s \in \{1, \dots, r\} \right\}. \quad (3.42)$$

Now we have all the necessary tools for deriving the results pertaining to the consistency of the MLE for fully dominated PMMs. These results are stated as follows.

**Proposition 3.3.9.** *Assume (F-1)–(F-3). Then*

- (i) *For all  $\theta \in \Theta$  and  $\chi \in \mathcal{M}^{\theta^*}(D, r)$ , there exists a kernel  $\ell^{\theta, \theta^*} \langle \cdot | \cdot \rangle$  on  $Y^{\mathbb{Z}^-} \times \mathcal{X}$  such that for all nonnegative and bounded measurable function  $f$  on  $(\mathcal{X}, \mathcal{X})$ ,*

$$\tilde{\mathbb{P}}^{\theta^*} \left[ \lim_{m \rightarrow \infty} \chi \ell^{\theta} \langle Y_1 | Y_{-m:0} \rangle f = \ell^{\theta, \theta^*} \langle Y_1 | Y_{-\infty:0} \rangle f \right] = 1. \quad (3.43)$$

Moreover,

$$\tilde{\mathbb{E}}^{\theta^*} [ |\ln \ell^{\theta, \theta^*} \langle Y_1 | Y_{-\infty:0} \rangle f| ] < \infty.$$

- (ii) *For all  $\theta \in \Theta$ ,  $\chi \in \mathcal{M}^{\theta^*}(D, r)$  and nonnegative and bounded measurable function  $f$  on  $(\mathcal{X}, \mathcal{X})$ ,*

$$\lim_{n \rightarrow \infty} n^{-1} \ln \chi \ell^{\theta} \langle Y_{0:n} \rangle f = \tilde{\mathbb{E}}^{\theta^*} [ \ln \ell^{\theta, \theta^*} \langle Y_1 | Y_{-\infty:0} \rangle f ], \quad \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.}$$

*Proof.* For convenience, the proof is postponed to [Section 3.4.4](#).  $\square$

As a consequence of [Proposition 3.3.9](#), we have the following result.

**Proposition 3.3.10.** *Assume (F-1)–(F-3). Then*

- (i) *For all  $\theta \in \Theta$  and  $\chi \in \mathcal{M}^{\theta^*}(D, r)$ ,*

$$\tilde{\mathbb{P}}^{\theta^*} \left( \lim_{m \rightarrow \infty} p_{\chi \otimes \delta_{Y_{-m}}}^{\theta} (Y_1 | Y_{-m+1:0}) = \ell^{\theta, \theta^*} \langle Y_1 | Y_{-\infty:0} \rangle \mathbb{1}_{\mathcal{X}} \right) = 1.$$

Moreover,

$$\tilde{\mathbb{E}}^{\theta^*} [ |\ln \ell^{\theta, \theta^*} \langle Y_1 | Y_{-\infty:0} \rangle \mathbb{1}_{\mathcal{X}}| ] < \infty.$$

- (ii) *For all  $\theta \in \Theta$  and  $\chi \in \mathcal{M}^{\theta^*}(D, r)$ ,  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.*

$$\lim_{n \rightarrow \infty} n^{-1} \ln p_{\chi \otimes \delta_{Y_0}}^{\theta} (Y_{1:n}) = \tilde{\mathbb{E}}^{\theta^*} [ \ln \ell^{\theta, \theta^*} \langle Y_1 | Y_{-\infty:0} \rangle \mathbb{1}_{\mathcal{X}} ].$$

*Proof.* The proof immediately follows by letting  $f = \mathbb{1}_{\mathcal{X}}$  in [Proposition 3.3.9](#).  $\square$

Before stating the equivalence-class consistency of the MLE, let us recall the kernel  $\Phi_{x,n}^{\theta}$  defined in [Definition 2.2.11](#) which in our case can be defined

as follows. For all  $\theta \in \Theta$ , positive integer  $n$  and  $x \in \mathsf{X}$ , define the kernel  $\Phi_{x,n}^\theta$  on  $\mathsf{Y}^{n+1} \times \mathcal{X}$  by, setting  $x_0 = x$ ,

$$\Phi_{x,n}^\theta(y_{0:n}; A) := \frac{\int_{\mathsf{X}^{n-1} \times A} \prod_{k=0}^{n-1} Q^\theta((x_k, y_k); dx_{k+1}) g^\theta((x_k, y_k, x_{k+1}); y_{k+1})}{\int_{\mathsf{X}^n} \prod_{k=0}^{n-1} Q^\theta((x_k, y_k); dx_{k+1}) g^\theta((x_k, y_k, x_{k+1}); y_{k+1})}.$$

In HMM case, the distribution  $\Phi_{x,n}^\theta(Y_{0:n}; \cdot)$  is usually referred to as the *filter distribution*. The following intermediate result is useful.

**Lemma 3.3.11.** *Assume (F-1)–(F-3). Then,*

(i) *For any  $\theta, \theta_\star \in \Theta$ , there exists a probability kernel  $\Phi^{\theta, \theta_\star}$  on  $\mathsf{Y}^{\mathbb{Z}^-} \times \mathcal{X}$  such that for any  $x \in \mathsf{X}$  such that  $\delta_x \in \mathcal{M}^{\theta_\star}(D, r)$ ,*

$$\tilde{\mathbb{P}}^{\theta_\star} \left( \text{for all bounded } f, \lim_{m \rightarrow \infty} \Phi_{x,m}^\theta(Y_{-m:0}; f) = \Phi^{\theta, \theta_\star}(Y_{-\infty:0}; f) \right) = 1.$$

(ii) *For any  $\theta, \theta_\star \in \Theta$  we have,  $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.,*

$$\ell^{\theta, \theta_\star} \langle Y_1 | Y_{-\infty:0} \rangle \mathbb{1}_\mathsf{X} = \int \Phi^{\theta, \theta_\star}(Y_{-\infty:0}; dx_0) \kappa^\theta \langle Y_0, Y_1 \rangle (x_0; \mathsf{X}). \quad (3.44)$$

*Proof.* Note that for all  $\theta \in \Theta$ ,  $y_{0:n} \in \mathsf{Y}^{n+1}$ ,  $n \in \mathbb{Z}_+^*$ ,  $x \in \mathsf{X}$  and (nonnegative) bounded measurable function  $f$ , we have

$$\Phi_{x,n}^\theta(y_{0:n}; f) = \frac{\delta_x \ell^\theta \langle y_{0:n} \rangle f}{\delta_x \ell^\theta \langle y_{0:n} \rangle \mathbb{1}_\mathsf{X}},$$

and so that

$$\begin{aligned} \Phi_{x,n}^\theta(Y_{-n:0}; f) &= \frac{\delta_x \ell^\theta \langle Y_{-n:0} \rangle f}{\delta_x \ell^\theta \langle Y_{-n:0} \rangle \mathbb{1}_\mathsf{X}} \\ &= \frac{(\delta_x \ell^\theta \langle Y_{-n:0} \rangle f) / (\delta_x \ell^\theta \langle Y_{-n:-1} \rangle \mathbb{1}_\mathsf{X})}{(\delta_x \ell^\theta \langle Y_{-n:0} \rangle \mathbb{1}_\mathsf{X}) / (\delta_x \ell^\theta \langle Y_{-n:-1} \rangle \mathbb{1}_\mathsf{X})} \\ &= \frac{\delta_x \ell^\theta \langle Y_0 | Y_{-n:-1} \rangle f}{\delta_x \ell^\theta \langle Y_0 | Y_{-n:-1} \rangle \mathbb{1}_\mathsf{X}}. \end{aligned}$$

Moreover, if  $x \in \mathsf{X}$  such that  $\delta_x \in \mathcal{M}^{\theta_\star}(D, r)$ , **Proposition 3.3.9** implies that  $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.,

$$\lim_{n \rightarrow \infty} \Phi_{x,n}^\theta(Y_{-n:0}; f) = \frac{\ell^{\theta, \theta_\star} \langle Y_0 | Y_{-\infty:-1} \rangle f}{\ell^{\theta, \theta_\star} \langle Y_0 | Y_{-\infty:-1} \rangle \mathbb{1}_\mathsf{X}} := \Phi^{\theta, \theta_\star}(Y_{-\infty:0}; f).$$

Indeed,  $\Phi^{\theta, \theta_\star}$  is a probability kernel on  $\mathbf{Y}^{\mathbb{Z}^-} \times \mathcal{X}$ . This proves [Lemma 3.3.11–\(i\)](#). To prove [Lemma 3.3.11–\(ii\)](#), observe that  $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.,

$$\begin{aligned}
 \ell^{\theta, \theta_\star} \langle Y_1 | Y_{-\infty:0} \rangle \mathbb{1}_{\mathbf{X}} &= \lim_{n \rightarrow \infty} \delta_x \ell^\theta \langle Y_1 | Y_{-n:0} \rangle \mathbb{1}_{\mathbf{X}} \\
 &= \lim_{n \rightarrow \infty} \frac{\delta_x \ell^\theta \langle Y_{-n:1} \rangle \mathbb{1}_{\mathbf{X}}}{\delta_x \ell^\theta \langle Y_{-n:0} \rangle \mathbb{1}_{\mathbf{X}}} \\
 &= \lim_{n \rightarrow \infty} \int \frac{\delta_x \ell^\theta \langle Y_{-n:0} \rangle (dx_0) \kappa^\theta \langle Y_0, Y_1 \rangle (x_0; \mathbf{X})}{\delta_x \ell^\theta \langle Y_{-n:0} \rangle \mathbb{1}_{\mathbf{X}}} \\
 &= \lim_{n \rightarrow \infty} \int \Phi_{x,n}^\theta (Y_{-n:0}; dx_0) \kappa^\theta \langle Y_0, Y_1 \rangle (x_0; \mathbf{X}) \\
 &= \lim_{n \rightarrow \infty} \Phi_{x,n}^\theta (Y_{-n:0}; \kappa^\theta \langle Y_0, Y_1 \rangle (\cdot; \mathbf{X})) \\
 &= \Phi^{\theta, \theta_\star} (Y_{-\infty:0}; \kappa^\theta \langle Y_0, Y_1 \rangle (\cdot; \mathbf{X})) \\
 &= \int \Phi^{\theta, \theta_\star} (Y_{-\infty:0}; dx_0) \kappa^\theta \langle Y_0, Y_1 \rangle (x_0; \mathbf{X}),
 \end{aligned}$$

where we have used, from [\(F-3\)](#),

$$\left| \kappa^\theta \langle Y_0, Y_1 \rangle (\cdot; \mathbf{X}) \right|_\infty \leq \sup_{\theta \in \Theta} \sup_{(x, x') \in \mathbf{X}^2} g^\theta ((x, Y_0, x'); Y_1) < \infty, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

□

**Remark 3.3.12.** From [\(3.44\)](#) and since  $\int \kappa^\theta \langle y_0, y_1 \rangle (x_0; \mathbf{X}) \nu(dy_1) = 1$  for all  $(x_0, y_0) \in \mathbf{X} \times \mathbf{Y}$ , it follows that, given  $y_{-\infty:0} \in \mathbf{Y}^{\mathbb{Z}^-}$ , for all  $\theta, \theta_\star, y_1 \mapsto \ell^{\theta, \theta_\star} \langle y_1 | y_{-\infty:0} \rangle \mathbb{1}_{\mathbf{X}}$  defines a probability density with respect to the measure  $\nu$  on  $(\mathbf{Y}, \mathcal{Y})$ .

Recall that since  $\mathbf{Y}$  is a Borel space, [[Kallenberg, 2002](#), Theorem 6.3] applies and the conditional distribution of  $Y_{1:n}$  given  $Y_{-\infty:0}$  defines a probability kernel. Since  $\tilde{\mathbb{P}}^\theta (Y_{1:n} \in \cdot)$  is dominated by  $\nu^{\otimes n}$ , this in turns defines a conditional density with respect to  $\nu^{\otimes n}$ , denoted by  $p_n^\theta(\cdot | \cdot)$ , so that for all  $B \in \mathcal{Y}^{\otimes n}$ ,

$$\tilde{\mathbb{P}}^\theta (Y_{1:n} \in B | Y_{-\infty:0}) = \int_B p_n^\theta (y_{1:n} | Y_{-\infty:0}) \nu(dy_1) \cdots \nu(dy_n), \quad \tilde{\mathbb{P}}^\theta\text{-a.s.} \quad (3.45)$$

Our main result on consistency of the MLE for the class of fully dominated PMMs is now stated.

**Theorem 3.3.13.** *Assume [\(F-1\)–\(F-4\)](#). Then for all  $\chi \in \mathcal{M}^{\theta_\star}(D, r)$ , denoting  $\xi = \chi \otimes \delta_{Y_0}$ , we have*

$$\lim_{n \rightarrow \infty} \Delta \left( \hat{\theta}_{\xi, n}, [\theta_\star] \right) = 0, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}, \quad (3.46)$$

where  $\hat{\theta}_{\xi,n}$  is defined by (3.30) and  $[\theta_\star]$  is the equivalence class of  $\theta_\star$  as defined in Definition 3.3.4.

*Proof.* Denote for all  $\theta, \theta_\star \in \Theta$ ,

$$\ell(\theta, \theta_\star) := \tilde{\mathbb{E}}^{\theta_\star} [\ln \ell^{\theta, \theta_\star} \langle Y_1 | Y_{-\infty:0} \rangle \mathbb{1}_X] \quad (3.47)$$

and

$$\Theta_\star := \operatorname{argmax}_{\theta \in \Theta} \ell(\theta, \theta_\star).$$

Next, we will show that

$$\lim_{n \rightarrow \infty} \Delta \left( \hat{\theta}_{\xi,n}, \Theta_\star \right) = 0, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.} \quad (3.48)$$

Observe that we have

$$\begin{aligned} \Theta_\star &= \operatorname{argmax}_{\theta \in \Theta} r \tilde{\mathbb{E}}^{\theta_\star} [\ln \ell^{\theta, \theta_\star} \langle Y_1 | Y_{-\infty:0} \rangle \mathbb{1}_X] \\ &= \operatorname{argmax}_{\theta \in \Theta} \tilde{\mathbb{E}}^{\theta_\star} [\ln \mathbf{L}^{\theta, \theta_\star} \langle Z_0 | Z_{-\infty:-1} \rangle \mathbb{1}_X]. \end{aligned}$$

The convergence (3.48) immediately follows from (3.74) with  $h = \mathbb{1}_X$  and Theorem 3.2.14.

It thus remains to show that  $\Theta_\star = [\theta_\star]$ . To this end, we apply Theorem 2.2.7 in Chapter 2, which corresponds to [Douc et al., 2014, Theorem 3]. To comply with the statement of Theorem 2.2.7, let us define  $p^{\theta, \theta_\star}$  by: for all  $y_{-\infty:0} \in \mathbf{Y}^{\mathbb{Z}^-}$ ,  $p^{\theta, \theta_\star}(y_1 | y_{-\infty:0}) := \ell^{\theta, \theta_\star} \langle y_1 | y_{-\infty:0} \rangle \mathbb{1}_X$  if  $\theta \neq \theta_\star$  and  $p^{\theta, \theta_\star}(y_1 | y_{-\infty:0}) := p_1^\theta(y_1 | y_{-\infty:0})$ , where  $p_1^\theta(\cdot | \cdot)$  is the conditional density defined in (3.45), otherwise. If Assumption (K-2) can be shown to hold, it will follow from Theorem 2.2.7 that

$$[\theta_\star] = \operatorname{argmax}_{\theta \in \Theta} \tilde{\mathbb{E}}^{\theta_\star} [\ln p^{\theta, \theta_\star}(Y_1 | Y_{-\infty:0})].$$

Thus the proof will immediately follow if we can moreover show that

$$\ell(\theta_\star, \theta_\star) = \tilde{\mathbb{E}}^{\theta_\star} [\ln p_1^{\theta_\star}(Y_1 | Y_{-\infty:0})]. \quad (3.49)$$

Now, let  $\mathcal{F}$  be a countable separating class of nonnegative bounded functions containing  $\mathbb{1}_X$ , see [Parthasarathy, 2005, Theorem 6.6, Chapter 6] for the existence of such a class. By Lemma 2.2.12, to check Assumption (K-2) it is sufficient to show that Assumption (K-3) is satisfied. Assumption (F-3) and (3.29) imply that for all bounded function  $f$ ,  $\mathcal{F}_f^\theta = \{x \mapsto \kappa^\theta \langle y, y' \rangle(x; f) : (y, y') \in \mathbf{Y}^2\}$  is a class of bounded functions,



and this in turn implies (K-3) by applying Lemma 3.3.11-(i) to some  $x$  such that  $\delta_x \in \mathcal{M}^{\theta^*}(D, r)$  (the existence of such a set  $D$  is guaranteed by (F-2)-(iii)). Thus Assumption (K-2) holds. To complete the proof, it only remains to show that (3.49) is satisfied. According to [Barron, 1985, Theorem 1], we have

$$\tilde{\mathbb{E}}^{\theta^*} [\ln p_1^{\theta^*}(Y_1|Y_{-\infty:0})] = \lim_{n \rightarrow \infty} n^{-1} \ln \pi_{1, Y_1}^{\theta^*} \ell^{\theta^*} \langle Y_{1:n} \rangle \mathbb{1}_{\mathcal{X}}, \quad \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.}, \quad (3.50)$$

where for all  $y_1 \in \mathcal{Y}$  and  $A \in \mathcal{X}$ ,

$$\pi_{1, y_1}^{\theta^*}(A) := \int \pi^{\theta^*}(dx_0 dy_0) \ell^{\theta^*} \langle y_{0:1} \rangle (x_0; A).$$

On the other hand, by applying Lemma 3.3.11-(ii) and by (3.47), it yields that

$$\ell(\theta_*, \theta_*) = \lim_{n \rightarrow \infty} n^{-1} \ln \chi_{Y_{0:1}}^{\theta^*} \ell^{\theta^*} \langle Y_{1:n} \rangle \mathbb{1}_{\mathcal{X}}, \quad \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.}, \quad (3.51)$$

where for all  $y_0, y_1 \in \mathcal{Y}$  and  $A \in \mathcal{X}$ ,

$$\chi_{y_{0:1}}^{\theta^*}(A) := \int \chi(dx_0) \ell^{\theta^*} \langle y_{0:1} \rangle (x_0; A).$$

Note that from (3.27), the measures  $\pi_{1, y_1}^{\theta^*}$  and  $\chi_{y_{0:1}}^{\theta^*}$  admit densities with respect to the same measure  $\mu$  given by

$$\frac{d\pi_{1, y_1}^{\theta^*}}{d\mu}(x_1) = \int \pi^{\theta^*}(dx_0 dy_0) q^{\theta^*}((x_0, y_0); x_1) g^{\theta^*}((x_0, y_0, x_1); y_1) \quad (3.52)$$

and

$$\frac{d\chi_{y_{0:1}}^{\theta^*}}{d\mu}(x_1) = \int \chi(dx_0) q^{\theta^*}((x_0, y_0); x_1) g^{\theta^*}((x_0, y_0, x_1); y_1), \quad (3.53)$$

respectively. Note also that, since  $q^\theta$  and  $g^\theta$  are positive, these densities are positive. Moreover, for all  $y_{0:n} \in \mathcal{Y}^{n+1}$ , we have

$$\chi \ell^{\theta^*} \langle y_{0:n} \rangle \mathbb{1}_{\mathcal{X}} = \int \frac{d\chi_{y_{0:1}}^{\theta^*}}{d\mu}(x_1) \times (\delta_{x_1} \ell^{\theta^*} \langle y_{1:n} \rangle \mathbb{1}_{\mathcal{X}}) \mu(dx_1).$$

The joint density of  $(X_1, Y_{1:n})$  under  $\mathbb{P}^{\theta^*}$  with respect to  $\mu \otimes \nu^{\otimes n}$  is given by

$$p_{1,n}^{\theta^*}(x_1, y_{1:n}) := \frac{d\pi_{1, y_1}^{\theta^*}}{d\mu}(x_1) \times (\delta_{x_1} \ell^{\theta^*} \langle y_{1:n} \rangle \mathbb{1}_{\mathcal{X}}).$$

Combining the four previous displays yields, for all  $y_{0:n} \in \mathbf{Y}^{n+1}$ ,

$$\begin{aligned} & \chi^{\ell^{\theta^*}} \langle y_{0:n} \rangle \mathbb{1}_{\mathbf{X}} \\ &= \int \frac{\int \chi(\mathrm{d}x_0) q^{\theta^*}((x_0, y_0); x_1) g^{\theta^*}((x_0, y_0, x_1); y_1)}{\int \pi^{\theta^*}(\mathrm{d}x' \mathrm{d}y') q^{\theta^*}((x', y'); x_1) g^{\theta^*}((x', y', x_1); y_1)} p_{1,n}^{\theta^*}(x_1, y_{1:n}) \mu(\mathrm{d}x_1). \end{aligned}$$

Dividing this by the density of  $Y_{1:n}$  with respect to  $\nu^{\otimes n}$  under  $\mathbb{P}^{\theta^*}$ , we get

$$\frac{\chi^{\ell^{\theta^*}} \langle Y_{0:n} \rangle \mathbb{1}_{\mathbf{X}}}{\pi_{1,Y_1}^{\theta^*} \ell^{\theta^*} \langle Y_{1:n} \rangle \mathbb{1}_{\mathbf{X}}} = \mathbb{E}^{\theta^*} [R(X_1, Y_{0:1}) | Y_{0:n}],$$

where  $R(x_1, y_{0:1})$  is the ratio between the densities (3.53) and (3.52). Since the denominator (3.52) is the joint probability density of  $X_1$  and  $Y_1$  with respect to  $\mu \otimes \nu$  under  $\mathbb{P}^{\theta^*}$ , we have

$$\mathbb{E}^{\theta^*} [R(X_1, Y_{0:1})] = \mathbb{E}^{\theta^*} [\mathbb{E}^{\theta^*} [R(X_1, Y_{0:1}) | Y_0]] = 1,$$

where we have used that given  $Y_0$ , the numerator (3.53) integrates to one with respect to  $\mu \otimes \nu$ . By Lévy's zero-one law, it thus implies that

$$\lim_{n \rightarrow \infty} \frac{\chi^{\ell^{\theta^*}} \langle Y_{0:n} \rangle \mathbb{1}_{\mathbf{X}}}{\pi_{1,Y_1}^{\theta^*} \ell^{\theta^*} \langle Y_{1:n} \rangle \mathbb{1}_{\mathbf{X}}} = \mathbb{E}^{\theta^*} [R(X_1, Y_{0:1}) | Y_{0:\infty}], \quad \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.},$$

and since by (3.27),  $R(x_1, y_{0:1})$  only takes positive values, this limit is positive. This implies that

$$\lim_{n \rightarrow \infty} n^{-1} \ln \frac{\chi^{\ell^{\theta^*}} \langle Y_{0:n} \rangle \mathbb{1}_{\mathbf{X}}}{\pi_{1,Y_1}^{\theta^*} \ell^{\theta^*} \langle Y_{1:n} \rangle \mathbb{1}_{\mathbf{X}}} = 0, \quad \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.}$$

Combining with (3.50) and (3.51), we finally get (3.49), and the proof therefore follows.  $\square$

All of the results above require that the initial probability  $\chi$  on  $(\mathbf{X}, \mathcal{X})$  should belong to the class  $\mathcal{M}^{\theta^*}(D, r)$ . To ensure this the following conditions are sufficient.

**Proposition 3.3.14.** *Assume that there exist a sequence of sets  $D_s \in \mathcal{X}$ ,  $s \in \{1, \dots, r\}$ , and some positive measurable function  $\delta$  on  $\mathbf{Y}$  such that*

$$\tilde{\mathbb{E}}^{\theta^*} [\ln^- \delta(Y)] < \infty$$

and that for all  $s \in \{1, \dots, r\}$ , with  $D_r = D$ ,

$$\inf_{\theta \in \Theta} \inf_{x \in D_{s-1}} Q^\theta((x, Y); D_s) \geq \delta(Y), \quad \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.} \quad (3.54)$$

and

$$\tilde{\mathbb{E}}^{\theta^*} \left[ \ln^- \inf_{\theta \in \Theta} \inf_{(x, x') \in D_{s-1} \times D_s} g^\theta((x_{s-1}, Y_0, x'); Y_1) \right] < \infty. \quad (3.55)$$

Then, for all probability  $\chi$  on  $(\mathbf{X}, \mathcal{X})$  with  $\chi(D_0) > 0$  belongs to  $\mathcal{M}^{\theta^*}(D, r)$ .

*Proof.* Let  $\chi$  be such that  $\chi(D_0) > 0$ . We have for all  $s \in \{1, \dots, r\}$  and  $\theta \in \Theta$ ,

$$\begin{aligned} & \chi \ell^\theta \langle Y_{0:s} \rangle \mathbb{1}_D \\ &= \int \chi(dx_0) \left( \prod_{i=1}^s Q^\theta((x_{i-1}, Y_{i-1}); dx_i) g^\theta((x_{i-1}, Y_{i-1}, x_i); Y_i) \right) \mathbb{1}_{D_s}(x_s) \\ &\geq \chi(D_0) \prod_{i=1}^s \inf_{\theta \in \Theta} \inf_{x \in D_{i-1}} Q^\theta((x, Y_{i-1}); D_i) \inf_{\theta \in \Theta} \inf_{(x, x') \in D_{i-1} \times D_i} g^\theta((x, Y_{i-1}, x'); Y_i) \\ &\geq \left( \prod_{i=1}^s \inf_{\theta \in \Theta} \inf_{(x, x') \in D_{i-1} \times D_i} g^\theta((x, Y_{i-1}, x'); Y_i) \right) \chi(D_0) \prod_{i=1}^s \delta(Y_{i-1}). \end{aligned}$$

The proof then follows by (3.54) and (3.55) above.  $\square$

When  $Q^\theta$  is a Markov kernel on  $\mathbf{X} \times \mathcal{X}$ , the following lemma, quoted from [Douc and Moulines \[2012\]](#), is useful.

**Lemma 3.3.15.** *Assume that  $\mathbf{X} = \mathbb{R}^d$  for some integer  $d \geq 1$ ,  $\mathcal{X}$  is the associated Borel  $\sigma$ -field and  $Q^\theta$  is a Markov kernel on  $\mathbf{X} \times \mathcal{X}$ . Suppose moreover that for any open subset  $U \in \mathcal{X}$ , the function  $(\theta, x) \mapsto Q^\theta(x; U)$  is lower semi-continuous on the product set  $\mathbf{X} \times \Theta$ . Then, for any  $\delta$  and compact subset  $D_0 \in \mathcal{X}$ , there exists a sequence of compact subsets  $D_s$ ,  $s \in \{1, \dots, r\}$ , satisfying (3.54).*

## 3.4 Postponed Proofs

### 3.4.1 Proof of Proposition 3.2.11

Before proving [Proposition 3.2.11](#), let us first show the following lemma.

**Lemma 3.4.1.** *Assume that (L-1) and (L-2) hold and let  $0 \leq \gamma^- < \gamma^+ \leq 1$ . Then for all  $\eta > 0$ , there exists  $\rho = \rho(\eta) \in (0, 1)$  such that for all positive integer  $n$ ,  $z_{0:n-1} \in \mathbf{Z}^n$  satisfying*

$$n^{-1} \sum_{i=0}^{n-1} \mathbb{1}_K(z_i) \geq \max(1 - \gamma^-, (1 + \gamma^+)/2), \quad (3.56)$$

for any  $\beta \in (\gamma^-, \gamma^+)$ , nonnegative bounded functions  $f$ , probability measures  $\chi$  and  $\chi'$  on  $(\mathbf{X}, \mathcal{X})$ ,  $A \in \mathcal{X}$  and  $\theta \in \Theta$ ,

$$\begin{aligned} & \left| \frac{\chi \mathbf{L}^\theta \langle z_{0:n-1} \rangle f}{\chi \mathbf{L}^\theta \langle z_{0:n-1} \rangle \mathbb{1}_{\mathbf{X}}} - \frac{\chi' \mathbf{L}^\theta \langle z_{0:n-1} \rangle f}{\chi' \mathbf{L}^\theta \langle z_{0:n-1} \rangle \mathbb{1}_{\mathbf{X}}} \right| \\ & \leq 2 \left( \rho^{\lfloor n(\beta - \gamma^-) \rfloor} + \frac{\eta^{\lfloor n(\gamma^+ - \beta) \rfloor / 2}}{\chi(A)\chi'(A)} \left[ \prod_{i=0}^{n-1} (F_{z_i}(A))^2 \right] \right) |f|_\infty, \end{aligned}$$

where, with the convention  $1/0 = \infty$ , for all  $z \in \mathbf{Z}$ ,

$$F_z(A) := \frac{\sup_{\theta \in \Theta} |\mathbf{L}^\theta \langle z \rangle (\cdot; \mathbf{X})|_\infty}{\inf_{\theta \in \Theta} \inf_{x \in A} \mathbf{L}^\theta \langle z \rangle (x; A)}. \quad (3.57)$$

*Proof.* By [Proposition 3.2.10](#), for all nonnegative bounded functions  $f$ ,  $\chi, \chi' \in \mathcal{P}(\mathbf{X}, \mathcal{X})$  and  $\theta \in \Theta$ ,

$$\begin{aligned} & \left| \frac{\chi \mathbf{L}^\theta \langle z_{0:n-1} \rangle f}{\chi \mathbf{L}^\theta \langle z_{0:n-1} \rangle \mathbb{1}_{\mathbf{X}}} - \frac{\chi' \mathbf{L}^\theta \langle z_{0:n-1} \rangle f}{\chi' \mathbf{L}^\theta \langle z_{0:n-1} \rangle \mathbb{1}_{\mathbf{X}}} \right| = \frac{|\Delta_{\chi, \chi'}^\theta \langle z_{0:n-1} \rangle (f, \mathbb{1}_{\mathbf{X}})|}{\chi \mathbf{L}^\theta \langle z_{0:n-1} \rangle \mathbb{1}_{\mathbf{X}} \cdot \chi' \mathbf{L}^\theta \langle z_{0:n-1} \rangle \mathbb{1}_{\mathbf{X}}} \\ & \leq 2\rho^{\lfloor n(\beta - \gamma^-) \rfloor} |f|_\infty + 2\eta^{\lfloor n(\gamma^+ - \beta) \rfloor / 2} \frac{\prod_{i=0}^{n-1} |\mathbf{L}^\theta \langle z_i \rangle (\cdot; \mathbf{X})|_\infty^2}{\chi \mathbf{L}^\theta \langle z_{0:n-1} \rangle \mathbb{1}_{\mathbf{X}} \cdot \chi' \mathbf{L}^\theta \langle z_{0:n-1} \rangle \mathbb{1}_{\mathbf{X}}} |f|_\infty, \end{aligned}$$

where we have used

$$\frac{\chi \mathbf{L}^\theta \langle z_{0:n-1} \rangle f}{\chi \mathbf{L}^\theta \langle z_{0:n-1} \rangle \mathbb{1}_{\mathbf{X}}} \vee \frac{\chi' \mathbf{L}^\theta \langle z_{0:n-1} \rangle f}{\chi' \mathbf{L}^\theta \langle z_{0:n-1} \rangle \mathbb{1}_{\mathbf{X}}} \leq |f|_\infty.$$

We conclude the proof by noting that for all  $A \in \mathcal{X}$ ,

$$\frac{\prod_{i=0}^{n-1} |\mathbf{L}^\theta \langle z_i \rangle (\cdot; \mathbf{X})|_\infty^2}{\chi \mathbf{L}^\theta \langle z_{0:n-1} \rangle \mathbb{1}_{\mathbf{X}} \cdot \chi' \mathbf{L}^\theta \langle z_{0:n-1} \rangle \mathbb{1}_{\mathbf{X}}} \leq \frac{\prod_{i=0}^{n-1} (F_{z_i}(A))^2}{\chi(A)\chi'(A)}.$$

□

**Proof of (3.16).** From (3.3), (3.13) and using that  $\ln(a) - \ln(b) \leq (a-b)/b$  for all  $a, b > 0$ , it follows that for all nonnegative integers  $m, k$  ( $m < k$ ) and sequence  $z_{-m:k} \in \mathbf{Z}^{m+k+1}$ , for all  $\chi, \chi' \in \mathcal{P}(\mathbf{X}, \mathcal{X})$  and bounded measurable function  $f \geq 1$ ,

$$\begin{aligned} & \ln \chi \mathbf{L}^\theta \langle z_k | z_{-m:k-1} \rangle f - \ln \chi' \mathbf{L}^\theta \langle z_k | z_{-m:k-1} \rangle f \\ & \leq \frac{\Delta_{\chi, \chi'}^\theta \langle z_{-m:k-1} \rangle (\mathbf{L}^\theta \langle z_k \rangle f, \mathbb{1}_{\mathbf{X}})}{\chi \mathbf{L}^\theta \langle z_{-m:k-1} \rangle \mathbb{1}_{\mathbf{X}} \cdot \chi' \mathbf{L}^\theta \langle z_{-m:k-1} \rangle (\mathbf{L}^\theta \langle z_k \rangle f)}. \quad (3.58) \end{aligned}$$

Now, let  $0 \leq \gamma^- < \gamma^+ \leq 1$ . By [Proposition 3.2.10](#), for all  $\eta > 0$  and  $\beta \in (\gamma^-, \gamma^+)$ , there exists  $\varrho \in (0, 1)$  such that, for all  $z_{-m:k-1} \in \mathbf{Z}^{m+k}$  satisfying

$$(m+k)^{-1} \sum_{i=-m}^{k-1} \mathbb{1}_{\mathbf{K}}(z_i) \geq \max(1 - \gamma^-, (1 + \gamma^+)/2), \quad (3.59)$$

we have

$$\begin{aligned} & \frac{\Delta_{\chi, \chi'}^{\theta} \langle z_{-m:k-1} \rangle (\mathbf{L}^{\theta} \langle z_k \rangle f, \mathbb{1}_{\mathbf{X}})}{\chi \mathbf{L}^{\theta} \langle z_{-m:k-1} \rangle \mathbb{1}_{\mathbf{X}} \cdot \chi' \mathbf{L}^{\theta} \langle z_{-m:k-1} \rangle (\mathbf{L}^{\theta} \langle z_k \rangle f)} \\ & \leq \varrho^{a_{m,k}} \left[ 1 + \frac{\chi \mathbf{L}^{\theta} \langle z_{-m:k-1} \rangle (\mathbf{L}^{\theta} \langle z_k \rangle f) \cdot \chi' \mathbf{L}^{\theta} \langle z_{-m:k-1} \rangle \mathbb{1}_{\mathbf{X}}}{\chi \mathbf{L}^{\theta} \langle z_{-m:k-1} \rangle \mathbb{1}_{\mathbf{X}} \cdot \chi' \mathbf{L}^{\theta} \langle z_{-m:k-1} \rangle (\mathbf{L}^{\theta} \langle z_k \rangle f)} \right] + 2\eta^{b_{m,k}} C_{m,k}, \end{aligned} \quad (3.60)$$

where  $a_{m,k} := \lfloor n(\beta - \gamma^-) \rfloor$ ,  $b_{m,k} := \lfloor n(\gamma^+ - \beta) \rfloor$  and

$$C_{m,k}^{\theta} := \frac{\prod_{i=0}^{k-1} |\mathbf{L}^{\theta} \langle z_i \rangle (\cdot; \mathbf{X})|_{\infty}^2}{\chi \mathbf{L}^{\theta} \langle z_{-m:k-1} \rangle \mathbb{1}_{\mathbf{X}} \cdot \chi' \mathbf{L}^{\theta} \langle z_{-m:k-1} \rangle (\mathbf{L}^{\theta} \langle z_k \rangle f)} |\mathbf{L}^{\theta} \langle z_k \rangle (\cdot; \mathbf{X})|_{\infty} |f|_{\infty}.$$

In addition by [\(3.13\)](#), we have

$$\begin{aligned} & \frac{\chi \mathbf{L}^{\theta} \langle z_{-m:k-1} \rangle (\mathbf{L}^{\theta} \langle z_k \rangle f) \cdot \chi' \mathbf{L}^{\theta} \langle z_{-m:k-1} \rangle \mathbb{1}_{\mathbf{X}}}{\chi \mathbf{L}^{\theta} \langle z_{-m:k-1} \rangle \mathbb{1}_{\mathbf{X}} \cdot \chi' \mathbf{L}^{\theta} \langle z_{-m:k-1} \rangle (\mathbf{L}^{\theta} \langle z_k \rangle f)} \\ & = 1 + \frac{\Delta_{\chi, \chi'}^{\theta} \langle z_{-m:k-1} \rangle (\mathbf{L}^{\theta} \langle z_k \rangle f, \mathbb{1}_{\mathbf{X}})}{\chi \mathbf{L}^{\theta} \langle z_{-m:k-1} \rangle \mathbb{1}_{\mathbf{X}} \cdot \chi' \mathbf{L}^{\theta} \langle z_{-m:k-1} \rangle (\mathbf{L}^{\theta} \langle z_k \rangle f)}. \end{aligned}$$

Plugging this into [\(3.60\)](#) and then using [\(3.58\)](#) yields

$$\begin{aligned} & \ln \chi \mathbf{L}^{\theta} \langle z_k | z_{-m:k-1} \rangle f - \ln \chi' \mathbf{L}^{\theta} \langle z_k | z_{-m:k-1} \rangle f \\ & \leq 2(1 - \varrho^{a_{m,k}})^{-1} (\varrho^{a_{m,k}} + \eta^{b_{m,k}} C_{m,k}^{\theta}). \end{aligned} \quad (3.61)$$

Note that for all  $z_{-m:k-1} \in \mathbf{Z}^{m+k}$ , we have

$$\begin{aligned} \chi \mathbf{L}^{\theta} \langle z_{-m:k-1} \rangle \mathbb{1}_{\mathbf{X}} & \geq \chi(D) \prod_{i=-m}^{k-1} \inf_{x \in D} \mathbf{L}^{\theta} \langle z_i \rangle (x; D), \\ \chi' \mathbf{L}^{\theta} \langle z_{-m:k-1} \rangle (\mathbf{L}^{\theta} \langle z_k \rangle f) & \geq \chi'(D) \prod_{i=-m}^k \inf_{x \in D} \mathbf{L}^{\theta} \langle z_i \rangle (x; D), \end{aligned}$$

where we have used  $|f|_{\infty} \geq 1$ . Then by exchanging  $\chi$  and  $\chi'$  in [\(3.61\)](#), we obtain the upper bound for  $|\ln \chi \mathbf{L}^{\theta} \langle z_k | z_{-m:k-1} \rangle f - \ln \chi' \mathbf{L}^{\theta} \langle z_k | z_{-m:k-1} \rangle f|$ .

That is, for all  $z_{-m:k-1} \in \mathbf{Z}^{m+k}$  satisfying (3.59), we have

$$\begin{aligned} & \sup_{\theta \in \Theta} \left| \ln \chi \mathbf{L}^\theta \langle z_k | z_{-m:k-1} \rangle f - \ln \chi' \mathbf{L}^\theta \langle z_k | z_{-m:k-1} \rangle f \right| \\ & \leq 2(1 - \varrho^{a_{m,k}})^{-1} \left\{ \varrho^{a_{m,k}} + \frac{\eta^{b_{m,k}} |f|_\infty}{\chi(D) \chi'(D)} \left[ \prod_{i=-m}^{k-1} (F_{z_i}(D))^2 \right] F_{z_k}(D) \right\}, \end{aligned} \quad (3.62)$$

where  $F_z$  is defined in (3.57). Using that  $\ln^+(a/b) \leq \ln^+(a) + \ln^-(b)$  for all positive real numbers  $a$  and  $b$ , we then have

$$\ln^+(F_z(D)) \leq \ln^+ \left( \sup_{\theta \in \Theta} |\mathbf{L}^\theta \langle z \rangle (\cdot; \mathbf{X})|_\infty \right) + \ln^- \left( \inf_{\theta \in \Theta} \inf_{x \in D} \mathbf{L}^\theta \langle z \rangle (x; D) \right). \quad (3.63)$$

By (L-3) and (L-4),

$$\mathbb{E}_Z [\ln^+ F_{Z_0}(D)] < \infty.$$

Now let  $\eta$  be small enough so that  $\mathbb{E}_Z [\ln^+ F_{Z_0}(D)] \leq -\ln \eta$ . Then, by Lemma 3.4.3, for all  $\chi, \chi' \in \mathcal{M}^+(D)$  there exists a  $\mathbb{P}_Z$ -a.s. finite random variable  $M = M(f) > 0$ , and a constant  $\tau \in (0, 1)$  such that, for all  $k \geq 1$  and  $m \geq 0$ ,  $\mathbb{P}_Z$ -a.s.

$$\frac{2}{1 - \varrho^{a_{m,k}}} \left\{ \varrho^{a_{m,k}} + \frac{\eta^{b_{m,k}} |f|_\infty}{\chi(D) \chi'(D)} \left[ \prod_{i=-m}^{k-1} (F_{z_i}(D))^2 \right] F_{z_k}(D) \right\} \leq \tau^{m+k} M,$$

and from (3.62), this implies

$$\sup_{\theta \in \Theta} \left| \ln \chi \mathbf{L}^\theta \langle Z_k | Z_{-m:k-1} \rangle f - \ln \chi' \mathbf{L}^\theta \langle Z_k | Z_{-m:k-1} \rangle f \right| \leq \tau^{m+k} M, \quad \mathbb{P}_Z\text{-a.s.},$$

provided that

$$(m+k)^{-1} \sum_{i=-m}^{k-1} \mathbb{1}_{\mathbf{K}}(Z_i) \geq \max(1 - \gamma^-, (1 + \gamma^+)/2), \quad \mathbb{P}_Z\text{-a.s.} \quad (3.64)$$

To complete, we need to show that there exists a  $\mathbb{P}_Z$ -a.s. finite random variable  $T$  such that for all  $k \geq T$  and for all  $m \geq 0$ , (3.64) holds. Note that under (L-2)(i),

$$1 - \mathbb{P}_Z(Z_0 \in \mathbf{K}) < 2\mathbb{P}_Z(Z_0 \in \mathbf{K}) - 1.$$

Now choose  $\tilde{\gamma}^-, \gamma^-, \gamma^+$  and  $\tilde{\gamma}^+$  such that

$$1 - \mathbb{P}_Z(Z_0 \in \mathbf{K}) < \tilde{\gamma}^- < \gamma^- < \gamma^+ < \tilde{\gamma}^+ < 2\mathbb{P}_Z(Z_0 \in \mathbf{K}) - 1.$$

By construction,

$$\mathbb{P}_Z(Z_0 \in \mathbb{K}) > \max(1 - \tilde{\gamma}^-, (1 + \tilde{\gamma}^+)/2).$$

By (L-1), the Birkhoff ergodic theorem ensures that there exists a  $\mathbb{P}_Z$ -a.s. finite random variable  $B$  such that for all  $k \geq B$  and  $m \geq B$ ,  $\mathbb{P}_Z$ -a.s.,

$$k^{-1} \sum_{i=0}^{k-1} \mathbb{1}_{\mathbb{K}}(Z_i) > \max(1 - \tilde{\gamma}^-, (1 + \tilde{\gamma}^+)/2) \quad (3.65)$$

$$m^{-1} \sum_{i=-m}^{-1} \mathbb{1}_{\mathbb{K}}(Z_i) > \max(1 - \tilde{\gamma}^-, (1 + \tilde{\gamma}^+)/2) \quad (3.66)$$

Now let  $T^+ := B(1 + \gamma^+)/(\tilde{\gamma}^+ - \gamma^+)$ . If  $m \geq B$  and  $k \geq T^+$ , then from (3.65) and (3.66) and using that  $T^+ \geq B$ ,  $\mathbb{P}_Z$ -a.s.,

$$\frac{\sum_{i=-m}^{k-1} \mathbb{1}_{\mathbb{K}}(Z_i)}{m+k} > \frac{\sum_{i=-m}^{-1} \mathbb{1}_{\mathbb{K}}(Z_i) + \sum_{i=0}^{k-1} \mathbb{1}_{\mathbb{K}}(Z_i)}{m+k} = (1 + \tilde{\gamma}^+)/2 > (1 + \gamma^+)/2.$$

On the other hand, if  $0 \leq m < B$  and  $k \geq T^+$ ,

$$\begin{aligned} \frac{\sum_{i=-m}^{k-1} \mathbb{1}_{\mathbb{K}}(Z_i)}{m+k} &\geq \frac{\sum_{i=0}^{k-1} \mathbb{1}_{\mathbb{K}}(Z_i)}{m+k} \\ &\geq \frac{k(1 + \tilde{\gamma}^+)/2}{m+k} > \frac{T^+(1 + \tilde{\gamma}^+)}{2(B + T^+)} = (1 + \gamma^+)/2. \end{aligned}$$

By similar argument, letting  $T^- := B(1 - \gamma^-)/(\gamma^- - \tilde{\gamma}^-)$ , we get that for all  $m \geq 0$  and  $k \geq T^-$ ,

$$\frac{\sum_{i=-m}^{k-1} \mathbb{1}_{\mathbb{K}}(Z_i)}{m+k} \geq 1 - \gamma^-, \quad \mathbb{P}_Z\text{-a.s.}$$

The proof follows by setting  $T = T^+ \vee T^-$ .

**Proof of (3.17).** Observe that for all  $\chi \in \mathcal{P}(\mathbb{X}, \mathcal{X})$ , bounded measurable function  $f \geq 1$  and  $m \in \mathbb{Z}_+, k \in \mathbb{Z}_+^*$ , we can write

$$\chi \mathbf{L}^\theta \langle Z_k | Z_{-m-1:k-1} \rangle f = \chi' \mathbf{L}^\theta \langle Z_k | Z_{-m:k-1} \rangle f,$$

where, for all  $A \in \mathcal{X}$ ,

$$\chi'(A) := \frac{\chi \mathbf{L}^\theta \langle Z_{-m-1} \rangle \mathbb{1}_A}{\chi \mathbf{L}^\theta \langle Z_{-m-1} \rangle \mathbb{1}_{\mathbb{X}}}.$$

Since

$$\chi'(D) = \frac{\chi \mathbf{L}^\theta \langle Z_{-m-1} \rangle \mathbb{1}_D}{\chi \mathbf{L}^\theta \langle Z_{-m-1} \rangle \mathbb{1}_X} \geq \frac{\chi(D)}{F_{Z_{-m-1}}(D)},$$

then (3.62) writes,  $\mathbb{P}_Z$ -a.s.,

$$\begin{aligned} & \sup_{\theta \in \Theta} \left| \ln \chi \mathbf{L}^\theta \langle Z_k | Z_{-m:k-1} \rangle f - \ln \chi \mathbf{L}^\theta \langle Z_k | Z_{-m-1:k-1} \rangle f \right| \\ & \leq \frac{2}{1 - \varrho^{a_{m,k}}} \left\{ \varrho^{a_{m,k}} + \frac{\eta^{b_{m,k}} |f|_\infty}{\chi(D)^2} F_{Z_{-m-1}}(D) \left[ \prod_{i=-m}^{k-1} (F_{Z_i}(D))^2 \right] F_{Z_k}(D) \right\}. \end{aligned}$$

The rest of the proof follows the same lines as the proof of (3.16) and is thus omitted.

**Proof of (3.18).** Note that when  $k = 0$ , (3.64) follows directly from (3.66). The proof is therefore concluded by following the same lines as the proof of (3.17).

### 3.4.2 Proof of Proposition 3.2.13

**Proof of Proposition 3.2.13–(i).** From (3.3), we have for all integer  $m > 0$ ,  $z_{-m:0} \in \mathbf{Z}^{m+1}$  and  $\chi \in \mathcal{M}^+(D)$ ,

$$\chi \mathbf{L}^\theta \langle z_0 | z_{-m:-1} \rangle \mathbb{1}_X \leq \left| \mathbf{L}^\theta \langle z_0 \rangle (\cdot; \mathbf{X}) \right|_\infty. \quad (3.67)$$

Thus, from (3.19) and (L-4), it follows that

$$\mathbb{E}_Z \left[ \ln^+ \mathbf{L}^\theta \langle Z_0 | Z_{-\infty:-1} \rangle \mathbb{1}_X \right] < \infty.$$

Now, for all  $\theta \in \Theta$ ,  $\chi \in \mathcal{M}^+(D)$  and positive integer  $m$ , define a measurable function  $V_{\chi,m}^\theta$  from  $(\mathbf{Z}^{m+1}, \mathcal{Z}^{\otimes(m+1)})$  to  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  by for all  $z_{-m:0} \in \mathbf{Z}^{m+1}$ ,

$$V_{\chi,m}^\theta(z_{-m:0}) := m^{-1} \sum_{k=1}^m \left( \ln^+ \left| \mathbf{L}^\theta \langle z_0 \rangle (\cdot; \mathbf{X}) \right|_\infty - \ln \chi \mathbf{L}^\theta \langle z_0 | z_{-k:-1} \rangle \mathbb{1}_X \right).$$

By (3.67),  $V_{\chi,m}^\theta$  is nonnegative for all positive integer  $m$ . Thus, Fatou's lemma implies

$$\liminf_{n \rightarrow \infty} \mathbb{E}_Z \left[ V_{\chi,m}^\theta \right] \geq \mathbb{E}_Z \left[ \liminf_{m \rightarrow \infty} V_{\chi,m}^\theta \right] \quad (3.68)$$

By definition,

$$\begin{aligned} \liminf_{n \rightarrow \infty} \mathbb{E}_Z \left[ V_{\chi,m}^\theta \right] &= \mathbb{E}_Z \left[ \ln^+ \left| \mathbf{L}^\theta \langle Z_0 \rangle (\cdot; \mathbf{X}) \right|_\infty \right] \\ &\quad - \limsup_{m \rightarrow \infty} m^{-1} \sum_{k=1}^m \mathbb{E}_Z \left[ \ln \chi \mathbf{L}^\theta \langle Z_0 | Z_{-k:-1} \rangle \mathbb{1}_X \right], \quad (3.69) \end{aligned}$$



and

$$\begin{aligned} \mathbb{E}_Z \left[ \liminf_{m \rightarrow \infty} V_{\chi, m}^\theta \right] &= \mathbb{E}_Z \left[ \ln^+ |\mathbf{L}^\theta \langle Z_0 \rangle (\cdot; \mathbf{X})|_\infty \right] \\ &\quad - \mathbb{E}_Z \left[ \limsup_{m \rightarrow \infty} m^{-1} \sum_{k=1}^m \ln \chi \mathbf{L}^\theta \langle Z_0 | Z_{-k:-1} \rangle \mathbb{1}_X \right]. \end{aligned} \quad (3.70)$$

Since the process  $\{Z_k : k \in \mathbb{Z}\}$  is stationary, this implies that for all  $k \in \mathbb{Z}_+$ ,

$$\mathbb{E}_Z \left[ \ln \chi \mathbf{L}^\theta \langle Z_0 | Z_{-k:-1} \rangle \mathbb{1}_X \right] = \mathbb{E}_Z \left[ \ln \chi \mathbf{L}^\theta \langle Z_k | Z_{0:k-1} \rangle \mathbb{1}_X \right],$$

showing that

$$m^{-1} \sum_{k=1}^m \mathbb{E}_Z \left[ \ln \chi \mathbf{L}^\theta \langle Z_0 | Z_{-k:-1} \rangle \mathbb{1}_X \right] = m^{-1} \sum_{k=1}^m \mathbb{E}_Z \left[ \ln \chi \mathbf{L}^\theta \langle Z_k | Z_{0:k-1} \rangle \mathbb{1}_X \right]. \quad (3.71)$$

Cesàro mean convergence moreover implies,  $\mathbb{P}_Z$ -a.s.,

$$\begin{aligned} \limsup_{m \rightarrow \infty} m^{-1} \sum_{k=1}^m \ln \chi \mathbf{L}^\theta \langle Z_0 | Z_{-k:-1} \rangle \mathbb{1}_X \\ = \lim_{m \rightarrow \infty} \ln \chi \mathbf{L}^\theta \langle Z_0 | Z_{-m:-1} \rangle \mathbb{1}_X = \ln \mathbf{L}^\theta \langle Z_0 | Z_{-\infty:-1} \rangle \mathbb{1}_X \end{aligned} \quad (3.72)$$

Thus combining (3.68), (3.69), (3.70), (3.71) and (3.72), we obtain

$$\begin{aligned} \mathbb{E}_Z \left[ \ln \mathbf{L}^\theta \langle Z_0 | Z_{-\infty:-1} \rangle \mathbb{1}_X \right] &\geq \limsup_{m \rightarrow \infty} m^{-1} \sum_{k=1}^m \mathbb{E}_Z \left[ \ln \chi \mathbf{L}^\theta \langle Z_k | Z_{0:k-1} \rangle \right] \mathbb{1}_X \\ &= \limsup_{m \rightarrow \infty} \left\{ \mathbb{E}_Z \left[ m^{-1} \ln \chi \mathbf{L}^\theta \langle Z_{0:m} \rangle \mathbb{1}_X \right] - m^{-1} \mathbb{E}_Z \left[ \ln \chi \mathbf{L}^\theta \langle Z_0 \rangle \mathbb{1}_X \right] \right\} > -\infty, \end{aligned}$$

where the last bound follows from (L-3) and the minorization

$$\ln \chi \mathbf{L}^\theta \langle Z_{0:m} \rangle \mathbb{1}_X \geq \ln \chi(D) + \sum_{k=0}^m \ln \inf_{x \in D} \mathbf{L}^\theta \langle Z_k \rangle (x; D),$$

establishing that

$$\mathbb{E}_Z \left[ \ln^- \mathbf{L}^\theta \langle Z_0 | Z_{-\infty:-1} \rangle \mathbb{1}_X \right] < \infty.$$

Thus (3.20) follows and this completes the proof of Proposition 3.2.13-(i).

**Proof of Proposition 3.2.13-(ii).** According to Proposition 3.2.11, there exist  $\mathbb{P}_Z$ -a.s. finite random variables  $M$  and  $T$  such that for all  $k \geq T$  and  $m \geq 0$ ,  $\mathbb{P}_Z$ -a.s.,

$$\sup_{\theta \in \Theta} \left| \ln \chi \mathbf{L}^\theta \langle Z_k | Z_{-m:k-1} \rangle \mathbb{1}_X - \ln \chi \mathbf{L}^\theta \langle Z_k | Z_{-m-1:k-1} \rangle \mathbb{1}_X \right| \leq M \tau^{m+k},$$

which implies,  $\mathbb{P}_Z$ -a.s.,

$$\sup_{\theta \in \Theta} |\ln \chi \mathbf{L}^\theta \langle Z_k | Z_{0:k-1} \rangle \mathbb{1}_X - \ln \chi \mathbf{L}^\theta \langle Z_k | Z_{-\infty:k-1} \rangle \mathbb{1}_X| \leq \frac{M\tau^k}{1-\tau}.$$

The proof of **Proposition 3.2.13–(ii)** then follows immediately by observing that  $\mathbb{P}_Z$ -a.s.,

$$\begin{aligned} n^{-1} \ln \chi \mathbf{L}^\theta \langle Z_{0:n-1} \rangle \mathbb{1}_X &= n^{-1} \sum_{k=1}^{n-1} \ln \chi \mathbf{L}^\theta \langle Z_k | Z_{0:k-1} \rangle \mathbb{1}_X + n^{-1} \ln \chi \mathbf{L}^\theta \langle Z_0 \rangle \mathbb{1}_X, \\ n^{-1} \ln \mathbf{L}^\theta \langle Z_{0:n-1} | Z_{-\infty:-1} \rangle \mathbb{1}_X &= n^{-1} \sum_{k=1}^{n-1} \ln \mathbf{L}^\theta \langle Z_k | Z_{-\infty:k-1} \rangle \mathbb{1}_X. \end{aligned}$$

**Proof of Proposition 3.2.13–(iii).** Under **(L-5)**, for any  $m \geq p$  and  $x \in D$ ,  $\mathbb{P}_Z$ -a.s., the function  $\theta \mapsto \ln \delta_x \mathbf{L}^\theta \langle Z_0 | Z_{-m:-1} \rangle \mathbb{1}_X$  is continuous on  $\Theta$ . Since

$$\lim_{m \rightarrow \infty} \sup_{\theta \in \Theta} |\ln \delta_x \mathbf{L}^\theta \langle Z_0 | Z_{-m:-1} \rangle \mathbb{1}_X - \ln \mathbf{L}^\theta \langle Z_0 | Z_{-\infty:-1} \rangle \mathbb{1}_X| = 0, \quad \mathbb{P}_Z\text{-a.s.},$$

and since the uniform limit of continuous function is continuous, then the function  $\theta \mapsto \ln \mathbf{L}^\theta \langle Z_0 | Z_{-\infty:-1} \rangle \mathbb{1}_X$  is  $\mathbb{P}_Z$ -a.s. continuous. Now let  $\theta_0 \in \Theta$ . For  $\rho > 0$ , let  $\mathbf{B}(\theta_0, \rho)$  be an open ball of center  $\theta_0$  and radius  $\rho$ . Then for all  $\theta \in \mathbf{B}(\theta_0, \rho)$ , we have

$$\begin{aligned} \ln \mathbf{L}^\theta \langle Z_0 | Z_{-\infty:-1} \rangle \mathbb{1}_X &\leq \sup_{\theta \in \mathbf{B}(\theta_0, \rho)} \ln \mathbf{L}^\theta \langle Z_0 | Z_{-\infty:-1} \rangle \mathbb{1}_X \\ &\leq \sup_{\theta \in \Theta} \sup_{x \in X} \ln \mathbf{L}^\theta \langle Z_0 \rangle (x; \mathbf{X}), \quad \mathbb{P}_Z\text{-a.s.}, \end{aligned}$$

which by **(L-4)** implies

$$\begin{aligned} \sup_{\theta \in \mathbf{B}(\theta_0, \rho)} \mathbb{E}_Z [\ln \mathbf{L}^\theta \langle Z_0 | Z_{-\infty:-1} \rangle \mathbb{1}_X] &\leq \mathbb{E}_Z \left[ \sup_{\theta \in \mathbf{B}(\theta_0, \rho)} \ln \mathbf{L}^\theta \langle Z_0 | Z_{-\infty:-1} \rangle \mathbb{1}_X \right] \\ &\leq \mathbb{E}_Z \left[ \sup_{\theta \in \Theta} \sup_{x \in X} \ln \mathbf{L}^\theta \langle Z_0 \rangle (x; \mathbf{X}) \right] < \infty. \end{aligned}$$

By monotone convergence theorem, it yields that

$$\begin{aligned} \lim_{\rho \downarrow 0} \sup_{\theta \in \mathbf{B}(\theta_0, \rho)} \mathbb{E}_Z [\ln \mathbf{L}^\theta \langle Z_0 | Z_{-\infty:-1} \rangle \mathbb{1}_X] &\leq \lim_{\rho \downarrow 0} \mathbb{E}_Z \left[ \sup_{\theta \in \mathbf{B}(\theta_0, \rho)} \ln \mathbf{L}^\theta \langle Z_0 | Z_{-\infty:-1} \rangle \mathbb{1}_X \right] \\ &= \mathbb{E}_Z \left[ \lim_{\rho \downarrow 0} \sup_{\theta \in \mathbf{B}(\theta_0, \rho)} \ln \mathbf{L}^\theta \langle Z_0 | Z_{-\infty:-1} \rangle \mathbb{1}_X \right] \\ &= \mathbb{E}_Z [\ln \mathbf{L}^{\theta_0} \langle Z_0 | Z_{-\infty:-1} \rangle \mathbb{1}_X], \end{aligned}$$

showing that

$$\theta \mapsto \mathbb{E}_Z [\ln \mathbf{L}^\theta \langle Z_0 | Z_{-\infty:-1} \rangle \mathbb{1}_X]$$

is  $\mathbb{P}_Z$ -a.s. upper semi-continuous at  $\theta_0 \in \Theta$ . The proof then follows since this holds with arbitrary  $\theta_0 \in \Theta$ .

### 3.4.3 Proof of Lemma 3.3.5

Let  $\bar{\zeta}^\theta$  is an invariant distribution for  $\bar{K}^\theta$ , i.e.  $\bar{\zeta}^\theta \bar{K}^\theta = \bar{\zeta}^\theta$ . We will show that for probability measure  $\zeta^\theta$  defined by for all  $A \in \mathcal{W}$ ,  $\zeta^\theta(A) = \bar{\zeta}^\theta(\mathbf{W}^r \times A)$ , then  $\zeta^\theta(K^\theta)^r = \zeta^\theta$ . Now from (3.36) and (3.37), we have for  $A \in \mathcal{W}$ ,

$$\begin{aligned} \zeta^\theta(K^\theta)^r(A) &= \int_{\mathbf{W}^r} \zeta^\theta(dw_0) K^\theta(w_0; dw_1) \cdots K^\theta(w_{r-2}; dw_{r-1}) K^\theta(w_{r-1}; A) \\ &= \int_{\mathbf{W}} \bar{\zeta}^\theta(\mathbf{W}^r \times dw_r) \mathbb{1}_{\mathbf{W}}(w_r) T^\theta(w_r; \mathbf{W}^{r-1} \times A) \\ &= \int_{\mathbf{W}^{r+1}} \bar{\zeta}^\theta(d\bar{w}) \bar{K}^\theta(\bar{w}; \mathbf{W}^r \times A) = \bar{\zeta}^\theta \bar{K}^\theta(\mathbf{W}^r \times A) \\ &= \bar{\zeta}^\theta(\mathbf{W}^r \times A) = \zeta^\theta(A). \end{aligned}$$

Conversely, suppose that  $\xi^\theta$  is an invariant distribution for  $(K^\theta)^r$ , that is,  $\xi^\theta(K^\theta)^r = \xi^\theta$ . Define a probability measure  $\bar{\xi}^\theta$  on  $(\mathbf{W}^{r+1}, \mathcal{W}^{\otimes(r+1)})$  by, for all  $A_0, \dots, A_r \in \mathcal{W}$ ,

$$\bar{\xi}^\theta(A_0 \times \cdots \times A_r) = \int_{A_0} \xi^\theta(dw) T^\theta(w; A_1 \times \cdots \times A_r).$$

Then from (3.36) and (3.37), for all  $A_0, \dots, A_r \in \mathcal{W}$ , we have

$$\begin{aligned} \bar{\xi}^\theta \bar{K}^\theta(A_0 \times \cdots \times A_r) &= \int_{\mathbf{W}^{r+1}} \bar{\xi}^\theta(d\bar{w}) \bar{K}^\theta(\bar{w}; A_0 \times \cdots \times A_r) \\ &= \int_{A_0} \bar{\xi}^\theta(\mathbf{W}^r \times dw_r) T^\theta(w_r; A_1 \times \cdots \times A_r) \\ &= \int_{A_0 \times \cdots \times A_r} \xi(dw_0) K^\theta(w_{r-1}; dw_1) \cdots K^\theta(w_0; dw_r) \\ &= \bar{\xi}^\theta(A_0 \times \cdots \times A_r), \end{aligned}$$

which implies that  $\bar{\xi}^\theta \bar{K}^\theta = \bar{\xi}^\theta$ .

For uniqueness, if  $\bar{K}^\theta$  admits a unique stationary probability measure  $\bar{\zeta}^\theta$ , then  $\bar{\zeta}^\theta = \bar{\xi}^\theta$ . Consequently, for all  $A \in \mathcal{W}$ ,

$$\begin{aligned} \zeta^\theta(A) &= \bar{\zeta}^\theta(\mathbf{W}^r \times A) = \bar{\xi}^\theta(\mathbf{W}^r \times A) \\ &= \int_{\mathbf{W}} \xi^\theta(dw) T^\theta(w; \mathbf{W}^{r-1} \times A) = \xi^\theta(K^\theta)^r(A) = \xi^\theta(A). \end{aligned}$$

To conclude, it is worth noting that if  $\bar{\xi}^\theta$  is also an invariant probability measure for  $\bar{K}^\theta$ , then the probability measure  $\xi^\theta$  defined by for all  $A \in \mathcal{W}$ ,

$$\xi^\theta(A) = \bar{\xi}^\theta(\mathbf{W}^r \times A)$$

is also an invariant probability measure for  $(K^\theta)^r$ . Then, if  $(K^\theta)^r$  admits a unique invariant probability measure, it follows that  $\bar{\xi}^\theta(\mathbf{W}^r \times \cdot) = \bar{\zeta}^\theta(\mathbf{W}^r \times \cdot)$  and this implies for all  $A_0, \dots, A_r \in \mathcal{W}$ ,

$$\begin{aligned} \bar{\zeta}^\theta(A_0 \times \dots \times A_r) &= \bar{\zeta}^\theta \bar{K}^\theta(A_0 \times \dots \times A_r) \\ &= \int_{A_0} \bar{\zeta}^\theta(\mathbf{W}^r \times dw_r) T^\theta(w_r; A_1 \times \dots \times A_r) \\ &= \int_{A_0} \bar{\xi}^\theta(\mathbf{W}^r \times dw_r) T^\theta(w_r; A_1 \times \dots \times A_r) \\ &= \bar{\xi}^\theta \bar{K}^\theta(A_0 \times \dots \times A_r) = \bar{\xi}^\theta(A_0 \times \dots \times A_r), \end{aligned}$$

completing the proof.

### 3.4.4 Proof of Proposition 3.3.9

We prove Proposition 3.3.9 by applying the results obtained in Section 3.2. First let us set, for all  $(\theta, z) \in \Theta \times \mathcal{Z}$ ,  $\mathbf{L}^\theta \langle z \rangle = \ell^\theta \langle z \rangle$  and define  $\mathbf{L}_n^\theta$  and  $\mathbf{L}^\theta \langle \cdot | \cdot \rangle$  as in Definition 3.2.1 and Definition 3.2.2, respectively. From these settings, we have for all  $\theta \in \Theta$  and  $y_{0:nr} \in \mathbf{Y}^{nr+1}$  with for all  $k \in \mathbb{Z}_+$ ,  $z_k = z_{kr:(k+1)r}$ ,

$$\mathbf{L}^\theta \langle z_{0:n-1} \rangle = \ell^\theta \langle y_{0:nr} \rangle$$

and for  $0 < p < n$ ,

$$\mathbf{L}^\theta \langle z_{p:n-1} | z_{0:p-1} \rangle = \ell^\theta \langle y_{pr+1:nr} | y_{0:pr} \rangle.$$

Obviously, (3.2) is satisfied. Comparing the assumptions, we have (F-1) implies (L-1); (F-2) corresponds to (L-2) and (L-3); (F-3) implies (L-4); and (F-4) yields (L-5). This implies that all the results in Section 3.2 can also be derived under the present Assumptions (F-1)–(F-4).

Now, let  $\chi \in \mathcal{P}(\mathbf{X}, \mathcal{X})$  such that  $\chi(D) > 0$ . We will show that for all non-negative and bounded measurable function  $f$ , the sequence  $\{\chi \ell^\theta \langle Y_1 | Y_{-m:0} \rangle f : m \in \mathbb{Z}_+^*\}$  converges  $\tilde{\mathbb{P}}^{\theta*}$ -a.s. For any  $t \in \{0, \dots, r-1\}$ , set

$$\chi_{m,t}^\theta(A) := \frac{\chi \ell^\theta \langle Y_{-mr-t:-mr} \rangle \mathbb{1}_A}{\chi \ell^\theta \langle Y_{-mr-t:-mr} \rangle \mathbb{1}_{\mathbf{X}}}.$$

From (3.32) and (3.35), for all  $m \in \mathbb{Z}_+$  and  $t \in \{0, \dots, r-1\}$ , we have

$$\begin{aligned} & \chi^{\ell^\theta} \langle Y_1 | Y_{-mr-t:0} \rangle f \\ &= \frac{\chi_{m,t}^\theta \mathbf{L}^\theta \langle Z_{-m:-1} \rangle \left[ \int_{\mathbf{X}} Q^\theta((\cdot, Y_0); dx_1) g^\theta((\cdot, Y_0, x_1); Y_1) f(x_1) \right]}{\chi_{m,t}^\theta \mathbf{L}^\theta \langle Z_{-m:-1} \rangle \mathbb{1}_{\mathbf{X}}}. \end{aligned}$$

Let  $0 \leq \gamma^- < \gamma^+ \leq 1$ . Lemma 3.4.1 shows that for all  $t \in \{0, \dots, r-1\}$  and  $\eta > 0$ , there exists  $\rho \in (0, 1)$  such that, if

$$m^{-1} \sum_{i=-m}^{-1} \mathbb{1}_{\mathbf{K}}(Z_i) \geq \max(1 - \gamma^-, (1 + \gamma^+)/2),$$

then for all  $\beta \in (\gamma^-, \gamma^+)$  and  $\theta \in \Theta$ , with  $F_z := F_z(D)$  defined by (3.57),

$$\begin{aligned} & \left| \chi^{\ell^\theta} \langle Y_1 | Y_{-mr-t:0} \rangle f - \chi^{\ell^\theta} \langle Y_1 | Y_{-mr:0} \rangle f \right| \\ & \leq c_f \left( \rho^{\lfloor m(\beta - \gamma^-) \rfloor} + \frac{\eta^{\lfloor m(\gamma^+ - \beta) \rfloor / 2}}{\chi_{m,t}^\theta(D)} \prod_{i=-m}^{-1} (F_{Z_i})^2 \right) \sup_{\theta \in \Theta} \sup_{(x, x') \in \mathbf{X}^2} g^\theta((x, Y_0, x'); Y_1) \\ & \leq c_f \left( \rho^{\lfloor m(\beta - \gamma^-) \rfloor} + \eta^{\lfloor m(\gamma^+ - \beta) \rfloor / 2} F'_{-m} \prod_{i=-m}^{-1} F_{Z_i}^2 \right) \sup_{\theta \in \Theta} \sup_{x, x' \in \mathbf{X}} g^\theta((x, Y_0, x'); Y_1) \end{aligned}$$

where  $c_f := 2|f|_\infty$  and

$$F'_{-m} := \max_{t \in \{0, \dots, r-1\}} \frac{1}{\inf_{\theta \in \Theta} \chi_{m,t}^\theta(D)}.$$

Note that the sequence  $\{F'_{-m} : m \in \mathbb{Z}_+\}$  is stationary. Then by the same argument as in the proof of (3.63), with  $\chi \in \mathcal{M}^{\theta^*}(D, r)$ , we have  $\tilde{\mathbb{E}}^{\theta^*}[\ln^+ F'_{-m}] < \infty$ . By choosing  $\gamma^-$  and  $\gamma^+$  such that  $\tilde{\mathbb{P}}^{\theta^*}(Z_0 \in \mathbf{K}) > \max(1 - \gamma^-, (1 + \gamma^+)/2)$  and by applying Lemma 3.4.3, it follows that there exist  $\varrho_\chi \in (0, 1)$  and a  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s. finite random variable  $C_\chi$  such that for all  $\ell > 1$ ,

$$\left| \chi^{\ell^\theta} \langle Y_1 | Y_{-\ell:0} \rangle f - \chi^{\ell^\theta} \langle Y_1 | Y_{-\ell-1:0} \rangle f \right| \leq C_\chi |f|_\infty \varrho_\chi^\ell, \quad \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.}$$

Similarly, for all  $\chi' \in \mathcal{P}(\mathbf{X}, \mathcal{X})$  such that  $\chi'(D) > 0$ , there exist  $\varrho_{\chi, \chi'} \in (0, 1)$  and a  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s. finite random variable  $C_{\chi, \chi'}$  such that for all  $\ell > 1$ ,

$$\left| \chi^{\ell^\theta} \langle Y_1 | Y_{-\ell:0} \rangle f - \chi'^{\ell^\theta} \langle Y_1 | Y_{-\ell:0} \rangle f \right| \leq C_{\chi, \chi'} |f|_\infty \varrho_{\chi, \chi'}^\ell, \quad \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.}$$

This yields that for all  $\chi \in \mathcal{P}(\mathbf{X}, \mathcal{X})$  satisfying  $\chi(D) > 0$ , the sequence  $\{\chi^{\ell^\theta} \langle Y_1 | Y_{-\ell:0} \rangle f : \ell \in \mathbb{Z}_+^*\}$  converges  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s. to a limit denoted

by  $\ell^{\theta, \theta^*} \langle Y_1 | Y_{-\infty:0} \rangle f$  which is measurable on  $(\mathbf{Y}^{\mathbb{Z}^-}, \mathcal{Y}^{\otimes \mathbb{Z}^-})$  and does not depend on  $\chi \in \mathcal{M}^{\theta^*}(D, r)$ . This establishes (3.43). On the other hand,  $\{\chi \ell^\theta \langle Y_1 | Y_{-\ell:0} \rangle : \ell \in \mathbb{Z}_+^*\}$  is a sequence of ( $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.) finite measure on  $(\mathbf{X}, \mathcal{X})$  and since (3.43) holds with any nonnegative and bounded measurable function  $f$ , this yield that  $\ell^{\theta, \theta^*} \langle Y_1 | Y_{-\infty:0} \rangle$  define a (nonnegative) measure on  $(\mathbf{X}, \mathcal{X})$ . This shows that  $\ell^{\theta, \theta^*}$  is a nonnegative kernel on  $\mathbf{Y}^{\mathbb{Z}^-} \times \mathcal{X}$ . Moreover, by stationarity of  $\{Y_k : k \in \mathbb{Z}\}$ , we obtain that for all  $k \in \mathbb{Z}_+$ ,  $\theta \in \Theta$ ,  $\chi \in \mathcal{M}^{\theta^*}(D, r)$  and nonnegative and bounded measurable function  $f$ ,

$$\lim_{m \rightarrow \infty} \chi \ell^\theta \langle Y_k | Y_{-m:k-1} \rangle f = \ell^{\theta, \theta^*} \langle Y_k | Y_{-\infty:k-1} \rangle f, \quad \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.},$$

showing the first part of Proposition 3.3.9–(i). To complete the proof of Proposition 3.3.9–(i), it remains to show that for all  $\theta \in \Theta$  and nonnegative and bounded measurable function  $f$ ,

$$\tilde{\mathbb{E}}^{\theta^*} [|\ln \ell^{\theta, \theta^*} \langle Y_k | Y_{-\infty:k-1} \rangle f|] < \infty.$$

Note that we have,  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\chi \ell^\theta \langle Y_k | Y_{-m:k-1} \rangle f \leq \sup_{(x, x') \in \mathbf{X}^2} g^\theta((x, Y_{k-1}, x'); Y_k) |f|_\infty,$$

and this implies that

$$\ell^{\theta, \theta^*} \langle Y_k | Y_{-\infty:k-1} \rangle f \leq \sup_{(x, x') \in \mathbf{X}^2} g^\theta((x, Y_{k-1}, x'); Y_k) |f|_\infty, \quad \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.}$$

Under (F-3), we get that

$$\tilde{\mathbb{E}}^{\theta^*} [\ln^+ \ell^{\theta, \theta^*} \langle Y_k | Y_{-\infty:k-1} \rangle f] < \infty.$$

Note also that if  $\mathbb{E}(A) < \infty$ , then we have  $\mathbb{E}(A - B) = \mathbb{E}(A) - \mathbb{E}(B)$ . Thus we can write,

$$\begin{aligned} & \tilde{\mathbb{E}}^{\theta^*} [\ln \ell^{\theta, \theta^*} \langle Y_k | Y_{-\infty:k-1} \rangle f] \\ &= \tilde{\mathbb{E}}^{\theta^*} [\ln^+ \ell^{\theta, \theta^*} \langle Y_k | Y_{-\infty:k-1} \rangle f] - \tilde{\mathbb{E}}^{\theta^*} [\ln^- \ell^{\theta, \theta^*} \langle Y_k | Y_{-\infty:k-1} \rangle f]. \end{aligned}$$

This will immediately imply

$$\tilde{\mathbb{E}}^{\theta^*} [\ln^- \ell^{\theta, \theta^*} \langle Y_k | Y_{-\infty:k-1} \rangle f] < \infty$$

if we can show that

$$\tilde{\mathbb{E}}^{\theta^*} [\ln \ell^{\theta, \theta^*} \langle Y_k | Y_{-\infty:k-1} \rangle f] > -\infty.$$

By stationarity of  $\{Y_k : k \in \mathbb{Z}\}$ , we have

$$\begin{aligned}
 & r\tilde{\mathbb{E}}^{\theta_\star} [\ln \ell^{\theta, \theta_\star} \langle Y_1 | Y_{-\infty:0} \rangle f] \\
 &= r\tilde{\mathbb{E}}^{\theta_\star} [\ln^+ \ell^{\theta, \theta_\star} \langle Y_1 | Y_{-\infty:0} \rangle f] - r\tilde{\mathbb{E}}^{\theta_\star} [\ln^- \ell^{\theta, \theta_\star} \langle Y_1 | Y_{-\infty:0} \rangle f] \\
 &= \tilde{\mathbb{E}}^{\theta_\star} \left[ \sum_{k=1}^r \ln^+ \ell^{\theta, \theta_\star} \langle Y_k | Y_{-\infty:k-1} \rangle f \right] - \tilde{\mathbb{E}}^{\theta_\star} \left[ \sum_{k=1}^r \ln^- \ell^{\theta, \theta_\star} \langle Y_k | Y_{-\infty:k-1} \rangle f \right] \\
 &= \tilde{\mathbb{E}}^{\theta_\star} \left[ \sum_{k=1}^r \ln \ell^{\theta, \theta_\star} \langle Y_k | Y_{-\infty:k-1} \rangle f \right].
 \end{aligned}$$

By [Corollary 3.2.12](#),  $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.,

$$\begin{aligned}
 \prod_{k=1}^r \ell^{\theta, \theta_\star} \langle Y_k | Y_{-\infty:k-1} \rangle f &= \prod_{k=1}^r \lim_{m \rightarrow \infty} \chi^{\ell^\theta} \langle Y_k | Y_{-mr:k-1} \rangle f \\
 &= \lim_{m \rightarrow \infty} \prod_{k=1}^r \chi^{\ell^\theta} \langle Y_k | Y_{-mr:k-1} \rangle f \\
 &= \lim_{m \rightarrow \infty} \chi^{\ell^\theta} \langle Y_{1:r} | Y_{-mr:k-1} \rangle f \\
 &= \lim_{m \rightarrow \infty} \chi^{\mathbf{L}^\theta} \langle Z_0 | Z_{-m:-1} \rangle f \\
 &= \mathbf{L}^{\theta, \theta_\star} \langle Z_0 | Z_{-\infty:-1} \rangle f.
 \end{aligned}$$

Thus,

$$\tilde{\mathbb{E}}^{\theta_\star} [\ln \ell^{\theta, \theta_\star} \langle Y_k | Y_{-\infty:k-1} \rangle f] = r^{-1} \tilde{\mathbb{E}}^{\theta_\star} [\ln \mathbf{L}^{\theta, \theta_\star} \langle Z_0 | Z_{-\infty:-1} \rangle f] > -\infty,$$

where we have used the result of [Proposition 3.2.13–\(i\)](#) such that

$$\tilde{\mathbb{E}}^{\theta_\star} [|\ln \mathbf{L}^{\theta, \theta_\star} \langle Z_0 | Z_{-\infty:-1} \rangle f|] < \infty.$$

This completes the proof of [Proposition 3.3.9–\(i\)](#).

To prove [Proposition 3.3.9–\(ii\)](#), observe that for all  $s$  and  $t$  and bounded measurable function  $h \geq 1$ , we have

$$\begin{aligned}
 \chi^{\ell^\theta} \langle Y_{0:s+t} \rangle h &\leq (\chi^{\ell^\theta} \langle Y_{0:s} \rangle \mathbb{1}_X) \left( \prod_{i=0}^{t-1} \sup_{\theta \in \Theta} \sup_{(x, x') \in X^2} g^\theta((x, Y_{s+i}, x'); Y_{s+i+1}) \right) |h|_\infty \\
 &\leq (\chi^{\ell^\theta} \langle Y_{0:s} \rangle h) \left( \prod_{i=0}^{t-1} \sup_{\theta \in \Theta} \sup_{(x, x') \in X^2} g^\theta((x, Y_{s+i}, x'); Y_{s+i+1}) \right) |h|_\infty.
 \end{aligned}$$

Thus,

$$\begin{aligned} n^{-1} \ln \chi^{\ell^\theta} \langle Y_{0:nr+r} \rangle h - \sum_{i=t}^{r-1} n^{-1} \ln^+ A_{n,t}^i - n^{-1} |h|_\infty &\leq n^{-1} \ln \chi^{\ell^\theta} \langle Y_{0:nr+t} \rangle h \\ &\leq n^{-1} \ln \chi^{\ell^\theta} \langle Y_{0:nr} \rangle h + \sum_{i=0}^{t-1} n^{-1} \ln^+ A_{n,t}^i + n^{-1} |h|_\infty, \end{aligned} \quad (3.73)$$

where  $A_{n,t}^i := \sup_{\theta \in \Theta} \sup_{(x', x'') \in \mathcal{X}^2} g^\theta((x', Y_{nr+i}, x''); Y_{nr+i+1})$ . Note that for all  $i$  and  $t$ ,  $\{A_{n,t}^i : n \in \mathbb{Z}_+\}$  is stationary. Thus from (F-3) and Lemma 3.4.2,

$$\lim_{n \rightarrow \infty} A_{n,t}^i = 0, \quad \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.}$$

From (3.73) and by noting that for all  $i$  and  $t$ ,  $\{A_{n,t}^i : n \in \mathbb{Z}_+\}$  no longer depend on  $\theta \in \Theta$ , for all  $t \in \{0, \dots, r\}$ ,  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in \Theta} n^{-1} |\ln \chi^{\ell^\theta} \langle Y_{0:nr-r+t} \rangle h - \ln \chi^{\mathbf{L}^\theta} \langle Z_{0:n-1} \rangle h| = 0. \quad (3.74)$$

Since  $t$  is arbitrary in  $\{0, \dots, r\}$ , using Proposition 3.2.13–(ii), we finally obtain,

$$\begin{aligned} \lim_{n \rightarrow \infty} n^{-1} \ln \chi^{\ell^\theta} \langle Y_{0:n} \rangle h &= r^{-1} \lim_{m \rightarrow \infty} m^{-1} \chi^{\mathbf{L}^\theta} \langle Z_{0:m-1} \rangle h \\ &= r^{-1} \tilde{\mathbb{E}}^{\theta^*} [\ln \mathbf{L}^{\theta, \theta^*} \langle Z_0 | Z_{-\infty:-1} \rangle h] \\ &= \tilde{\mathbb{E}}^{\theta^*} [\ln \ell^{\theta, \theta^*} \langle Y_1 | Y_{-\infty:0} \rangle h], \quad \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.} \end{aligned}$$

This result holds with all bounded measurable function  $h \geq 1$ , and thus propagates to all bounded measurable function  $f \geq 0$ , concluding the proof.

### 3.4.5 Some Useful Lemmas

The following lemmas are taken from Douc and Moulines [2012], and thus stated without proofs.

**Lemma 3.4.2.** *Let  $\{Z_k : k \in \mathbb{Z}\}$  be a sequence of nonnegative random variables on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  having the same marginal distribution, that is, for any  $k \in \mathbb{Z}$  and any nonnegative and measurable function  $f$ ,  $\mathbb{E}[f(Z_k)] = \mathbb{E}[f(Z_0)]$ .*

- (i) *Assume that  $\mathbb{E}[\ln^+ Z_0] < \infty$ . Then, for all  $\beta \in (0, 1)$ ,  $\sup_{k \geq 0} \beta^k Z_k < \infty$ ,  $\mathbb{P}$ -a.s. Moreover,  $\lim_{n \rightarrow \infty} n^{-1} \ln^+ Z_n = 0$ ,  $\mathbb{P}$ -a.s.*
- (ii) *Assume that  $\mathbb{E}[|\ln Z_0|] < \infty$ . Then, for all  $\beta \in (0, 1)$ ,  $\sup_{k \in \mathbb{Z}} \beta^{|k|} Z_k < \infty$  and  $\inf_{k \in \mathbb{Z}} \beta^{-|k|} Z_k > 0$ ,  $\mathbb{P}$ -a.s.*



**Lemma 3.4.3.** *Let  $\{U_k : k \in \mathbb{Z}\}$ ,  $\{V_k : k \in \mathbb{Z}\}$  and  $\{W_k : k \in \mathbb{Z}\}$  be stationary sequences such that*

$$\mathbb{E}[\ln^+ U_0] < \infty, \quad \mathbb{E}[\ln^+ V_0] < \infty, \quad \mathbb{E}[\ln^+ W_0] < \infty.$$

*Assume moreover that  $\{V_k : k \in \mathbb{Z}\}$  is ergodic. Then, for all  $\eta$  in  $(0, 1)$  such that  $\mathbb{E}[\ln^+ V_0] < -\ln \eta$  and for all  $\rho$  in  $(0, 1)$ , there exists a  $\mathbb{P}$ -a.s. finite random variable  $M$  and a constant  $\varrho$  in  $(0, 1)$  such that for all  $k \geq 1$ ,  $m \geq 0$ ,*

$$\rho^{m+k} + \eta^{m+k} W_{-m} \left( \prod_{i=-m}^{k-1} V_i \right) U_k \leq \varrho^{m+k} M, \quad \mathbb{P}\text{-a.s.}$$

# Chapter 4

## Observation-driven Models: Handy Sufficient Conditions for the Convergence of the MLE

**Summary.** This chapter, largely inspired by [Douc et al. \[2015\]](#), aims at establishing asymptotic properties for a general class of first-order observation-driven times series models. We generalize the class of observation-driven models considered by [Douc et al. \[2013\]](#) in the sense that we allow the conditional law of each observation to depend on the parameter. The existence of ergodic solutions and the consistency of the maximum likelihood estimator (MLE) are derived under easy-to-check conditions. The obtained conditions appear to apply for a wide class of models. We illustrate our results with specific observation-driven times series, including the recently introduced NBIN-GARCH and NM-GARCH models considered in [Section 2.4 of Chapter 2](#), demonstrating the consistency of the MLE for these two models.

### 4.1 Introduction

Observation-driven models (ODMs) introduced by [Cox \[1981\]](#) have been widely used in various disciplines such as in economics ([Pindyck and Rubinfeld \[1998\]](#)), environmental study ([Bhaskaran et al. \[2013\]](#)), epidemiology and public health study ([Zeger \[1988\]](#), [Davis et al. \[1999\]](#), [Ferland et al. \[2006\]](#)), finance ([Liesenfeld and Richard \[2003\]](#), [Rydberg and Shephard \[2003\]](#), [Fokianos and Tjøstheim \[2011\]](#), [Francq and Zakoian \[2011\]](#)) and population dynamics ([Ives et al. \[2003\]](#)). The celebrated GARCH(1, 1) model, see [Bollerslev \[1986\]](#), as well as most of the models derived from this one, see [Boller-](#)

slev [2008] for a list of some of them, are typical examples of ODMs. A number of contributions on this class of models dealing with countable data includes Streett [2000], Davis et al. [2003], Heinen [2003], Fokianos et al. [2009], Neumann [2011], Doukhan et al. [2012], Davis and Liu [2012] and Douc et al. [2013]. ODMs have the nice feature that the associated (conditional) likelihood and its derivatives are easy to compute and the prediction is straightforward. The consistency of the maximum likelihood estimator (in short, MLE) for the class of these models can be cumbersome, except when it can be derived using computations specific to the studied model (the GARCH(1, 1) case being one of the most celebrated example). When the observed variable is discrete, general consistency results have been obtained only recently in Davis and Liu [2012] or Douc et al. [2013] (see also in Henderson et al. [2011] for the existence of stationary and ergodic solutions to some observation-driven time series models). However, the consistency result of Douc et al. [2013] applies to some restricted class of models and does not cover the case where the distribution of the observations given the hidden variable also depends on an unknown parameter. We now introduce three simple examples, to which the results of Douc et al. [2013] can not be directly applied. The first one is the negative binomial integer-valued GARCH (NBIN-GARCH) model, which was first introduced by Zhu [2011] as a generalization of the Poisson IN-GARCH model. The NBIN-GARCH model belongs to the class of integer-valued GARCH models that account for overdispersion (namely, variability is larger than mean) and potential heavy tails in the high values. In Zhu [2011], the author applied this model to treat the data of counts of poliomyelitis cases in the USA from 1970 to 1983 reported by the Centres for Disease Control, where data overdispersion was detected. The estimation result showed that NBIN-GARCH(1, 1) outperformed among some commonly used models such as Poisson and Double Poisson models. The NBIN-GARCH(1, 1) model is formally defined as follows.

**Example 4.1.1** (NBIN-GARCH(1, 1) model). Consider the following recursion.

$$\begin{aligned} X_{k+1} &= \omega + aX_k + bY_k, \\ Y_{k+1} | X_{0:k+1}, Y_{0:k} &\sim \mathcal{NB}\left(r, \frac{X_{k+1}}{1 + X_{k+1}}\right), \end{aligned} \tag{4.1}$$

where  $X_k$  takes values in  $\mathsf{X} = \mathbb{R}_+$ ,  $Y_k$  takes values in  $\mathbb{Z}_+$  and  $\theta = (\omega, a, b, r) \in (0, \infty)^4$  is an unknown parameter. In (4.1),  $\mathcal{NB}(r, p)$  denotes the negative binomial distribution with parameters  $r > 0$  and  $p \in (0, 1)$ , that is, if  $Y \sim \mathcal{NB}(r, p)$ , then  $\mathbb{P}(Y = k) = \frac{\Gamma(k+r)}{k!\Gamma(r)}(1-p)^r p^k$  for all  $k \geq 0$ , where  $\Gamma$  stands for

the Gamma function. Though substantial analysis on this model has been carried out in the literature, to the best of our knowledge, the consistency of the MLE has not been treated; see the end of the discussions of Section 6 in [Zhu \[2011\]](#).

The second example is the univariate normal mixture GARCH (NM-GARCH) proposed by [Haas et al. \[2004\]](#) and later considered by [Alexander and Lazar \[2006\]](#). The NM-GARCH model is another natural extension of GARCH processes, where the usual Gaussian conditional distribution of the observations given the hidden volatility variable is replaced by a mixture of Gaussian distributions given a hidden vector volatility variable. The NM-GARCH model has the ability of capturing time variation in both conditional skewness and kurtosis, while the classical GARCH cannot. In [Alexander and Lazar \[2006\]](#), the NM-GARCH(1, 1) model was applied to examine the data of exchange rates consisting of daily prices in US dollars of three different currencies (British pound, euro and Japanese yen) from 2 January 1989 to 31 December 2002. The empirical evidence suggested the best performance of NM(2)-GARCH(1, 1) when compared to the classical GARCH(1, 1), standardized symmetric and skewed  $t$ -GARCH(1, 1) models applied to this same data. The definition of this model is formally stated as follows.

**Example 4.1.2** (NM( $d$ )-GARCH(1, 1) model). Let  $d \in \mathbb{Z}_+^*$  and consider the following recursion:

$$\begin{aligned} \mathbf{X}_{k+1} &= \boldsymbol{\omega} + \mathbf{A}\mathbf{X}_k + Y_k^2 \mathbf{b}, \\ Y_{k+1} | \mathbf{X}_{0:k+1}, Y_{0:k} &\sim G^\theta(\mathbf{X}_{k+1}; \cdot), \\ \frac{dG^\theta(\mathbf{x}; \cdot)}{d\nu}(y) &= \sum_{\ell=1}^d \gamma_\ell \frac{e^{-y^2/2x_\ell}}{(2\pi x_\ell)^{1/2}}, \quad \mathbf{x} \in (0, \infty)^d, y \in \mathbb{R}, \end{aligned} \tag{4.2}$$

where  $\nu$  is the Lebesgue measure on  $\mathbb{R}$ ,  $\mathbf{X}_k = [X_{1,k} \dots X_{d,k}]^T$  takes values in  $\mathbf{X} = \mathbb{R}_+^d$ ;  $\boldsymbol{\gamma} = [\gamma_1 \dots \gamma_d]^T$  a  $d$ -dimensional vector of mixture coefficients belonging to the  $d$ -dimensional simplex

$$\mathbf{P}_d = \left\{ \boldsymbol{\gamma} \in \mathbb{R}_+^d : \sum_{\ell=1}^d \gamma_\ell = 1 \right\}; \tag{4.3}$$

$\boldsymbol{\omega}$ ,  $\mathbf{b}$  are  $d$ -dimensional vector parameters with positive and nonnegative entries, respectively; and  $\mathbf{A}$  is a  $d \times d$  matrix parameter with nonnegative entries. Here we have  $\theta = (\boldsymbol{\gamma}, \boldsymbol{\omega}, \mathbf{A}, \mathbf{b})$ . Note that  $G^\theta$  depends on  $\theta$  only through the mixture coefficients  $\gamma_1, \dots, \gamma_d$ . If  $d = 1$ , we obtain the usual conditionally Gaussian GARCH(1, 1) process. In such a case, since  $\boldsymbol{\gamma} = \gamma_1 =$

1,  $G^\theta$  no longer depends on  $\theta$ . To our knowledge, the usual consistency proof of the MLE for the GARCH cannot be directly adapted to this model.

Finally, we consider the following new example, where a threshold is added to the usual IN-GARCH model in the conditional distribution.

**Example 4.1.3** (Threshold IN-GARCH model). Consider the following recursion.

$$\begin{aligned} X_{k+1} &= \omega + aX_k + bY_k, \\ Y_{k+1}|X_{0:k+1}, Y_{0:k} &\sim \mathcal{P}(X_{k+1} \wedge \tau), \end{aligned} \tag{4.4}$$

where  $X_k$  takes values in  $\mathsf{X} = (0, \infty)$ ;  $Y_k$  takes values in  $\mathbb{Z}_+$  and  $\theta = (\omega, a, b, \tau) \in (0, \infty)^4$  is an unknown parameter. Comparing with the usual IN-GARCH model, a threshold  $\tau$  has been added in the conditional observation distribution. This corresponds to the practical case where the hidden variable has an influence on the observation up to this threshold.

For a well-specified model, a classical approach to establish the consistency of the MLE generally involves two main steps: first the MLE converges to the maximizing set  $\Theta_\star$  of a limit criterion, and second the maximizing set indeed reduces to the true parameter  $\theta_\star$ , which is usually referred to as solving the *identifiability* problem. In this chapter, we are interested in solving the problem involved in the first step, that is, the convergence of MLE. We extend the convergence result of MLE obtained in Douc et al. [2013], which is valid for a restricted class of models, to a larger class of models in which the three examples introduced above are embedded. More precisely, we show the convergence of MLE in observation-driven models where the probability distributions of observations explicitly depend on the unknown parameters. Moreover, we provide very simple conditions that are easy to check, as shown by the three illustrating examples.

The chapter is organized as follows. Specific definitions and notation are introduced in Section 4.2. Then, Section 4.3 contains the main result of this chapter, that is, sufficient conditions for the existence of ergodic solutions and for the consistency of the MLE. These results are then applied in Section 4.4 to the three examples introduced above. Numerical experiments for the NBIN-GARCH(1, 1) model are given in Section 4.5. Finally, Section 4.6 provides the proofs of the main results, inspired from Douc et al. [2013], and of a technical lemma.

## 4.2 Definitions and Notation

Consider a bivariate stochastic process  $\{(X_k, Y_k) : k \in \mathbb{Z}_+\}$  on  $\mathsf{X} \times \mathsf{Y}$ , where  $(\mathsf{X}, d_X)$  is a complete and separable metric space endowed with the

associated Borel  $\sigma$ -field  $\mathcal{X}$  and  $(Y, \mathcal{Y})$  is a Borel space. Let  $(\Theta, \Delta)$ , the set of parameters, be a compact metric space,  $\{G^\theta : \theta \in \Theta\}$  be a family of probability kernels on  $X \times Y$  and  $\{(x, y) \mapsto \psi_y^\theta(x) : \theta \in \Theta\}$  be a family of measurable functions from  $(X \times Y, \mathcal{X} \otimes \mathcal{Y})$  to  $(X, \mathcal{X})$ . The observation-driven time series model can be formally defined as follows.

**Definition 4.2.1.** A time series  $\{Y_k : k \in \mathbb{Z}_+\}$  valued in  $Y$  is said to be distributed according to an *observation-driven model*, in short ODM, with parameter  $\theta \in \Theta$  if there is a bivariate Markov chain  $\{(X_k, Y_k) : k \in \mathbb{Z}_+\}$  on  $X \times Y$  whose transition kernel  $K^\theta$  satisfies

$$K^\theta((x, y); dx' dy') = \delta_{\psi_y^\theta(x)}(dx') G^\theta(x'; dy'), \quad (4.5)$$

where  $\delta_a$  denotes the Dirac mass at point  $a$ . Moreover, we will say that the observation-driven time series model is dominated by some  $\sigma$ -finite measure  $\nu$  on  $(Y, \mathcal{Y})$  if for all  $x \in X$ , the probability kernel  $G^\theta(x; \cdot)$  is dominated by  $\nu$ . In this case we denote by  $g^\theta(x; \cdot)$  its Radon–Nikodym derivative,  $g^\theta(x; y) = \frac{dG^\theta(x; \cdot)}{d\nu}(y)$ , and we always assume that for all  $(x, y) \in X \times Y$  and for all  $\theta \in \Theta$ ,

$$g^\theta(x; y) > 0.$$

A dominated parametric ODM is thus characterized by the collection  $\{(g^\theta, \psi^\theta) : \theta \in \Theta\}$ . The class of ODMs is a particular case of *partially-observed Markov chains* since only  $Y_k$ 's are observed, whereas  $X_k$ 's are *hidden* variables. Note that our notation for observation-driven models is slightly different from that of Douc et al. [2013] where their sequence  $\{Y_k\}$  corresponds to our sequence  $\{Y_{k-1}\}$ . Note also that the process  $\{X_k : k \geq 1\}$  by itself is a Markov chain with transition kernel defined by

$$R^\theta(x; A) = \int \mathbb{1}_A(\psi_y^\theta(x)) G^\theta(x; dy), \quad x \in X, A \in \mathcal{X}. \quad (4.6)$$

However, ODMs do not belong to the class of hidden Markov models. This can be seen in the following recursive relation, which holds for all  $k \geq 0$ ,

$$\begin{aligned} X_{k+1} &= \psi_{Y_k}^\theta(X_k), \\ Y_{k+1} | \mathcal{F}_k &\sim G^\theta(X_{k+1}; \cdot), \end{aligned} \quad (4.7)$$

where  $\mathcal{F}_k = \sigma(X_\ell, X_{\ell+1}, Y_\ell : \ell \leq k, \ell \in \mathbb{Z}_+)$ , and which can be represented graphically in [Figure 4.1](#).

The most popular example is the GARCH(1, 1) process, where  $G^\theta(x; \cdot)$  is a centered (say Gaussian) distribution with variance  $x$  and  $\psi_y^\theta(x)$  is an affine

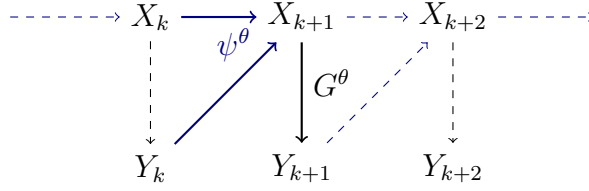


Figure 4.1 – Graphical Representation of the Observation-driven Model.

function of  $x$  and  $y^2$ . One can readily check that Examples 4.1.1 and 4.1.2 are other instances of dominated observation-driven models.

The inference about model parameter is carried out by relying on the conditional likelihood of the observations  $(Y_1, \dots, Y_n)$  given  $X_1 = x$  for an arbitrary  $x \in \mathbf{X}$ . The corresponding conditional density function with respect to  $\nu^{\otimes n}$  is, under parameter  $\theta$ , for all  $x \in \mathbf{X}$ ,

$$y_{1:n} \mapsto \prod_{k=1}^n g^\theta(\psi^\theta\langle y_{1:k-1} \rangle(x); y_k), \quad (4.8)$$

where, for any vector  $y_{1:p} = (y_1, \dots, y_p) \in \mathbf{Y}^p$ ,  $\psi^\theta\langle y_{1:p} \rangle$  is the  $\mathbf{X} \rightarrow \mathbf{X}$  function obtained as the successive composition of  $\psi_{y_1}^\theta$ ,  $\psi_{y_2}^\theta$ , ..., and  $\psi_{y_p}^\theta$ ,

$$\psi^\theta\langle y_{1:p} \rangle = \psi_{y_p}^\theta \circ \psi_{y_{p-1}}^\theta \circ \dots \circ \psi_{y_1}^\theta, \quad (4.9)$$

with the convention  $\psi^\theta\langle y_{s:t} \rangle(x) = x$  for  $s > t$ . Then, the corresponding (conditional) Maximum Likelihood Estimator (MLE)  $\hat{\theta}_{x,n}$  of the parameter  $\theta$ , is defined by

$$\hat{\theta}_{x,n} \in \operatorname{argmax}_{\theta \in \Theta} \mathbb{L}_{x,n}^\theta\langle Y_{1:n} \rangle, \quad (4.10)$$

where

$$\mathbb{L}_{x,n}^\theta\langle y_{1:n} \rangle := n^{-1} \sum_{k=1}^n \ln g^\theta(\psi^\theta\langle y_{1:k-1} \rangle(x); y_k). \quad (4.11)$$

In this contribution, we study the convergence of  $\hat{\theta}_{x,n}$  as  $n \rightarrow \infty$  for some well-chosen value of  $x$  under the assumption that the model is well specified and the observations are in a steady state. This means that we assume that the observations  $\{Y_k : k \in \mathbb{Z}_+\}$  are distributed according to  $\tilde{\mathbb{P}}^{\theta_\star}$  with  $\theta_\star \in \Theta$ , where, for all  $\theta \in \Theta$ ,  $\tilde{\mathbb{P}}^\theta$  denotes the stationary distribution of the observation-driven time series corresponding to the parameter  $\theta$ . However whether such a distribution is well defined is not always obvious. We will use the following ergodicity assumption.

(A-1) For all  $\theta \in \Theta$ , the transition kernel  $K^\theta$  of the complete chain admits a unique stationary distribution  $\pi^\theta$  on  $\mathsf{X} \times \mathsf{Y}$ .

With this assumption, we can now define  $\tilde{\mathbb{P}}^\theta$ . The following notation and definitions will be used throughout the chapter.

**Definition 4.2.2.** For any probability distribution  $\mu$  on  $\mathsf{X} \times \mathsf{Y}$ , we denote by  $\mathbb{P}_\mu^\theta$  the distribution of the Markov chain  $\{(X_k, Y_k) : k \in \mathbb{Z}_+\}$  with kernel  $K^\theta$  and initial probability measure  $\mu$ . Under Assumption (A-1), we denote by  $\pi_1^\theta$  and  $\pi_2^\theta$  the marginal distributions of  $\pi^\theta$  on  $\mathsf{X}$  and  $\mathsf{Y}$ , respectively and by  $\mathbb{P}^\theta$  and  $\tilde{\mathbb{P}}^\theta$  the probability distributions defined respectively as follows.

- a)  $\mathbb{P}^\theta$  denotes the extension of  $\mathbb{P}_{\pi^\theta}^\theta$  on the whole line  $(\mathsf{X} \times \mathsf{Y})^\mathbb{Z}$ .
- b)  $\tilde{\mathbb{P}}^\theta$  is the corresponding projection on the component  $\mathsf{Y}^\mathbb{Z}$ .

We also use the symbols  $\mathbb{E}^\theta$  and  $\tilde{\mathbb{E}}^\theta$  to denote the expectations corresponding to  $\mathbb{P}^\theta$  and  $\tilde{\mathbb{P}}^\theta$ , respectively.

## 4.3 Main Results

### 4.3.1 Preliminaries

In this section, we follow the same lines as in Douc et al. [2013] to derive the convergence of the MLE  $\hat{\theta}_{x,n}$  for a general class of observation-driven models. The approach is to establish that, as the number of observations  $n \rightarrow \infty$ , there exists a  $(\mathsf{Y}^\mathbb{Z}, \mathcal{Y}^{\otimes \mathbb{Z}}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  measurable function  $p^\theta(\cdot|\cdot)$  such that the normalized log-likelihood  $\mathsf{L}_{x,n}^\theta(Y_{1:n})$  defined in (4.11), for some appropriate value of  $x$ , can be approximated by

$$n^{-1} \sum_{k=1}^n \ln p^\theta(Y_k | Y_{-\infty:k-1}).$$

To define  $p^\theta(\cdot|\cdot)$ , we set, for all  $y_{-\infty:1} \in \mathsf{Y}^{\mathbb{Z}^-}$ , whenever the following limit is well defined,

$$p^\theta(y_1 | y_{-\infty:0}) = \begin{cases} \lim_{m \rightarrow \infty} g^\theta(\psi^\theta\langle y_{-m:0} \rangle(x); y_1) & \text{if the limit exists,} \\ \infty & \text{otherwise.} \end{cases} \quad (4.12)$$

By (A-1), the process  $Y$  is ergodic under  $\tilde{\mathbb{P}}^{\theta^*}$  and provided that

$$\tilde{\mathbb{E}}^{\theta^*} [\ln^+ p^\theta(Y_1 | Y_{-\infty:0})] < \infty,$$



it follows that

$$\lim_{n \rightarrow \infty} \mathbb{L}_{x,n}^{\theta} \langle Y_{1:n} \rangle = \tilde{\mathbb{E}}^{\theta_{\star}} [\ln p^{\theta}(Y_1 | Y_{-\infty:0})], \quad \tilde{\mathbb{P}}^{\theta_{\star}}\text{-a.s.}$$

In this chapter, we show that with probability tending to one, the MLE  $\hat{\theta}_{x,n}$  eventually lies in a neighborhood of the set

$$\Theta_{\star} = \operatorname{argmax}_{\theta \in \Theta} \tilde{\mathbb{E}}^{\theta_{\star}} [\ln p^{\theta}(Y_1 | Y_{-\infty:0})], \quad (4.13)$$

which only depends on  $\theta_{\star}$ . In this contribution, we provide easy-to-check sufficient conditions implying

$$\lim_{n \rightarrow \infty} \Delta(\hat{\theta}_{x,n}, \Theta_{\star}) = 0, \quad \tilde{\mathbb{P}}^{\theta_{\star}}\text{-a.s.}, \quad (4.14)$$

but, for the sake of brevity, we do not precisely determine the set  $\Theta_{\star}$ . Many approaches have been proposed to investigate this problem, which is often referred to as the *identifiability* problem. In particular cases, one can prove that  $\Theta_{\star} = \{\theta_{\star}\}$ , in which case the strong consistency of the MLE follows from (4.14). We will mention a general result which precises how the set  $\Theta_{\star}$  is related to the true parameter  $\theta_{\star}$  in [Remark 4.3.4](#). For the moment, let us mention that we have

$$\theta_{\star} \in \Theta_{\star}, \quad (4.15)$$

provided that the following assumption holds:

(B-1) For all  $\theta, \theta_{\star} \in \Theta$ , we have

- (i) If  $\theta \neq \theta_{\star}$ ,  $y \mapsto p^{\theta}(y | Y_{-\infty:0})$  is a density function  $\tilde{\mathbb{P}}^{\theta_{\star}}$ -a.s.
- (ii) Under  $\tilde{\mathbb{P}}^{\theta_{\star}}$ , the function  $y \mapsto p^{\theta_{\star}}(y | Y_{-\infty:0})$  is the conditional density function of  $Y_1$  given  $Y_{-\infty:0}$ .

Indeed, (4.15) follows by writing for all  $\theta \in \Theta$ ,

$$\begin{aligned} \tilde{\mathbb{E}}^{\theta_{\star}} [\ln p^{\theta_{\star}}(Y_1 | Y_{-\infty:0}) - \ln p^{\theta}(Y_1 | Y_{-\infty:0})] &= \tilde{\mathbb{E}}^{\theta_{\star}} \left[ \ln \frac{p^{\theta_{\star}}(Y_1 | Y_{-\infty:0})}{p^{\theta}(Y_1 | Y_{-\infty:0})} \right] \\ &= \tilde{\mathbb{E}}^{\theta_{\star}} \left[ \tilde{\mathbb{E}}^{\theta_{\star}} \left[ \ln \frac{p^{\theta_{\star}}(Y_1 | Y_{-\infty:0})}{p^{\theta}(Y_1 | Y_{-\infty:0})} \middle| Y_{-\infty:0} \right] \right], \end{aligned}$$

which is nonnegative under (B-1) since it is the expectation of a conditional Kullback-Leibler divergence.

### 4.3.2 Convergence of the MLE

In this part, we always assume that (A-1) holds. The following is a list of additional assumptions on which our convergence result relies.

(A-2) There exists a function  $\bar{V} : \mathsf{X} \rightarrow \mathbb{R}_+$  such that, for all  $\theta \in \Theta$ ,  $\pi_1^\theta(\bar{V}) < \infty$ .

**Remark 4.3.1.** Assumption (A-2) is usually obtained as a byproduct of the proof of Assumption (A-1), see Section 4.3.3. It is here stated as an assumption for convenience.

The following set of conditions can readily be checked on  $g^\theta$  and  $\psi^\theta$ .

(B-2) For all  $y \in \mathsf{Y}$ , the function  $(\theta, x) \mapsto g^\theta(x; y)$  is continuous on  $\Theta \times \mathsf{X}$ .

(B-3) For all  $y \in \mathsf{Y}$ , the function  $(\theta, x) \mapsto \psi_y^\theta(x)$  is continuous on  $\Theta \times \mathsf{X}$ .

The function  $\bar{V}$  appearing in (B-4)(viii) below is the same one as in Assumption (A-2). Moreover, in this condition and throughout the chapter we write  $f \lesssim V$  for a real-valued function  $f$  and a nonnegative function  $V$  defined on the same space  $\mathsf{X}$ , whenever there exists a positive constant  $c$  such that  $|f(x)| \leq cV(x)$  for all  $x \in \mathsf{X}$ .

(B-4) There exist  $x_1 \in \mathsf{X}$ , a closed set  $\mathsf{X}_1 \subseteq \mathsf{X}$ ,  $\varrho \in (0, 1)$ ,  $C \geq 0$  and measurable functions  $\bar{\psi} : \mathsf{X} \rightarrow \mathbb{R}_+$ ,  $H : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  and  $\bar{\phi} : \mathsf{Y} \rightarrow \mathbb{R}_+$  such that the following assertions hold.

(i) For all  $\theta \in \Theta$  and  $(x, y) \in \mathsf{X} \times \mathsf{Y}$ ,  $\psi_y^\theta(x) \in \mathsf{X}_1$ .

(ii) 
$$\sup_{(\theta, x, y) \in \Theta \times \mathsf{X}_1 \times \mathsf{Y}} g^\theta(x; y) < \infty.$$

(iii) For all  $\theta \in \Theta$ ,  $n \in \mathbb{Z}_+$ ,  $x \in \mathsf{X}$ , and  $y_{1:n} \in \mathsf{Y}^n$ ,

$$d_{\mathsf{X}}(\psi^\theta\langle y_{1:n} \rangle(x_1), \psi^\theta\langle y_{1:n} \rangle(x)) \leq \varrho^n \bar{\psi}(x). \quad (4.16)$$

(iv)  $\bar{\psi}$  is locally bounded.

(v) For all  $\theta \in \Theta$  and  $y \in \mathsf{Y}$ ,  $\bar{\psi}(\psi_y^\theta(x_1)) \leq \bar{\phi}(y)$ .

(vi) For all  $\theta \in \Theta$  and  $(x, x', y) \in \mathsf{X}_1 \times \mathsf{X}_1 \times \mathsf{Y}$ ,

$$\left| \ln \frac{g^\theta(x; y)}{g^\theta(x'; y)} \right| \leq H(d_{\mathsf{X}}(x, x')) e^{C(d_{\mathsf{X}}(x_1, x) \vee d_{\mathsf{X}}(x_1, x'))} \bar{\phi}(y). \quad (4.17)$$

(vii)  $H(u) = O(u)$  as  $u \rightarrow 0$ .

(viii) If  $C = 0$ , then, for all  $\theta \in \Theta$ ,

$$G^\theta \ln^+ \bar{\phi} \lesssim \bar{V}; \quad (4.18)$$

otherwise, for all  $\theta \in \Theta$ ,

$$G^\theta \bar{\phi} \lesssim \bar{V}. \quad (4.19)$$

Let us now state our main result as follows.

**Theorem 4.3.2.** *Assume that (A-1), (A-2), (B-2), (B-3) and (B-4) hold. Then, letting  $x_1 \in \mathsf{X}$  as in (B-4), the function  $p^\theta(\cdot|\cdot)$  defined by (4.12) with  $x = x_1$  satisfies (B-1) and the convergence (4.14) of the MLE holds with the set  $\Theta_\star$  defined by (4.13).*

For convenience, the proof is postponed to [Section 4.6.1](#).

**Remark 4.3.3.** As noticed in [Douc et al. \[2013\]](#), the techniques used to prove [Theorem 4.3.2](#) also apply in the misspecified case, where  $Y$  is not distributed according to  $\tilde{\mathbb{P}}^{\theta_\star}$ . We do not pursue in this direction in this contribution.

The consistency of the MLE then follows from [Theorem 4.3.2](#) by the following remark.

**Remark 4.3.4.** In many specific cases, one can show that  $\Theta_\star$  defined by (4.13) is the singleton  $\{\theta_\star\}$ . However this task appears to be quite difficult in some cases such as [Example 4.1.3](#). Instead one can use [[Douc et al., 2014](#), Section 4.2], where it is shown that the assumptions of [Theorem 4.3.2](#) imply that  $\Theta_\star$  is exactly the set of parameters  $\theta$  such that  $\tilde{\mathbb{P}}^\theta = \tilde{\mathbb{P}}^{\theta_\star}$ . Thus we can conclude that the MLE converges to the *equivalence class* of the true parameter. This type of consistency has been introduced by [Leroux \[1992\]](#) in the context of hidden Markov models in order to disentangle the proof of the consistency from the problem of identifiability. Recall that the model is identifiable if and only if the equivalent classes  $\{\theta \in \Theta : \tilde{\mathbb{P}}^\theta = \tilde{\mathbb{P}}^{\theta_\star}\}$  reduce to singletons  $\{\theta_\star\}$  for all  $\theta_\star \in \Theta$ .

### 4.3.3 Ergodicity

In this section, the ODM is studied to prove the condition (A-1). Since this is a “for all  $\theta$  (...)” condition, to save space and alleviate the notational burden, we will drop the superscript  $\theta$  from, for example,  $G^\theta$ ,  $R^\theta$  and  $\psi^\theta$  and respectively write, instead,  $G$ ,  $R$  and  $\psi$ .

Ergodicity of Markov chains are usually studied using  $\varphi$ -irreducibility. This approach is well known to be quite efficient when dealing with fully dominated models, see [Meyn and Tweedie \[2009\]](#). It is not at all the same picture for ODMs, where other tools need to be invoked, see [Fokianos and Tjøstheim \[2011\]](#), [Douc et al. \[2013\]](#). Since the ergodicity is studied for a given parameter  $\theta$ , the ergodicity results of [Douc et al. \[2013\]](#) directly apply, even though ODMs are restricted to the case where  $g$  does not depend on the unknown parameter  $\theta$  in this reference. Our main contribution here is to focus on an easy-to-check list of assumptions yielding the ergodicity

conditions (A-1) and (A-2). We also provide a lemma (Lemma 4.3.6) which gives the construction of the instrumental functions  $\alpha$  and  $\phi$  used in the list of assumptions.

(A-3) The measurable space  $(X, d_X)$  is a locally compact, complete and separable metric space and its associated  $\sigma$ -field  $\mathcal{X}$  is the Borel  $\sigma$ -field.

(A-4) There exist  $(\lambda, \beta) \in (0, 1) \times \mathbb{R}_+$  and a measurable function  $V : X \rightarrow \mathbb{R}_+$  such that  $RV \leq \lambda V + \beta$  and  $\{V \leq M\}$  is compact for any  $M > 0$ .

(A-5) The Markov kernel  $R$  is weak Feller, that is, for any continuous and bounded function  $f$  defined on  $X$ ,  $Rf$  is continuous and bounded on  $X$ .

(A-6) The Markov kernel  $R$  has a reachable point, that is, there exists  $x_0 \in X$  such that, for any  $x \in X$  and any neighborhood  $\mathcal{N}$  of  $x_0$ ,  $R^m(x; \mathcal{N}) > 0$  for at least one positive integer  $m$ .

(A-7) We have 
$$\sup_{\substack{(x, x', y) \in X^2 \times Y \\ x \neq x'}} \frac{d_X(\psi_y(x), \psi_y(x'))}{d_X(x, x')} < 1.$$

(A-8) There exist a measurable function  $\alpha$  from  $X^2$  to  $[0, 1]$ , a measurable function  $\phi : X^2 \rightarrow X$  and a measurable function  $W : X^2 \rightarrow [1, \infty)$  such that the following assertions hold.

(i) For all  $(x, x') \in X^2$  and  $y \in Y$ ,

$$\min \{g(x; y), g(x'; y)\} \geq \alpha(x, x')g(\phi(x, x'); y). \quad (4.20)$$

(ii) For all  $x \in X$ ,  $W(x, \cdot)$  is finitely bounded in a neighborhood of  $x$ , that is, there exists  $\gamma_x > 0$  such that 
$$\sup_{x' \in B(x, \gamma_x)} W(x, x') < \infty.$$

(iii) For all  $(x, x') \in X^2$ ,  $1 - \alpha(x, x') \leq d_X(x, x')W(x, x')$ .

(iv) 
$$\sup \left( \int_Y W(\psi_y(x), \psi_y(x')) G(\phi(x, x'); dy) - W(x, x') \right) < \infty,$$
 where the sup is taken over all  $(x, x') \in X^2$ .

We can now state the main ergodicity result.

**Theorem 4.3.5.** *Conditions (A-3), (A-4), (A-5), (A-6), (A-7) and (A-8) imply that  $K$  admits a unique stationary distribution  $\pi$  on  $X \times Y$ . Moreover,  $\pi_1 \bar{V} < \infty$  for every  $\bar{V} : X \rightarrow \mathbb{R}_+$  such that  $\bar{V} \lesssim V$ .*

The proof of Theorem 4.3.5 is postponed to Section 4.6.2 for convenience.

The first conclusion of Theorem 4.3.5 can directly be applied for all  $\theta \in \Theta$  to check (A-1). The second conclusion can be used to check (A-2). In doing so, one must take care of the fact that although  $V$  may depend on  $\theta$ ,  $\bar{V}$  does not.

Assumptions (A-4), (A-5) and (A-6) have to be checked directly on the Markov kernel  $R$  defined by (4.6). To this end it can be useful to define, for any given  $x \in \mathsf{X}$ , the distribution

$$\bar{\mathbb{P}}_x := \mathbb{P}_{\delta_x \otimes G(x; \cdot)} \quad (4.21)$$

on  $(\mathsf{X} \times \mathsf{Y})^{\mathbb{Z}_+}$ , where  $\mathbb{P}_\mu$  is defined for any distribution  $\mu$  on  $\mathsf{X} \times \mathsf{Y}$  as in Definition 4.2.2. Then the first component process  $\{X_k : k \in \mathbb{Z}_+\}$  associated to  $\bar{\mathbb{P}}_x$  is a Markov chain with Markov kernel  $R$  and initial distribution  $\delta_x$ .

We now provide a general framework for constructing  $\alpha$  and  $\phi$  that appear in (A-8).

**Lemma 4.3.6.** *Suppose that  $\mathsf{X} = \mathsf{C}^{\mathsf{S}}$  for some measurable space  $(\mathsf{S}, \mathcal{S})$  and  $\mathsf{C} \subseteq \mathbb{R}$ . Thus, for all  $x \in \mathsf{X}$ , we write  $x = (x_s)_{s \in \mathsf{S}}$ , where  $x_s \in \mathsf{C}$  for all  $s \in \mathsf{S}$ . Suppose moreover that for all  $x = (x_s)_{s \in \mathsf{S}} \in \mathsf{X}$ , we can express the conditional density  $g(x; \cdot)$  as a mixture of densities of the form  $j(x_s)h(x_s; \cdot)$  over  $s \in \mathsf{S}$ . This means that for all  $t \in \mathsf{C}$ ,  $y \mapsto j(t)h(t; y)$  is a density with respect to  $\nu$  and there exists a probability measure  $\mu$  on  $(\mathsf{S}, \mathcal{S})$  such that*

$$g(x; y) = \int_{\mathsf{S}} j(x_s)h(x_s; y)\mu(ds), \quad y \in \mathsf{Y}. \quad (4.22)$$

We moreover assume that  $h$  takes nonnegative values and that one of the two following assumptions holds.

(H-1) For all  $y \in \mathsf{Y}$ , the function  $h(\cdot; y) : t \mapsto h(t; y)$  is nondecreasing.

(H-2) For all  $y \in \mathsf{Y}$ , the function  $h(\cdot; y) : t \mapsto h(t; y)$  is nonincreasing.

For all  $(x, x') \in \mathsf{X}^2$ , denoting  $x \wedge x' := (\min\{x_s, x'_s\})_{s \in \mathsf{S}}$  and  $x \vee x' := (\max\{x_s, x'_s\})_{s \in \mathsf{S}}$ , we define  $\alpha(x, x')$  and  $\phi(x, x')$  as

$$\begin{cases} \alpha(x, x') = \inf_{s \in \mathsf{S}} \left\{ \frac{j(x_s \vee x'_s)}{j(x_s \wedge x'_s)} \right\} & \text{and } \phi(x, x') = x \wedge x' & \text{under (H-1);} \\ \alpha(x, x') = \inf_{s \in \mathsf{S}} \left\{ \frac{j(x_s \wedge x'_s)}{j(x_s \vee x'_s)} \right\} & \text{and } \phi(x, x') = x \vee x' & \text{under (H-2).} \end{cases}$$

Then  $\alpha$  and  $\phi$  defined above satisfy (A-8)(i).

*Proof.* We only prove this result under Condition (H-1). The proof is similar under (H-2).

Since for all  $t \in \mathsf{C}$ ,  $y \mapsto j(t)h(t; y)$  is a density with respect to  $\nu$ , we have

$$j(t) = \left( \int h(t; y)\nu(dy) \right)^{-1} > 0.$$

Thus  $j$  is nonincreasing on  $\mathbb{C}$ . Clearly, the defined  $\alpha$  takes values on  $[0, 1]$  and  $\phi$  defines a function from  $\mathbb{X}^2$  to  $\mathbb{X}$ . For all  $(x, x') \in \mathbb{X}^2$  and  $y \in \mathbb{Y}$ , we have

$$\begin{aligned}
 g(x; y) &= \int_{\mathbb{S}} j(x_s) h(x_s; y) \mu(ds) \\
 &\geq \int_{\mathbb{S}} j(x_s \vee x'_s) h(x_s \wedge x'_s; y) \mu(ds) \\
 &\geq \int_{\mathbb{S}} \frac{j(x_s \vee x'_s)}{j(x_s \wedge x'_s)} j(x_s \wedge x'_s) h(x_s \wedge x'_s; y) \mu(ds) \\
 &\geq \int_{\mathbb{S}} \inf_{s \in \mathbb{S}} \left\{ \frac{j(x_s \vee x'_s)}{j(x_s \wedge x'_s)} \right\} j(x_s \wedge x'_s) h(x_s \wedge x'_s; y) \mu(ds) \\
 &= \alpha(x, x') g(\phi(x, x'); y).
 \end{aligned}$$

By symmetry of  $\alpha$  and  $\phi$ , we get (4.20) and thus (A-8)(i) holds.  $\square$

## 4.4 Examples

Let us now apply these results to prove the convergence of MLE of Examples 4.1.1, 4.1.2 and 4.1.3.

### 4.4.1 NBIN-GARCH Model

Example 4.1.1 is a specific case of Definition 4.2.1 where  $\nu$  is the counting measure on  $\mathbb{Y} = \mathbb{N}$ ,

$$\psi_y^\theta(x) = \omega + ax + by, \quad (4.23)$$

$$g^\theta(x; y) = \frac{\Gamma(y+r)}{y! \Gamma(r)} \left( \frac{1}{1+x} \right)^r \left( \frac{x}{1+x} \right)^y, \quad (4.24)$$

with  $\theta = (\omega, a, b, r)$  in a compact subset  $\Theta$  of  $(0, \infty)^4$  and  $\mathbb{X} = (0, \infty)$ .

In [Zhu, 2011, Theorem 1], the equation satisfied by the mean of the observations  $\mu_k = \mathbb{E}[Y_k]$  is derived and is shown to admit a constant solution if and only if

$$rb + a < 1. \quad (4.25)$$

This clearly implies that this condition is necessary to have a stationary solution  $\{Y_k\}$  with finite mean. However it does not imply the existence of such a solution. In fact, the following result shows that (4.25) is indeed a necessary and sufficient condition to have a stationary solution  $\{Y_k\}$  with finite mean. It also shows that all the assumptions of Theorem 4.3.2 hold,

which, with [Remark 4.3.4](#), provides the consistency of the MLE  $\hat{\theta}_{x_1, n}$  for any  $x_1 \in \mathsf{X}$ .

**Theorem 4.4.1.** *Suppose that all  $\theta = (\omega, a, b, r)$  in  $\Theta$  satisfy Condition [\(4.25\)](#). Then Assumptions [\(A-1\)](#), [\(A-2\)](#), [\(B-2\)](#), [\(B-3\)](#) and [\(B-4\)](#) hold with  $\bar{V}$  defined as the identity function on  $\mathsf{X}$  and with any  $x_1 \in \mathsf{X}$ .*

*Proof.* For convenience, we divide the proof into two steps.

**Step 1.** We first prove Assumptions [\(A-1\)](#) and [\(A-2\)](#) by applying [Theorem 4.3.5](#). We set  $\bar{V}(x) = V(x) = x$  and thus we only need to check [\(A-3\)](#), [\(A-4\)](#), [\(A-5\)](#), [\(A-6\)](#), [\(A-7\)](#) and [\(A-8\)](#). Condition [\(A-3\)](#) holds. We have for all  $\theta \in \Theta$ ,

$$RV(x) = \omega + (a + br)x = (a + br)V(x) + \omega,$$

which yields [\(A-4\)](#). The fact that the kernel  $R$  is weak Feller easily follows by observing that, as  $p \rightarrow p'$ ,  $\mathcal{NB}(r, p)$  converges weakly to  $\mathcal{NB}(r, p')$ , so [\(A-5\)](#) holds.

We now prove [\(A-6\)](#). Let  $x_\infty = \omega/(1-a)$ . Let  $x \in \mathbb{R}$  and define recursively the sequence  $x_0 = x, x_k = \omega + ax_{k-1}$  for all positive integers  $k$ . Since  $0 < a < 1$ , this sequence converges to the fixed point  $x_\infty$ . Therefore, defining  $\mathbb{P}_x$  as in [\(4.21\)](#), for any neighborhood  $\mathcal{N}$  of  $x_\infty$ , there exists some  $n$  such that  $x_n \in \mathcal{N}$  and we have

$$\begin{aligned} R^n(x; \mathcal{N}) &= \bar{\mathbb{P}}_x(X_n \in \mathcal{N}) \geq \bar{\mathbb{P}}_x(X_k = x_k \text{ for all } k = 1, \dots, n) \\ &= \bar{\mathbb{P}}_x(Y_0 = \dots = Y_{n-1} = 0) > 0. \end{aligned}$$

So [\(A-6\)](#) holds. Assumption [\(A-7\)](#) holds since we have for all  $(x, x', y) \in \mathsf{X}^2 \times \mathsf{Y}$  with  $x \neq x'$ ,

$$\frac{|\psi_y(x) - \psi_y(x')|}{|x - x'|} = a < 1.$$

To prove [\(A-8\)](#), we apply [Lemma 4.3.6](#) with  $\mathsf{C} = \mathsf{X}, \mathsf{S} = \{1\}$  (so  $\mu$  boils down to the Dirac measure on  $\{1\}$ ). For all  $(x, y) \in \mathsf{X} \times \mathsf{Y}$ , let  $j(x) = \left(\frac{1}{1+x}\right)^r$  and  $h(x; y) = \frac{\Gamma(y+r)}{y!\Gamma(r)} \left(\frac{x}{1+x}\right)^y$ . Indeed,  $h$  satisfies [\(H-1\)](#). Thus by [Lemma 4.3.6](#), for all  $(x, x') \in \mathsf{X}^2$  and  $y \in \mathsf{Y}$ , we get that

$$\alpha(x, x') = \left(\frac{1 + x \wedge x'}{1 + x \vee x'}\right)^r \in (0, 1] \quad \text{and} \quad \phi(x, x') = x \wedge x'$$

satisfy [\(A-8\)\(i\)](#). For any given  $r > 0$ , let a function  $W : \mathsf{X}^2 \rightarrow [1, \infty)$  be defined by, for all  $(x, x') \in \mathsf{X}^2$ ,  $W(x, x') = 1 \vee r$ . By definition of  $W$ , as a constant function, [\(A-8\)\(ii\)](#) and [\(A-8\)\(iv\)](#) clearly hold. Moreover, [\(A-8\)\(iii\)](#) holds since for all  $(x, x') \in \mathsf{X}^2$ , we have that

$$1 - \alpha(x, x') \leq (1 \vee r)|x - x'| = W(x, x')|x - x'|.$$

Therefore, (A-8) holds, which completes **Step 1**.

**Step 2.** We now prove (B-2), (B-3) and (B-4). By assumption on  $\Theta$ , then there exists  $(\underline{\omega}, \bar{\omega}, \underline{b}, \bar{b}, \underline{r}, \bar{r}, \underline{\alpha}, \bar{\alpha}) \in (0, \infty)^6 \times (0, 1)^2$  such that

$$\underline{\omega} \leq \omega \leq \bar{\omega}, \quad \underline{b} \leq b \leq \bar{b}, \quad \underline{r} \leq r \leq \bar{r}, \quad \underline{\alpha} \leq a + br \leq \bar{\alpha}.$$

Clearly, (B-2) and (B-3) hold by definitions of  $\psi_y^\theta(x)$  and  $g^\theta(x; y)$ . It remains to check (B-4) for a well-chosen closed subset  $\mathbf{X}_1$  and any  $x_1 \in \mathbf{X}$ . Let  $\mathbf{X}_1 = [\underline{\omega}, \infty) \subset \mathbf{X}$  so that (B-4)(i) holds. By noting that for all  $(\theta, x, y) \in \Theta \times \mathbf{X} \times \mathbf{Y}$ ,  $g^\theta(x; y) \leq 1$ , we have (B-4)(ii). From (4.9) and (4.23), we have for all  $s \leq t$ ,  $y_{s:t} \in \mathbf{Y}^{t-s+1}$ ,  $x \in \mathbf{X}$  and  $\theta \in \Theta$ ,

$$\psi^\theta \langle y_{s:t} \rangle (x) = \omega \left( \frac{1 - a^{t-s+1}}{1 - a} \right) + a^{t-s+1}x + b \sum_{j=0}^{t-s} a^j y_{t-j}. \quad (4.26)$$

Using (4.26), we have, for all  $\theta \in \Theta$ ,  $x \in \mathbf{X}$  and  $y_{1:n} \in \mathbf{Y}^n$ ,

$$|\psi^\theta \langle y_{1:n} \rangle (x_1) - \psi^\theta \langle y_{1:n} \rangle (x)| = a^n |x_1 - x| \leq \bar{\alpha}^n |x_1 - x|.$$

This gives (B-4)(iii) and (B-4)(iv) by setting  $\varrho = \bar{\alpha} < 1$  and  $\bar{\psi}(x) = |x_1 - x|$ . Next we set  $\bar{\phi}$ ,  $H$  and  $C$  to meet Conditions (B-4)(v) and (B-4)(vi) and (B-4)(vii). Let us write, for all  $\theta \in \Theta$  and  $y \in \mathbf{Y}$ ,

$$|x_1 - \psi_y^\theta(x_1)| \leq \omega + (1 + a)x_1 + by \leq \bar{\omega} + (1 + \bar{\alpha})x_1 + \bar{b}y$$

and, for all  $(x, x') \in \mathbf{X}_1^2 = [\underline{\omega}, \infty)^2$ ,

$$\begin{aligned} |\ln g^\theta(x; y) - \ln g^\theta(x'; y)| &= |(r + y) [\ln(1 + x') - \ln(1 + x)] + y [\ln x - \ln x']| \\ &\leq [(r + y)(1 + \underline{\omega})^{-1} + y \underline{\omega}^{-1}] |x - x'| \\ &\leq [\bar{r} + y(1 + \underline{\omega}^{-1})] |x - x'|. \end{aligned}$$

Setting  $\bar{\phi}(y) = \bar{\omega} \vee \bar{r} + (1 + \bar{\alpha})x_1 + (\bar{b} \vee (1 + \underline{\omega}^{-1}))y$ ,  $H(x) = x$  and  $C = 0$  then yield Conditions (B-4)(v), (B-4)(vi) and (B-4)(vii). Now (B-4)(viii) follows from

$$\int \ln^+ y G^\theta(x; dy) \leq \int y G^\theta(x; dy) = rx \leq \bar{r} \bar{V}(x).$$

This concludes the proof.  $\square$

#### 4.4.2 NM-GARCH Model

The NM( $d$ )-GARCH(1, 1) of Example 4.1.2 is a specific case of Definition 4.2.1 where  $\mathbf{X} = \mathbb{R}_+^d$  and  $\nu$  is the Lebesgue measure on  $\mathbf{Y} = \mathbb{R}$ ,

$$\psi_y^\theta(\mathbf{x}) = \boldsymbol{\omega} + \mathbf{A}\mathbf{x} + y^2 \mathbf{b}, \quad (4.27)$$

$$g^\theta(\mathbf{x}; y) = \sum_{\ell=1}^d \gamma_\ell \frac{e^{-y^2/2x_\ell}}{(2\pi x_\ell)^{1/2}}, \quad (\mathbf{x}, y) \in \mathbf{X} \times \mathbf{Y}, \quad (4.28)$$



and  $\theta = (\boldsymbol{\gamma}, \boldsymbol{\omega}, \mathbf{A}, \mathbf{b}) \in \Theta$ , a compact subset of  $\mathbf{P}_d \times (0, \infty)^d \times \mathbb{R}_+^{d \times d} \times \mathbb{R}_+^d$ , with  $\mathbf{P}_d$  defined by (4.3).

In Haas et al. [2004], it is shown that the equation satisfied by the variance of a univariate NM( $d$ )-GARCH(1, 1) process admits a constant solution if and only if

$$|\lambda|_{\max}(\mathbf{A} + \mathbf{b}\boldsymbol{\gamma}^T) < 1, \quad (4.29)$$

where, for any square matrix  $\mathbf{M}$ ,  $|\lambda|_{\max}(\mathbf{M})$  denotes the spectral radius of  $\mathbf{M}$ . It follows that the existence of a weakly stationary solution implies (4.29) but it does not say anything about the existence of stationary or weakly stationary solution. The result below shows that (4.29) is indeed a sufficient condition for the existence of a stationary solution with finite variance. It moreover provides with Theorem 4.3.2 and Remark 4.3.4 the consistency of the MLE  $\hat{\theta}_{\mathbf{x}_1, n}$  for any  $\mathbf{x}_1 \in \mathbf{X}$ .

**Theorem 4.4.2.** *Suppose that all  $\theta = (\boldsymbol{\gamma}, \boldsymbol{\omega}, \mathbf{A}, \mathbf{b})$  in  $\Theta$  satisfy (4.29). Then Assumptions (A-1), (A-2), (B-2), (B-3) and (B-4) hold with  $\bar{V}$  defined as any norm on  $\mathbf{X}$ .*

*Proof.* . In this proof section, we set

$$\bar{V}(\mathbf{x}) = |\mathbf{x}| = \sum_{\ell=1}^d |x_\ell|, \quad (4.30)$$

for all  $\mathbf{x} = (x_\ell) \in \mathbf{X}$ . As in Theorem 4.4.1, we divide the proof into two steps.

**Step 1.** We first show that Assumptions (A-1) and (A-2) hold with the above  $\bar{V}$  by applying Theorem 4.3.5. Define  $V$  on  $\mathbf{X}$  by setting

$$V(\mathbf{x}) = (\mathbf{1} + \mathbf{x}_0)^T \mathbf{x},$$

where  $\mathbf{1}$  is the vector of  $\mathbf{X}$  with all entries equal to 1 and  $\mathbf{x}_0$  is defined by

$$\mathbf{1} + \mathbf{x}_0 = (\mathbf{I} - (\mathbf{A} + \mathbf{b}\boldsymbol{\gamma}^T)^T)^{-1} \mathbf{1}.$$

We indeed note that by Condition (4.29) the above inversion is well defined and moreover

$$(\mathbf{I} - (\mathbf{A} + \mathbf{b}\boldsymbol{\gamma}^T)^T)^{-1} = \mathbf{I} + \sum_{k \geq 1} (\mathbf{A}^T + \boldsymbol{\gamma}\mathbf{b}^T)^k,$$

and, since  $\mathbf{A}$ ,  $\mathbf{b}$ ,  $\boldsymbol{\gamma}$  all have non-negative entries, it follows that  $\mathbf{x}_0$  has non-negative entries. Thus, for all  $\mathbf{x} = (x_\ell) \in \mathbf{X}$ ,

$$\bar{V}(\mathbf{x}) = \mathbf{1}^T \mathbf{x} \leq V(\mathbf{x}),$$

so that  $\bar{V} \lesssim V$ . Hence by [Theorem 4.3.5](#), we thus only need to check [\(A-3\)](#), [\(A-4\)](#), [\(A-5\)](#), [\(A-6\)](#), [\(A-7\)](#) and [\(A-8\)](#) with  $V$  defined as above for a given  $\theta = (\boldsymbol{\gamma}, \boldsymbol{\omega}, \mathbf{A}, \mathbf{b}) \in \Theta$  (so we drop  $\theta$  in the notation in the remaining of **Step 1**). Condition [\(A-3\)](#) holds for any metric  $d_X$  associated to a norm on the finite dimensional space  $\mathbf{X}$ . (The precise choice of  $d$  is postponed to the verification of [\(A-7\)](#).) We have

$$\begin{aligned} RV(\mathbf{x}) &= \int V(\boldsymbol{\omega} + \mathbf{A}\mathbf{x} + y^2\mathbf{b}) G(\mathbf{x}, dy) \\ &= (\mathbf{1} + \mathbf{x}_0)^T \boldsymbol{\omega} + (\mathbf{1} + \mathbf{x}_0)^T (\mathbf{A} + \mathbf{b}\boldsymbol{\gamma}^T) \mathbf{x} \\ &= V(\boldsymbol{\omega}) + \mathbf{1}^T (\mathbf{I} - (\mathbf{A} + \mathbf{b}\boldsymbol{\gamma}^T))^{-1} (\mathbf{A} + \mathbf{b}\boldsymbol{\gamma}^T - \mathbf{I} + \mathbf{I}) \mathbf{x} \\ &= V(\boldsymbol{\omega}) + \mathbf{x}_0^T \mathbf{x} \\ &\leq V(\boldsymbol{\omega}) + \lambda V(\mathbf{x}), \end{aligned}$$

where we set  $\lambda = \max_\ell \{\mathbf{x}_{0,\ell}/(1 + \mathbf{x}_{0,\ell})\} < 1$ . Hence [\(A-4\)](#) holds. Condition [\(A-5\)](#) easily follows from the continuity of the Gaussian distribution with respect to its variance parameter. We now prove [\(A-6\)](#). From [\(4.9\)](#) and [\(4.27\)](#), we have for all  $n \geq 1$ ,  $y_{0:n-1} \in \mathbf{Y}^n$  and  $\mathbf{x} \in \mathbf{X}$ ,

$$\psi^\theta \langle y_{0:n-1} \rangle (\mathbf{x}) = \mathbf{A}^n \mathbf{x} + \sum_{j=0}^{n-1} \mathbf{A}^j (\boldsymbol{\omega} + y_{n-1-j}^2 \mathbf{b}). \quad (4.31)$$

Let us use the norm

$$\|\mathbf{M}\| = \max_j \sum_i |\mathbf{M}_{i,j}| = \sup_{|\mathbf{x}| \leq 1} |\mathbf{M}\mathbf{x}|$$

on  $d \times d$  matrices. Note that by [\(4.29\)](#), there exists  $\delta \in (0, 1)$  and  $c > 0$  such that, for any  $k \geq 1$ ,

$$\left\| (\mathbf{A} + \mathbf{b}\boldsymbol{\gamma}^T)^k \right\| \leq c \delta^k. \quad (4.32)$$

Using that  $\mathbf{A}$ ,  $\mathbf{b}$ ,  $\boldsymbol{\gamma}$  all have nonnegative entries, we have

$$\|\mathbf{A}^k\| \leq \left\| (\mathbf{A} + \mathbf{b}\boldsymbol{\gamma}^T)^k \right\|. \quad (4.33)$$

Hence  $(\mathbf{I} - \mathbf{A})^{-1} = \mathbf{I} + \sum_{k \geq 1} \mathbf{A}^k$  is well defined and we set  $\mathbf{x}_\infty = (\mathbf{I} - \mathbf{A})^{-1} \boldsymbol{\omega}$  so that, with [\(4.27\)](#), we have

$$\psi^\theta \langle y_{0:n-1} \rangle (\mathbf{x}) - \mathbf{x}_\infty = \mathbf{A}^n \mathbf{x} + \sum_{j \geq n} \mathbf{A}^j \boldsymbol{\omega} + \sum_{j=0}^{n-1} y_{n-1-j}^2 \mathbf{A}^j \mathbf{b}.$$

Then, using definition (4.21), we get that,  $\bar{\mathbb{P}}_{\mathbf{x}}$ -a.s., for all  $n \geq 1$ ,

$$\begin{aligned} |\mathbf{X}_n - \mathbf{x}_\infty| &= |\psi \langle Y_{0:n-1} \rangle(\mathbf{x}) - \mathbf{x}_\infty| \\ &\leq |\mathbf{A}^n(\mathbf{x} - \mathbf{x}_\infty)| + \sum_{j \geq n} |\mathbf{A}^j \boldsymbol{\omega}| + \left( \max_{0 \leq j \leq n-1} Y_j^2 \right) \sum_{j=0}^{n-1} |\mathbf{A}^j \mathbf{b}|. \end{aligned}$$

With (4.32) and (4.33), this implies

$$\bar{\mathbb{P}}_{\mathbf{x}} \left( |\mathbf{X}_n - \mathbf{x}_\infty| \leq c \left[ \delta^n \left( |\mathbf{x} - \mathbf{x}_\infty| + \frac{|\boldsymbol{\omega}|}{1 - \delta} \right) + \frac{|\mathbf{b}|}{1 - \delta} \max_{0 \leq j \leq n-1} Y_j^2 \right] \right) = 1.$$

To obtain (A-6), it is sufficient to observe that, since  $g$  takes positive values in (4.28), for any positive  $\epsilon$ ,  $\mathbf{x} \in \mathbf{X}$  and any  $n \geq 1$ ,

$$\bar{\mathbb{P}}_{\mathbf{x}} \left( \max_{0 \leq j \leq n-1} Y_j^2 < \epsilon \right) > 0.$$

Next we prove (A-7). We have

$$\psi_y(\mathbf{x}) - \psi_y(\mathbf{x}') = \mathbf{A}(\mathbf{x} - \mathbf{x}').$$

Since (4.32) and (4.33) imply that  $|\lambda|_{\max}(\mathbf{A}) < 1$ , there exists a vector norm which makes  $\mathbf{A}$  strictly contracting. Choosing the metric  $d_{\mathbf{X}}$  on  $\mathbf{X}$  as the one derived from this norm, we get (A-7). To show (A-8), we again rely on Lemma 4.3.6. Let us set  $\mathbf{C} = (0, \infty)$  and  $\mathbf{S} = \{1, \dots, d\}$  and define the probability measure  $\mu$  on  $\mathbf{S}$  by  $\mu(\{s\}) = \gamma_s$ , for all  $s \in \mathbf{S}$ . For all  $(t, y) \in \mathbf{C} \times \mathbf{Y}$ , let  $j(t) = \frac{1}{(2\pi t)^{1/2}}$  and  $h(t; y) = \exp(-y^2/2t)$ . Obviously, Relation (4.22) holds and  $h$  satisfies (H-1). Hence, Lemma 4.3.6 implies that  $\alpha$  and  $\phi$  defined respectively for all  $\mathbf{x} = (x_1, \dots, x_d)$ ,  $\mathbf{x}' = (x'_1, \dots, x'_d) \in \mathbf{X}$  by

$$\alpha(\mathbf{x}, \mathbf{x}') = \min_{1 \leq \ell \leq d} \left\{ \left( \frac{x_\ell \wedge x'_\ell}{x_\ell \vee x'_\ell} \right)^{\frac{1}{2}} \right\} \in (0, 1] \text{ and } \phi(\mathbf{x}, \mathbf{x}') = (x_1 \wedge x'_1, \dots, x_d \wedge x'_d),$$

satisfy (A-8)(i). For  $\mathbf{x} = (x_1, \dots, x_d)$ ,  $\mathbf{x}' = (x'_1, \dots, x'_d) \in \mathbf{X}$ , we have

$$\begin{aligned} 1 - \alpha(\mathbf{x}, \mathbf{x}') &= 1 - \min_{1 \leq \ell \leq d} \left\{ \left( 1 - \frac{|x_\ell - x'_\ell|}{x_\ell \vee x'_\ell} \right)^{\frac{1}{2}} \right\} \\ &\leq \max_{1 \leq \ell \leq d} \left\{ \frac{|x_\ell - x'_\ell|}{x_\ell \vee x'_\ell} \right\} \\ &\leq \min_{1 \leq \ell \leq d} (x_\ell^{-1} \wedge x'^{-1}_\ell) |\mathbf{x} - \mathbf{x}'| \\ &\leq W(\mathbf{x}, \mathbf{x}') d_{\mathbf{X}}(\mathbf{x}, \mathbf{x}'), \end{aligned}$$

where  $d_{\mathbf{X}}$  is the metric previously defined and  $W$  is defined by  $W(\mathbf{x}, \mathbf{x}') = 1 \vee (c_{d_{\mathbf{X}}} \min_{1 \leq \ell \leq d} (x_{\ell}^{-1} \wedge x'_{\ell}{}^{-1}))$  with  $c_{d_{\mathbf{X}}} > 0$  is conveniently chosen (such a constant exists since  $d_{\mathbf{X}}$  is the metric associated to a norm and  $\mathbf{X}$  has finite dimension). Then (A-8)(ii) and (A-8)(iii) hold and, since for all  $y \in \mathbf{Y}$  and  $x \in \mathbf{X}$ ,  $\psi_y(\mathbf{x})$  has all its entries bounded from below by the positive entries of  $\boldsymbol{\omega}$ ,  $W(\psi_y(\mathbf{x}), \psi_y(\mathbf{x}'))$  is uniformly bounded over  $(\mathbf{x}, \mathbf{x}', y) \in \mathbf{X} \times \mathbf{X} \times \mathbf{Y}$  and (A-8)(iv) holds. This completes **Step 1**.

**Step 2** We now show that Assumptions (B-2), (B-3) and (B-4) hold.

Clearly, (B-2) and (B-3) hold by definitions of  $\psi_y^{\theta}(\mathbf{x})$  and  $g^{\theta}(\mathbf{x}; y)$ . It remains to show (B-4). Since  $\Theta$  is compact, then

$$\underline{\omega} \leq \min_{1 \leq \ell \leq d} \omega_{\ell}, \quad |\boldsymbol{\omega}| \leq \bar{\omega}, \quad \underline{b} \leq |\mathbf{b}| \leq \bar{b}, \quad |\lambda|_{\max}(\mathbf{A} + \mathbf{b}\boldsymbol{\gamma}^T) \leq \bar{\rho}, \quad \|\mathbf{A} + \mathbf{b}\boldsymbol{\gamma}^T\| \leq L$$

for some  $(\underline{\omega}, \bar{\omega}, \underline{b}, \bar{b}, \bar{\rho}) \in (0, \infty)^4 \times (0, 1)$  and  $L > 0$ . By [Moulines et al., 2005, Lemma 12], we note that this implies that, for all  $\bar{\delta} \in (\bar{\rho}, 1)$ , there exists  $\bar{C} > 0$  such that for all  $k \geq 1$  and all  $\theta \in \Theta$ ,

$$\|(\mathbf{A} + \mathbf{b}\boldsymbol{\gamma}^T)^k\| \leq \bar{C} \bar{\delta}^k. \quad (4.34)$$

We set  $\mathbf{X}_1 = [\underline{\omega}, \infty)^d \subset \mathbf{X}$  so that (B-4)(i) holds. Moreover, for all  $(\theta, \mathbf{x}, y) \in \Theta \times \mathbf{X}_1 \times \mathbf{Y}$ ,  $g^{\theta}(\mathbf{x}; y) \leq (2\pi\underline{\omega})^{-1/2}$ . Thus, Condition (B-4)(ii) holds. Now let  $x_1 \in \mathbf{X}$ . Using (4.31), (4.34) and (4.33), we have, for all  $\mathbf{x} \in \mathbf{X}$ ,  $y_{1:n} \in \mathbf{Y}^n$  and  $\theta \in \Theta$ ,

$$\begin{aligned} |\psi^{\theta}\langle y_{1:n} \rangle(\mathbf{x}_1) - \psi^{\theta}\langle y_{1:n} \rangle(\mathbf{x})| &= |\mathbf{A}^n(\mathbf{x}_1 - \mathbf{x})| \\ &\leq \bar{C} \bar{\delta}^n |\mathbf{x}_1 - \mathbf{x}|. \end{aligned}$$

Using that the norm defining  $d_{\mathbf{X}}$  is equivalent to the norm  $|\cdot|$ , we get (B-4)(iii) with

$$\bar{\psi}(\mathbf{x}) = \bar{C}' |\mathbf{x}_1 - \mathbf{x}|,$$

for some positive constant  $\bar{C}'$ . Hence (B-4)(iv) holds and since

$$|\mathbf{x}_1 - \psi_y^{\theta}(\mathbf{x}_1)| \leq (L + 1) |\mathbf{x}_1| + \bar{\omega} + y^2 \bar{b},$$

we also get (B-4)(v) provided that

$$\bar{\phi}(y) \geq (L + 1) |\mathbf{x}_1| + \bar{\omega} + y^2 \bar{b}. \quad (4.35)$$

It is straightforward to show that, for all  $\theta \in \Theta$ ,  $\mathbf{x} \in \mathbf{X}_1$ ,  $y \in \mathbb{R}$ , and  $\ell \in \{1, \dots, d\}$ ,

$$\left| \frac{\partial \ln g^{\theta}}{\partial x_{\ell}}(\mathbf{x}; y) \right| \leq \frac{1}{2} \left( \frac{y^2}{\underline{\omega}^2} + \frac{1}{\underline{\omega}} \right).$$

Thus, by the mean value theorem, for all  $\theta \in \Theta$ ,  $(\mathbf{x}, \mathbf{x}') \in \mathbf{X}_1 \times \mathbf{X}_1$  and  $y \in \mathbf{Y}$ ,

$$|\ln g^\theta(\mathbf{x}; y) - \ln g^\theta(\mathbf{x}'; y)| \leq \frac{1}{2} \left( \frac{y^2}{\underline{\omega}^2} + \frac{1}{\underline{\omega}} \right) |\mathbf{x} - \mathbf{x}'|$$

We thus obtain (B-4)(v), (B-4)(vi) and (B-4)(vii) by setting  $C = 0$ ,

$$H(u) = \sup_{d(\mathbf{x}, \mathbf{x}') \leq u} |\mathbf{x} - \mathbf{x}'|,$$

and

$$\bar{\phi}(y) = (L + 1) |\mathbf{x}_1| + \bar{\omega} + 1/(2\underline{\omega}) + y^2(\bar{b} + \underline{\omega}^2).$$

In addition, for all  $\theta \in \Theta$  and  $\mathbf{x} \in \mathbf{X}$ , we have

$$\int y^2 G^\theta(x; dy) = \gamma^T \mathbf{x}.$$

Hence, using (4.30) with the above definitions, we obtain (B-4)(viii) and the proof is concluded.  $\square$

#### 4.4.3 Threshold IN-GARCH Model

The threshold IN-GARCH(1,1) in Example 4.1.3 is a specific case of Definition 4.2.1 where  $\nu$  is the counting measure on  $\mathbf{Y} = \mathbb{Z}_+$ ,

$$\psi_y^\theta(x) = \omega + ax + by, \quad (4.36)$$

$$g^\theta(x; y) = e^{-(x \wedge \tau)} \frac{(x \wedge \tau)^y}{y!}, \quad (4.37)$$

with  $\theta = (\omega, a, b, \tau)$  in a compact subset  $\Theta$  of  $(0, \infty)^4$  and  $\mathbf{X} = (0, \infty)$ . In this model, if  $a < 1$ , we then have the ergodicity and consistency results as stated in Theorem 4.4.3 below.

**Theorem 4.4.3.** *Suppose that all  $\theta = (\omega, a, b, \tau)$  in  $\Theta$  satisfy  $a < 1$ . Then Assumptions (A-1), (A-2), (B-2), (B-3) and (B-4) hold with  $\bar{V}$  defined as the identity function on  $\mathbf{X}$  and with any  $x_1 \in \mathbf{X}$ .*

*Proof.* . As in the proofs of the two theorems above, for convenience, we divide the proof into two steps.

**Step 1.** We first prove Assumptions (A-1) and (A-2) by applying Theorem 4.3.5. We set  $\bar{V}(x) = V(x) = x$  and thus we only need to check (A-3), (A-4), (A-5), (A-6), (A-7) and (A-8). Condition (A-3) holds with the usual metric on  $\mathbb{R}$ . We have for all  $\theta \in \Theta$ ,

$$RV(x) = \omega + ax + b(x \wedge \tau) \leq aV(x) + (\omega + b\tau),$$

which yields (A-4). The fact that the kernel  $R$  is weak Feller easily follows by observing that, as  $x \rightarrow x'$ ,  $\mathcal{P}(x)$  converges weakly to  $\mathcal{P}(x')$  and the map  $x \mapsto x \wedge \tau$  is continuous, so (A-5) holds.

The proof of (A-6) is similar to the NBIN-GARCH case of Theorem 4.4.1 and is thus omitted. Assumption (A-7) holds since we have for all  $(x, x', y) \in \mathsf{X}^2 \times \mathsf{Y}$  with  $x \neq x'$ ,

$$\frac{|\psi_y(x) - \psi_y(x')|}{|x - x'|} = a < 1.$$

To prove (A-8), we apply Lemma 4.3.6 with  $\mathsf{C} = \mathsf{X}$ ,  $\mathsf{S} = \{1\}$  (so  $\mu$  boils down to the Dirac measure on  $\{1\}$ ). For all  $(x, y) \in \mathsf{X} \times \mathsf{Y}$ , let  $j(x) = e^{-(x \wedge \tau)}$  and  $h(x; y) = \frac{(x \wedge \tau)^y}{y!}$ . Then  $h$  indeed satisfies (H-1). Thus by Lemma 4.3.6, for all  $(x, x') \in \mathsf{X}^2$  and  $y \in \mathsf{Y}$ , we get that

$$\alpha(x, x') = e^{-(x \vee x') \wedge \tau + (x \wedge x') \wedge \tau} \in (0, 1] \quad \text{and} \quad \phi(x, x') = x \wedge x'$$

satisfy (A-8)(i).

Let  $W(x, x') = 1$  for all  $(x, x') \in \mathsf{X}^2$ , which is a constant function. Thus (A-8)(ii) and (A-8)(iv) clearly hold. Moreover, (A-8)(iii) holds since for all  $(x, x') \in \mathsf{X}^2$ , we have that

$$1 - \alpha(x, x') \leq x \vee x' - x \wedge x' = |x - x'| = W(x, x')|x - x'|.$$

Therefore, (A-8) holds, which completes **Step 1**.

**Step 2.** We now prove (B-2), (B-3) and (B-4). By assumption on  $\Theta$ , then there exists  $(\underline{\omega}, \bar{\omega}, \underline{b}, \bar{b}, \underline{\tau}, \bar{\tau}, \underline{\alpha}, \bar{\alpha}) \in (0, \infty)^6 \times (0, 1)^2$  such that

$$\underline{\omega} \leq \omega \leq \bar{\omega}, \quad \underline{b} \leq b \leq \bar{b}, \quad \underline{\tau} \leq \tau \leq \bar{\tau}, \quad \underline{\alpha} \leq a \leq \bar{\alpha}.$$

Clearly, (B-2) and (B-3) hold by definitions of  $\psi_y^\theta(x)$  and  $g^\theta(x; y)$ . It remains to check (B-4) for a well-chosen closed subset  $\mathsf{X}_1$  and any  $x_1 \in \mathsf{X}$ . Let  $\mathsf{X}_1 = [\underline{\omega}, \infty) \subset \mathsf{X}$  so that (B-4)(i) holds. By noting that for all  $(\theta, x, y) \in \Theta \times \mathsf{X} \times \mathsf{Y}$ ,  $g^\theta(x; y) \leq 1$ , we have (B-4)(ii). From (4.9) and (4.36), we have for all  $s \leq t$ ,  $y_{s:t} \in \mathsf{Y}^{t-s+1}$ ,  $x \in \mathsf{X}$  and  $\theta \in \Theta$ ,

$$\psi^\theta \langle y_{s:t} \rangle(x) = \omega \left( \frac{1 - a^{t-s+1}}{1 - a} \right) + a^{t-s+1}x + b \sum_{j=0}^{t-s} a^j y_{t-j}. \quad (4.38)$$

Using (4.38), we have, for all  $\theta \in \Theta$ ,  $x \in \mathsf{X}$  and  $y_{1:n} \in \mathsf{Y}^n$ ,

$$|\psi^\theta \langle y_{1:n} \rangle(x_1) - \psi^\theta \langle y_{1:n} \rangle(x)| = a^n |x_1 - x| \leq \bar{\alpha}^n |x_1 - x|.$$

This gives (B-4)(iii) and (B-4)(iv) by setting  $\rho = \bar{\alpha} < 1$  and  $\bar{\psi}(x) = |x_1 - x|$ . Next we set  $\bar{\phi}$ ,  $H$  and  $C$  to meet Conditions (B-4)(v) and (B-4)(vi) and (B-4)(vii). Let us write, for all  $\theta \in \Theta$  and  $y \in \mathcal{Y}$ ,

$$|x_1 - \psi_y^\theta(x_1)| \leq \omega + (1 + a)x_1 + by \leq \bar{\omega} + (1 + \bar{\alpha})x_1 + \bar{b}y$$

and, for all  $(x, x') \in \mathcal{X}_1^2 = [\underline{\omega}, \infty)^2$ ,

$$\begin{aligned} |\ln g^\theta(x; y) - \ln g^\theta(x'; y)| &= |(x' \wedge \tau - x \wedge \tau) + y(\ln(x \wedge \tau) - \ln(x' \wedge \tau))| \\ &\leq (1 + (\underline{\omega} \wedge \underline{\tau})^{-1}y) |x - x'|. \end{aligned}$$

Setting  $\bar{\phi}(y) = 1 + \bar{\omega} + (1 + \bar{\alpha})x_1 + (\bar{b} \vee (\underline{\omega} \wedge \underline{\tau})^{-1})y$ ,  $H(x) = x$  and  $C = 0$  then yield Conditions (B-4)(v), (B-4)(vi) and (B-4)(vii). Now (B-4)(viii) follows from

$$\int \ln^+ y G^\theta(x; dy) \leq \int y G^\theta(x; dy) = x \wedge \tau \leq \bar{V}(x).$$

This concludes the proof.  $\square$

## 4.5 Numerical Experiments

### 4.5.1 Numerical Procedure

In this part, we provide a numerical method for computing the (conditional) MLE  $\hat{\theta}_{x,n}$  for the parameter  $\theta = (\omega, a, b, r)$  in the NBIN-GARCH(1, 1) model introduced in [Example 4.1.1](#) and studied in [Section 4.4.1](#). It is convenient to write  $\theta = (\vartheta, r)$  with  $\vartheta = (\omega, a, b)$  and then to write  $\psi_y^\vartheta(x)$  and  $g^r(x; y)$  instead of  $\psi_y^\theta(x)$  and  $g^\theta(x; y)$  in [\(4.23\)](#) and [\(4.24\)](#), respectively. In contrast to the approach used in [Zhu \[2011\]](#), we allow the component  $r$  to be any positive real number, rather than a discrete one and to be unknown as well. We thus maximize jointly with respect to the parameters  $\vartheta$  and  $r$  the log-likelihood function  $\mathbb{L}_{x,n}^\theta \langle y_{1:n} \rangle = \mathbb{L}_{x,n}^{(\vartheta,r)} \langle y_{1:n} \rangle$ . In practice, one does not rely on a compact set  $\Theta$  of parameters as in [Theorem 4.4.1](#). Instead the maximization is performed over all parameters  $\omega > 0$ ,  $a > 0$ ,  $b > 0$ ,  $r > 0$  such that the stability constraint  $a + br < 1$  holds (taken from [\(4.25\)](#)). We use the constrained nonlinear optimization function `auglag` (Augmented Lagrangian Minimization Algorithm) from the package `alabama` (Augmented Lagrangian Adaptive Barrier Minimization Algorithm) in R. For this purpose we provide an initial parameter point and a numerical computation of the normalized log-likelihood function  $\mathbb{L}_{x,n}^\theta \langle y_{1:n} \rangle$  and of its gradient. The initial point is obtained by applying a conditional least square (CLS) estimation based on an ARMA(1, 1) representation of the model; see [\[Zhu, 2011, Section 3\]](#). The

computation of the log-likelihood and of its derivatives are derived as follows. For all  $x \in \mathbf{X}$ , denoting  $u_k^\vartheta = \psi^\vartheta \langle y_{1:k-1} \rangle (x)$  for all  $k \geq 2$  and  $u_1^\vartheta = x$ , we have

$$\begin{aligned} \mathbf{L}_{x,n}^{(\vartheta,r)} \langle y_{1:n} \rangle &= n^{-1} \sum_{k=1}^n \ln g^r (\psi^\vartheta \langle y_{1:k-1} \rangle (x); y_k) \\ &= n^{-1} \ln g^r (x, y_1) + n^{-1} \sum_{k=2}^n \ln g^r (u_k^\vartheta; y_k). \end{aligned}$$

The computation of  $u_k^\vartheta$  for all  $k \geq 2$  is done iteratively by observing that  $u_k^\vartheta = \psi^\vartheta \langle y_{k-1} \rangle (u_{k-1}^\vartheta)$  and the computation of  $\mathbf{L}_{x,n}^{(\vartheta,r)} \langle y_{1:n} \rangle$  is deduced. The computation of the derivatives with respect to parameter  $\theta = (\vartheta, r)$  of the function  $\mathbf{L}_{x,n}^{(\vartheta,r)} \langle y_{1:n} \rangle$  are then obtained in two steps. First, for  $k \geq 2$ , the derivative of  $u_k^\vartheta$  with respect to  $\vartheta$  are obtained iteratively by  $\partial u_1^\vartheta / \partial \vartheta = 0$  and

$$\frac{\partial u_k^\vartheta}{\partial \vartheta} = (1, u_{k-1}^\vartheta, Y_{k-1}) + a \frac{\partial u_{k-1}^\vartheta}{\partial \vartheta}.$$

Then the derivatives of  $\mathbf{L}_{x,n}^{(\vartheta,r)} \langle y_{1:n} \rangle$  with respect to  $\vartheta$  and  $r$  are given by

$$\begin{aligned} \frac{\partial \mathbf{L}_{x,n}^{(\vartheta,r)}}{\partial \vartheta} &= n^{-1} \sum_{k=1}^n \frac{\partial \ln g^r}{\partial x} (u_k^\vartheta; y_k) \frac{\partial u_k^\vartheta}{\partial \vartheta} \\ &= n^{-1} \sum_{k=2}^n \left( \frac{y_k}{u_k^\vartheta} - \frac{y_k + r}{1 + u_k^\vartheta} \right) \frac{\partial u_k^\vartheta}{\partial \vartheta} \end{aligned}$$

and

$$\begin{aligned} \frac{\partial \mathbf{L}_{x,n}^{(\vartheta,r)}}{\partial r} &= n^{-1} \sum_{k=1}^n \frac{\partial \ln g^r}{\partial r} (u_k^\vartheta; y_k) \\ &= n^{-1} \sum_{k=1}^n (\Gamma_2(r + y_k) - \ln(1 + u_k^\vartheta)) - \Gamma_2(r), \end{aligned}$$

respectively, where  $\Gamma_2$  is the digamma function  $\Gamma_2(r) = \frac{d}{dr} \ln \Gamma(r)$ ,  $r > 0$ .

### 4.5.2 Simulation Study

We consider two NBIN-GARCH(1, 1) models with parameters:

$$(M.1) \quad \theta_\star = (\omega_\star, a_\star, b_\star, r_\star) = (3, .2, .2, 2) \text{ and}$$

$$(M.2) \quad \theta_\star = (\omega_\star, a_\star, b_\star, r_\star) = (3, .35, .1, 1.5).$$



---

4. OBSERVATION-DRIVEN MODELS

---

We simulated  $m = 200$  data sets for each sample size  $n = 2^7, 2^8, 2^9$  and  $2^{10}$ . In [Figure 4.2](#), we display the obtained boxplots of the difference of the normalized log-likelihood functions evaluated respectively at MLE and at the true value  $\theta_*$ . As predicted by the theory, this difference appears to converge to 0 as the number of observations  $n \rightarrow \infty$ . For the NBIN-GARCH(1, 1) model, it can be shown that  $\Theta_* = \{\theta_*\}$ , which implies the convergence of the MLE to the true parameter. We can observe this behavior for each component of the MLE for the two models in [Figure 4.3](#) and [Figure 4.4](#). We also report the Monte Carlo mean along with the mean absolute deviation error (MADE):  $\text{MADE} = m^{-1} \sum_{j=1}^m |\hat{\theta}_{x,n}^j - \theta_*^j|$  as an evaluation criterion for the estimated parameter in [Table 4.1](#).

Table 4.1 – Mean of estimates, MADEs (within parentheses) for the NBIN-GARCH(1, 1) models

Model	Parameter	Sample size $n$			
		$n = 2^7$	$n = 2^8$	$n = 2^9$	$n = 2^{10}$
(M.1)	$\hat{\omega}$	3.311(.973)	3.212(.719)	3.108(.507)	3.062(.372)
	$\hat{a}$	.165(.138)	.173(.113)	.187(.076)	.193(.055)
	$\hat{b}$	.194(.049)	.195(.034)	.197(.025)	.200(.018)
	$\hat{r}$	2.045(.241)	2.035(.166)	2.020(.112)	2.011(.074)
(M.2)	$\hat{\omega}$	3.525(1.325)	3.362(1.258)	3.326(1.041)	3.167(.761)
	$\hat{a}$	.252(.227)	.290(.213)	.296(.170)	.319(.136)
	$\hat{b}$	.092(.056)	.097(.039)	.098(.028)	.100(.022)
	$\hat{r}$	1.563(.175)	1.539(.129)	1.520(.093)	1.513(.066)

#### 4. OBSERVATION-DRIVEN MODELS

---

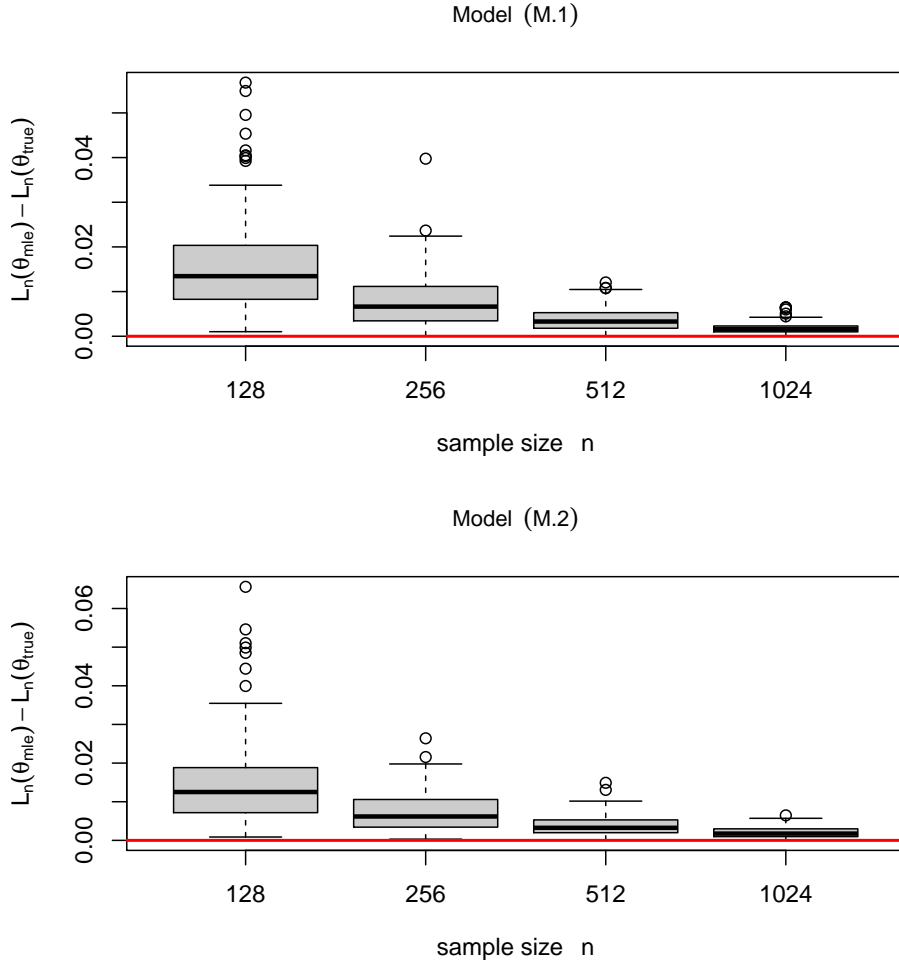


Figure 4.2 – Boxplots of the differences of log-likelihood functions evaluated at the estimated MLE and the true value for Models (M.1) and (M.2) with sample sizes  $n = 2^7, 2^8, 2^9$  and  $n = 2^{10}$ , respectively. The red “continuous” line indicates the position of zero.

#### 4. OBSERVATION-DRIVEN MODELS

---

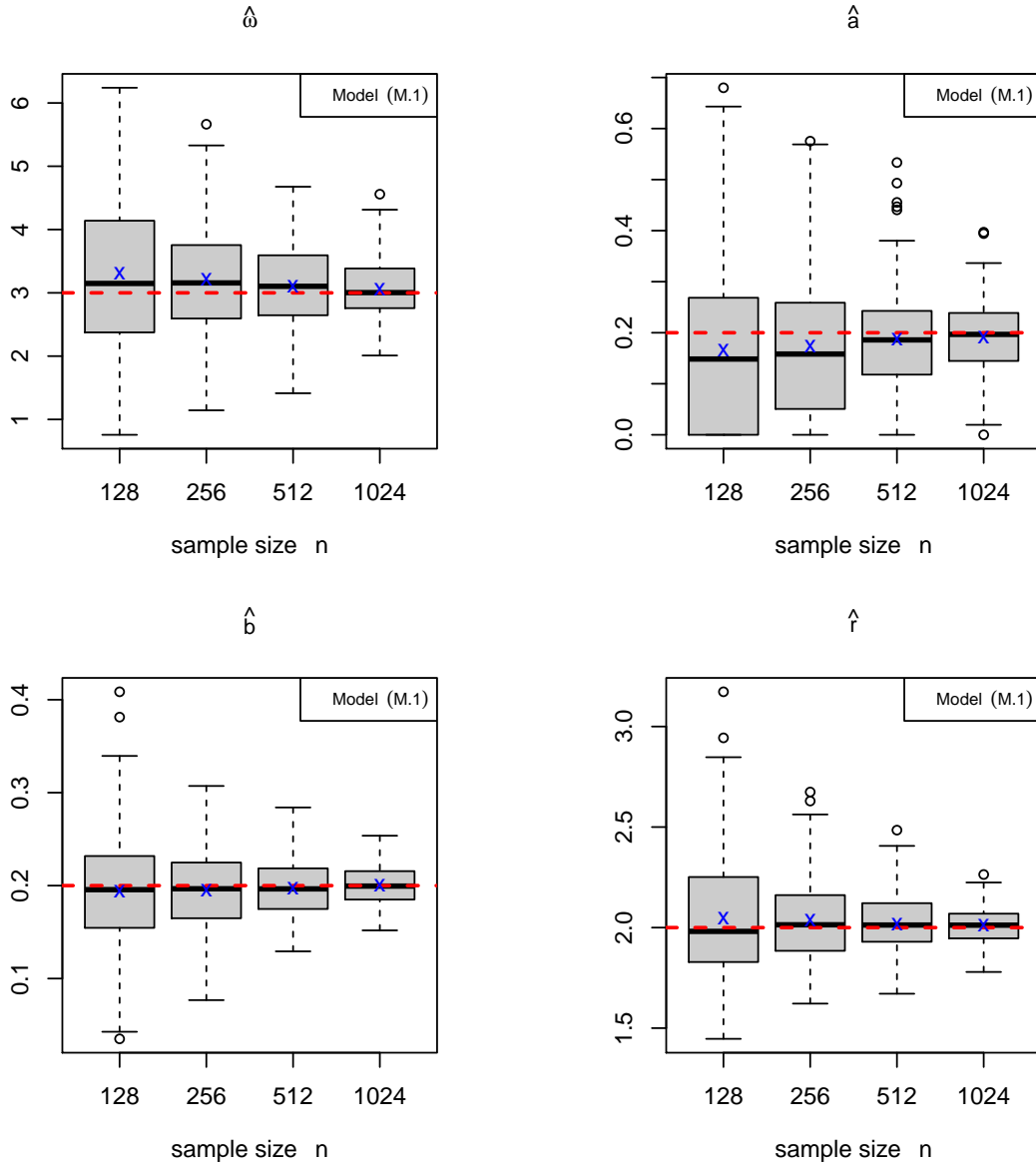


Figure 4.3 – Boxplots of the estimated MLE for Model (M.1) with sample sizes  $n = 2^7, 2^8, 2^9$  and  $n = 2^{10}$ , respectively. The red “dashed” line indicates the true value of the parameter and the blue “x” indicates the location of the Monte Carlo mean of the MLE.

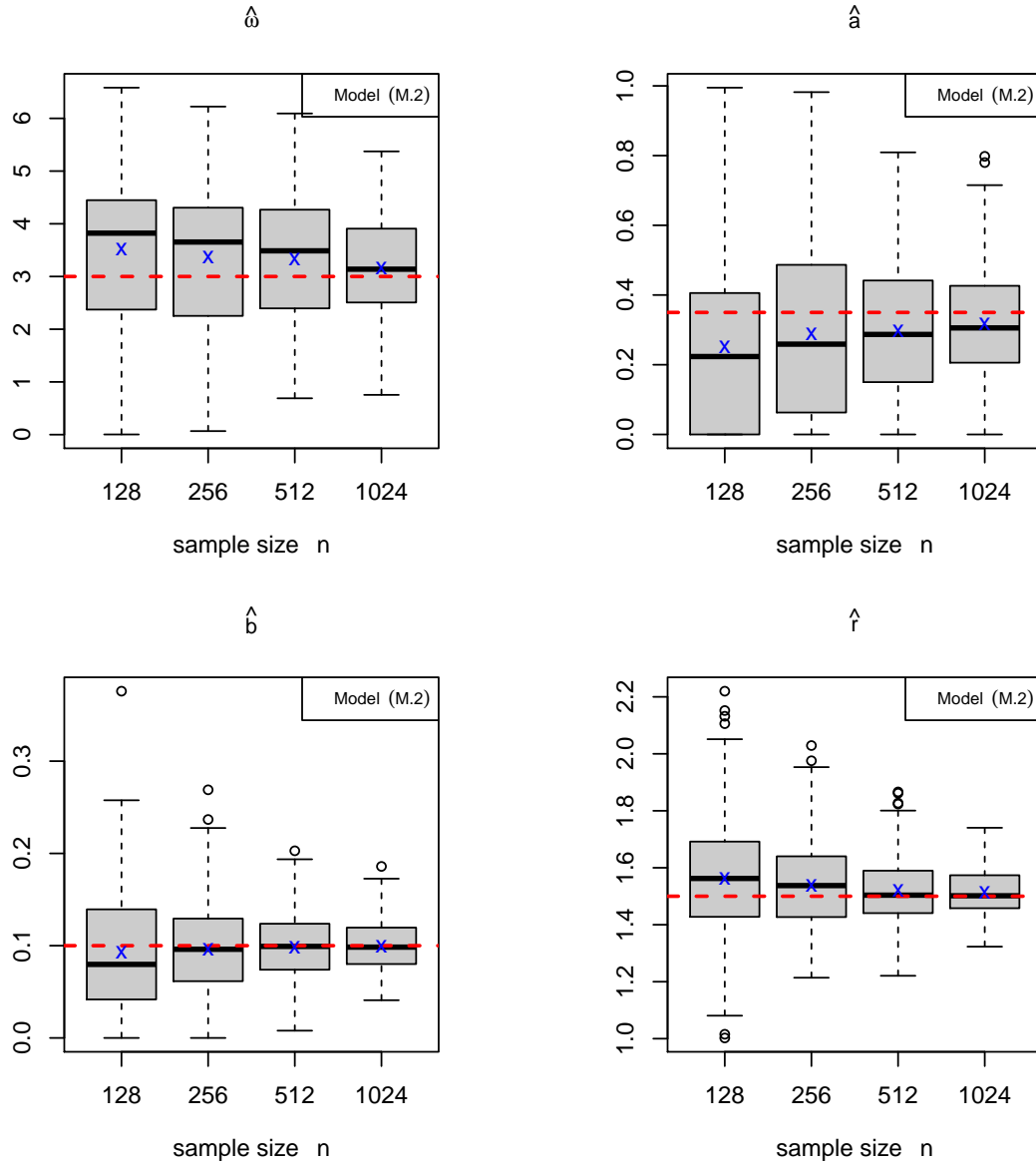


Figure 4.4 – Same as Figure 4.3 but for Model (M.2).

## 4.6 Postponed Proofs

### 4.6.1 Convergence of the MLE

Assumptions (A-1) and (A-2) are supposed to hold throughout this section. The proof of Theorem 4.3.2 relies on the approach introduced in Pfan-

zagl [1969], which was already used in Douc et al. [2013] for a restricted class of ODMs. Our main contribution here is to provide the handy conditions listed in Assumption (B-4). We first show that our conditions imply (B-1) and the following one.

(B-5) There exists  $x_1 \in X$  such that, for all  $\theta, \theta_* \in \Theta$ ,  $p^\theta(Y_1 | Y_{-\infty:0})$  defined as in (4.12) with  $x = x_1$  is finite  $\tilde{\mathbb{P}}^{\theta_*}$ -a.s. Moreover, for all  $\theta_* \in \Theta$ , we have

$$\lim_{k \rightarrow \infty} \sup_{\theta \in \Theta} \left| \ln \frac{g^\theta(\psi^\theta \langle Y_{1:k-1} \rangle(x_1); Y_k)}{p^\theta(Y_k | Y_{-\infty:k-1})} \right| = 0 \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.} \quad (4.39)$$

Indeed we have the following lemma.

**Lemma 4.6.1.** *Assumptions (B-2), (B-3) and (B-4) imply (B-5) and (B-1).*

*Proof.* See Section 4.6.3.  $\square$

Now the proof of Theorem 4.3.2 directly follows from the following lemma.

**Lemma 4.6.2.** *Assume that (B-2), (B-3) and (B-4)(i)–(ii) hold and that  $x_1$  satisfies (B-5). Then,  $\Theta_*$  defined by (4.13) is a non-empty closed subset of  $\Theta$  and (4.14) holds.*

*Proof.* By [Douc et al., 2013, Theorem 33], to obtain (4.14), it is sufficient to show that, for all  $\theta_* \in \Theta$ , the two following assertions hold.

(a)  $\tilde{\mathbb{E}}^{\theta_*} [\sup_{\theta \in \Theta} \ln^+ p^\theta(Y_1 | Y_{-\infty:0})] < \infty$ ,

(b) the function  $\theta \mapsto \ln p^\theta(Y_1 | Y_{-\infty:0})$  is continuous on  $\Theta$ ,  $\tilde{\mathbb{P}}^{\theta_*}$ -a.s.

In (B-5),  $p^\theta(Y_1 | Y_{-\infty:0})$  is defined  $\tilde{\mathbb{P}}^{\theta_*}$ -a.s. as the limit in (4.12) with  $x = x_1$ . So,  $\tilde{\mathbb{P}}^{\theta_*}$ -a.s., by (B-4)(i)–(ii),  $p^\theta(Y_1 | Y_{-\infty:0})$  is bounded by the finite constant appearing in (B-4)(ii). Hence (a) holds.

Condition (b) then follows from (4.39). Since almost sure convergence implies the convergence in probability and  $\tilde{\mathbb{P}}^{\theta_*}$  is shift invariant, the random sequence

$$U_m := \sup_{\theta \in \Theta} \left| \ln \frac{g^\theta(\psi^\theta \langle Y_{-m:0} \rangle(x_1); Y_1)}{p^\theta(Y_1 | Y_{-\infty:0})} \right|, \quad m \in \mathbb{Z}_+,$$

converges to zero in  $\tilde{\mathbb{P}}^{\theta_*}$ -probability. Then there exists a subsequence of  $\{U_m : m \in \mathbb{Z}_+\}$  which converges  $\tilde{\mathbb{P}}^{\theta_*}$ -a.s. to zero. Hence, interpreting this convergence as a uniform (in  $\theta$ ) convergence of  $\ln g^\theta(\psi^\theta \langle Y_{-m:0} \rangle(x_1); Y_1)$  to  $\ln p^\theta(Y_1 | Y_{-\infty:0})$  to conclude that (b) holds, it is sufficient to show that  $\theta \mapsto \ln g^\theta(\psi^\theta \langle Y_{-m:0} \rangle(x_1); Y_1)$  is continuous for all  $m$   $\tilde{\mathbb{P}}^{\theta_*}$ -a.s. This is indeed the case by (B-2) and (B-3) and since  $g^\theta(x; y)$  is positive.  $\square$

### 4.6.2 Ergodicity

For proving [Theorem 4.3.5](#), we first recall a more general set of conditions derived in [Douc et al. \[2013\]](#), which are based on the following definition.

**Definition 4.6.3.** Let  $\bar{G}$  be a probability kernel from  $\mathsf{X}^2$  to  $\mathcal{Y}^{\otimes 2} \otimes \mathcal{P}(\{0, 1\})$  satisfying the following marginal conditions, for all  $(x, x') \in \mathsf{X}^2$  and  $B \in \mathcal{Y}$ ,

$$\begin{cases} \bar{G}((x, x'); B \times \mathsf{Y} \times \{0, 1\}) = G(x; B), \\ \bar{G}((x, x'); \mathsf{Y} \times B \times \{0, 1\}) = G(x'; B), \end{cases} \quad (4.40)$$

and such that the following coupling condition holds

$$\bar{G}((x, x'); \{(y, y) : y \in \mathsf{Y}\} \times \{1\}) = \bar{G}((x, x'); \mathsf{Y}^2 \times \{1\}). \quad (4.41)$$

Define the following quantities successively.

- The trace measure of  $\bar{G}((x, x'); \cdot)$  on the set  $\{(y, y) : y \in \mathsf{Y}\} \times \{1\}$  is denoted by

$$\check{G}((x, x'); B) = \bar{G}((x, x'); \{(y, y) : y \in B\} \times \{1\}), \quad B \in \mathcal{Y}. \quad (4.42)$$

- The kernel  $\bar{R}$  from  $(\mathsf{X}^2, \mathcal{X}^{\otimes 2})$  to  $(\mathsf{X}^2 \times \{0, 1\}, \mathcal{X}^{\otimes 2} \otimes \mathcal{P}(\{0, 1\}))$  is defined for all  $x, x' \in \mathsf{X}^2$  and  $A \in \mathcal{X}^{\otimes 2}$  by

$$\bar{R}((x, x'); A \times \{1\}) = \int_{\mathsf{Y}} \mathbb{1}_A(\psi_y(x), \psi_y(x')) \check{G}((x, x'); dy). \quad (4.43)$$

- The measurable function  $\alpha$  from  $\mathsf{X}^2$  to  $[0, 1]$  is defined by

$$\alpha(x, x') = \bar{R}((x, x'); \mathsf{X}^2 \times \{1\}) = \bar{G}((x, x'); \mathsf{Y}^2 \times \{1\}). \quad (4.44)$$

- The kernel  $\hat{R}$  is defined for all  $(x, x') \in \mathsf{X}^2$  and  $A \in \mathcal{X}^{\otimes 2}$  by

$$\hat{R}((x, x'); A) = \begin{cases} \frac{\bar{R}((x, x'); A \times \{1\})}{\alpha(x, x')} & \text{if } \alpha(x, x') > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (4.45)$$

We can now introduce the so-called *contracting condition* which yields ergodicity.

- (A-9) There exists a kernel  $\bar{G}$  yielding  $\alpha$  and  $\hat{R}$  as in [Definition 4.6.3](#), a measurable function  $W : \mathsf{X}^2 \rightarrow [1, \infty)$  satisfying [Conditions \(A-8\)\(ii\)](#) and [\(A-8\)\(iii\)](#) and real numbers  $(D, \zeta_1, \zeta_2, \rho) \in (\mathbb{R}_+)^3 \times (0, 1)$  such that, for all  $(x, x') \in \mathsf{X}^2$  and for all  $n \geq 1$ ,

$$\hat{R}^n((x, x'); d_{\mathsf{X}}) \leq D \rho^n d_{\mathsf{X}}(x, x'), \quad (4.46)$$

$$\hat{R}^n((x, x'); d_{\mathsf{X}} \times W) \leq D \rho^n d_{\mathsf{X}}^{\zeta_1}(x, x') W^{\zeta_2}(x, x'). \quad (4.47)$$

Under Conditions (A-3), (A-4), (A-5), (A-6) and (A-9) and by combining Theorem 6, Proposition 8 and Lemma 7 in Douc et al. [2013], we immediately obtain the following result.

**Theorem 4.6.4.** *Assume (A-3), (A-4), (A-5), (A-6) and (A-9). Then the Markov kernel  $K$  admits a unique invariant distribution  $\pi$  and  $\pi_1(\bar{V}) < \infty$  for any  $\bar{V} : \mathsf{X} \rightarrow \mathbb{R}_+$  such that  $\bar{V} \lesssim V$ .*

Assumptions (A-3), (A-4), (A-5) and (A-6) are quite usual and easy to check. The key point to obtain ergodicity is thus to construct  $\bar{G}$  satisfying (A-9). For this, we can also rely on the following result which is quoted from [Douc et al., 2013, Lemma 9].

**Lemma 4.6.5.** *Assume that there exists  $(\rho, \beta) \in (0, 1) \times \mathbb{R}$  such that for all  $(x, x') \in \mathsf{X}^2$ ,*

$$\hat{R}((x, x'); \{(x_1, x'_1) \in \mathsf{X}^2 : d_{\mathsf{X}}(x_1, x'_1) > \rho d_{\mathsf{X}}(x, x')\}) = 0, \quad (4.48)$$

$$\hat{R}W \leq W + \beta. \quad (4.49)$$

Then, (4.46) and (4.47) hold.

Now we can prove that our set of conditions is sufficient.

*Proof of Theorem 4.3.5.* We only need to show that (A-7) and (A-8) imply (A-9). We preface our proof by the following lemma.

**Lemma 4.6.6.** *Assume (A-8)(i). Then one can define a kernel  $\bar{G}$  as in Definition 4.6.3 with the same  $\alpha$  given in (4.44). Moreover, the kernel  $\hat{R}$  defined by (4.45) satisfies, for all  $(x, x') \in \mathsf{X}^2$  such that  $\alpha(x, x') > 0$  and all measurable functions  $f : \mathsf{X}^2 \rightarrow \mathbb{R}_+$ ,*

$$\hat{R}((x, x'); f) = G(\phi(x, x'); \tilde{f}) \quad \text{with} \quad \tilde{f}(y) = f(\psi_y(x), \psi_y(x')). \quad (4.50)$$

Let us conclude the proof of Theorem 4.3.5 before proving this lemma. By Lemma 4.6.6 and Lemma 4.6.5, it remains to check that (4.48) and (4.49) hold for all  $(x, x') \in \mathsf{X}^2$ . Observe that by definition of  $\hat{R}$ , Condition (A-8)(iv) is equivalent to

$$\sup_{(x, x') \in \mathsf{X}^2} \left( \hat{R}W(x, x') - W(x, x') \right) < \infty.$$

so we can find  $\beta \in \mathbb{R}$  such that (4.49) holds for all  $(x, x') \in \mathsf{X}^2$ .

Now, let  $(x, x') \in \mathsf{X}^2$  and let  $(X, X')$  be distributed according to  $\hat{R}((x, x'); \cdot)$  which is defined in (4.50). When  $x = x'$ , then  $d_{\mathsf{X}}(X, X') = 0$ ,

implying that Condition (4.48) holds with any nonnegative  $\rho$ . For  $x \neq x'$ , let  $\rho$  be defined by

$$\rho = \sup_{\substack{(x,x',y) \in \mathsf{X}^2 \times \mathsf{Y} \\ x \neq x'}} \frac{d_{\mathsf{X}}(\psi_y(x), \psi_y(x'))}{d_{\mathsf{X}}(x, x')}, \quad (4.51)$$

which is in  $(0, 1)$  by (A-7). Then

$$\frac{d_{\mathsf{X}}(X, X')}{d_{\mathsf{X}}(x, x')} = \frac{d_{\mathsf{X}}(\psi_Y(x), \psi_Y(x'))}{d_{\mathsf{X}}(x, x')} \leq \rho.$$

Therefore, Condition (4.48) holds for all  $(x, x') \in \mathsf{X}^2$  with  $\rho$  as in (4.51).  $\square$

We conclude this section with the postponed

*Proof of Lemma 4.6.6.* Let  $(x, x') \in \mathsf{X}^2$ . We define  $\bar{G}((x, x'); \cdot)$  as the distribution of  $(Y, Y', U)$  drawn as follows. We first draw a random variable  $\bar{Y}$  taking values in  $\mathsf{Y}$  with density  $g(\phi(x, x'); \cdot)$  with respect to  $\nu$ . Then we define  $(Y, Y', U)$  by separating the two cases,  $\alpha(x, x') = 1$  and  $\alpha(x, x') < 1$ .

- Suppose that  $\alpha(x, x') = 1$ . Then from (A-8)(i), we have

$$G(x; \cdot) = G(x'; \cdot) = G(\phi(x, x'); \cdot).$$

In this case, we set  $(Y, Y', U) = (\bar{Y}, \bar{Y}, 1)$ .

- Suppose now that  $\alpha(x, x') < 1$ . Then, using (4.20), the functions

$$(1 - \alpha(x, x'))^{-1} [g(x; \cdot) - \alpha(x, x')g(\phi(x, x'); \cdot)]$$

and

$$(1 - \alpha(x, x'))^{-1} [g(x'; \cdot) - \alpha(x, x')g(\phi(x, x'); \cdot)],$$

are probability density functions with respect to  $\nu$  and we let  $\Lambda$  and  $\Lambda'$  be two independent random variables taking values in  $\mathsf{Y}$  drawn with these two density functions, respectively. In this case, we draw  $U$  independently according to a Bernoulli variable with mean  $\alpha(x, x')$  and set

$$(Y, Y') = \begin{cases} (\bar{Y}, \bar{Y}) & \text{if } U = 1, \\ (\Lambda, \Lambda') & \text{if } U = 0. \end{cases}$$

One can easily check that the so defined kernel  $\bar{G}$  satisfies (4.40) and (4.41). Moreover, for all  $(x, x') \in \mathsf{X}^2$ ,

$$\bar{G}((x, x'); \mathsf{Y}^2 \times \{1\}) = \mathbb{P}(U = 1) = \alpha(x, x'),$$



which is compatible with (4.44). The kernel  $\hat{R}$  is defined by setting  $\hat{R}((x, x'); \cdot)$  as the conditional distribution of  $(X, X') = (\psi_Y(x), \psi_Y(x'))$  given that  $U = 1$ . To complete the proof of Lemma 4.6.6, observe that for any measurable  $f : \mathsf{X}^2 \rightarrow \mathbb{R}_+$ , we have, for all  $(x, x') \in \mathsf{X}^2$  such that  $\alpha(x, x') > 0$ ,

$$\begin{aligned} \hat{R}((x, x'); f) &= \mathbb{E}[f(\psi_Y(x), \psi_Y(x')) \mid U = 1] \\ &= \mathbb{E}[f(\psi_{\bar{Y}}(x), \psi_{\bar{Y}}(x'))] \\ &= G(\phi(x, x'); \tilde{f}), \end{aligned}$$

where  $\tilde{f}(y) = f(\psi_y(x), \psi_y(x'))$  for all  $y \in \mathsf{Y}$ . □

### 4.6.3 Proof of Lemma 4.6.1.

Under (A-2), Assumptions (B-4)(viii) implies that for all  $\theta \in \Theta$ ,

$$\pi_2^\theta(\ln^+(\bar{\phi})) < \infty, \quad (4.52)$$

and if moreover  $C > 0$ ,

$$\pi_2^\theta(\bar{\phi}) < \infty. \quad (4.53)$$

For proving Lemma 4.6.1, we will also make use of [Douc et al., 2013, Lemma 34] which we quote here for convenience.

**Lemma 4.6.7.** *Let  $\{U_k : k \in \mathbb{Z}_+\}$  be a stationary sequence of real-valued random variables on  $(\Omega, \mathcal{F}, \mathbb{P})$ . Assume that  $\mathbb{E}(\ln^+ |U_0|) < \infty$ . Then, for all  $\eta \in (0, 1)$ ,*

$$\lim_{k \rightarrow \infty} \eta^k U_k = 0, \quad \mathbb{P}\text{-a.s.}$$

*Proof of Lemma 4.6.1.* We first show that  $p^\theta(y|Y_{-\infty:0})$  in (4.12) is finite for  $x = x_1$   $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s. By (B-2), this follows by writing

$$p^\theta(y_1 | y_{-\infty:0}) = g^\theta(\psi^\theta\langle y_{-\infty:0} \rangle; y_1), \quad (4.54)$$

if, for all  $\theta, \theta_\star \in \Theta$ , the limit

$$\psi^\theta\langle Y_{-\infty:0} \rangle = \lim_{m \rightarrow \infty} \psi^\theta\langle Y_{-m:0} \rangle(x_1) \quad \text{is well defined} \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.} \quad (4.55)$$

For all  $\theta \in \Theta$ ,  $m \geq 0$ ,  $x \in \mathsf{X}$  and  $y_{-m:0} \in \mathsf{Y}^{m+1}$ , using (B-4)(iii), we have

$$d_{\mathsf{X}}(\psi^\theta\langle y_{-m:0} \rangle(x_1), \psi^\theta\langle y_{-m:0} \rangle(x)) \leq \varrho^{m+1} \bar{\psi}(x). \quad (4.56)$$

Taking  $x = \psi_{y_{-m-1}}^\theta(x_1)$  and using (B-4)(v), we obtain, for all  $y_{-m-1:0} \in \mathsf{Y}^{m+2}$ ,

$$d_{\mathsf{X}}(\psi^\theta\langle y_{-m:0} \rangle(x_1), \psi^\theta\langle y_{-m-1:0} \rangle(x_1)) \leq \varrho^{m+1} \bar{\phi}(y_{-m-1}).$$

Using (4.52) and Lemma 4.6.7, we have that

$$\forall \eta \in (0, 1), \quad \sum_{k \in \mathbb{Z}} \eta^{|k|} \bar{\phi}(Y_k) < \infty, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}, \quad (4.57)$$

and thus  $\{\psi^\theta \langle Y_{-m:0} \rangle(x_1) : m \in \mathbb{Z}_+\}$  is a Cauchy sequence  $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s. Its limit exists  $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s., since  $(\mathbf{X}, d_X)$  is assumed to be complete, which defines the  $\mathbf{X}$ -valued random variable  $\psi^\theta \langle Y_{-\infty:0} \rangle$  for all  $\theta, \theta_\star \in \Theta$  when  $Y$  has distribution  $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s. Thus (4.55) holds and we further obtain that

$$\begin{aligned} \sup_{\theta \in \Theta} d_X(\psi^\theta \langle Y_{-k:0} \rangle(x_1), x_1) &\leq \sup_{\theta \in \Theta} \sum_{m=0}^k d_X(\psi^\theta \langle Y_{-m:0} \rangle(x_1), \psi^\theta \langle Y_{-m+1:0} \rangle(x_1)) \\ &\leq \sum_{m \geq 0} \varrho^m \bar{\phi}(Y_{-m}) < \infty, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.} \end{aligned} \quad (4.58)$$

so that, letting  $k \rightarrow \infty$ ,

$$\sup_{\theta \in \Theta} d_X(\psi^\theta \langle Y_{-\infty:0} \rangle, x_1) \leq \sum_{m \geq 0} \varrho^m \bar{\phi}(Y_{-m}) < \infty, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.} \quad (4.59)$$

Let us now prove (B-1). Relation (4.54) directly yields (B-1)(i). Let us prove (B-1)(ii), hence consider the case  $\theta = \theta_\star$ . Using (4.56), we have

$$d_X(\psi^{\theta_\star} \langle Y_{-m:0} \rangle(x_1), \psi^{\theta_\star} \langle Y_{-m:0} \rangle(X_{-m})) \leq \varrho^{m+1} \bar{\psi}(X_{-m}), \quad \mathbb{P}^{\theta_\star}\text{-a.s.}$$

Since  $\{\bar{\psi}(X_{-m}) : m \in \mathbb{Z}_+\}$  is stationary under  $\mathbb{P}^{\theta_\star}$ , it is bounded in probability, and since  $\varrho < 1$ , for all  $\epsilon > 0$ , we have

$$\lim_{m \rightarrow \infty} \mathbb{P}^{\theta_\star} (d_X(\psi^{\theta_\star} \langle Y_{-m:0} \rangle(X_{-m}), \psi^{\theta_\star} \langle Y_{-m:0} \rangle(x)) > \epsilon) = 0. \quad (4.60)$$

Note that for all  $m \geq 1$ ,  $\psi^{\theta_\star} \langle Y_{-m:0} \rangle(X_{-m}) = X_1$ ,  $\mathbb{P}^{\theta_\star}$ -a.s., hence we get that

$$\psi^{\theta_\star} \langle Y_{-\infty:0} \rangle = X_1, \quad \mathbb{P}^{\theta_\star}\text{-a.s.} \quad (4.61)$$

To complete the proof of (B-1)(ii), we need to show that, under  $\tilde{\mathbb{P}}^{\theta_\star}$ ,  $y \mapsto g^{\theta_\star}(\psi^{\theta_\star} \langle Y_{-\infty:0} \rangle; y) = g^{\theta_\star}(X_1; y)$  is the conditional density of  $Y_1$  given  $Y_{-\infty:0}$ , that is, for any  $B \in \mathcal{Y}$ ,

$$\int \mathbb{1}_B(y) g^{\theta_\star}(X_1; y) \nu(dy) = \mathbb{P}^{\theta_\star}(Y_1 \in B | Y_{-\infty:0}).$$

Now, note that, by definition of  $\mathbb{P}^{\theta_\star}$ ,

$$\int \mathbb{1}_B(y) g^{\theta_\star}(X_1; y) \nu(dy) = \mathbb{P}^{\theta_\star}(Y_1 \in B | X_1) = \mathbb{P}^{\theta_\star}(Y_1 \in B | X_1, Y_{-\infty:0}).$$

But since (4.61) implies that  $X_1$  is  $\sigma(Y_{-\infty:0})$ -measurable,  $X_1$  can be removed in the last conditioning, which concludes the proof (B-1)(ii).

Finally, it remains to show the uniform convergence (4.39) in (B-5). By (B-3) and (4.55), we have, for all  $\theta, \theta_* \in \Theta$ ,  $k \in \mathbb{Z}_+$ ,

$$\psi^\theta \langle Y_{-\infty:k-1} \rangle = \psi^\theta \langle Y_{1:k-1} \rangle (\psi^\theta \langle Y_{-\infty:0} \rangle), \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.} \quad (4.62)$$

From (B-4)(iii) and (4.62), we get

$$d_X(\psi^\theta \langle Y_{1:k-1} \rangle(x_1), \psi^\theta \langle Y_{-\infty:k-1} \rangle) \leq \varrho^{k-1} \bar{\psi}(\psi^\theta \langle Y_{-\infty:0} \rangle), \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.}$$

On the other hand (B-4)(iv) and (4.59) imply

$$\sup_{\theta \in \Theta} \bar{\psi}(\psi^\theta \langle Y_{-\infty:0} \rangle) < \infty, \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.}, \quad (4.63)$$

which, with the previous display, yields,

$$\sup_{\theta \in \Theta} d_X(\psi^\theta \langle Y_{1:k-1} \rangle(x_1), \psi^\theta \langle Y_{-\infty:k-1} \rangle) = O_{k \rightarrow \infty}(\varrho^k), \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.} \quad (4.64)$$

Since  $\mathbf{X}_1$  is closed and satisfies Condition (B-4)(i), we have that,  $\psi^\theta \langle Y_{1:k-1} \rangle(x_1)$  and  $\psi^\theta \langle Y_{-\infty:k-1} \rangle$  are in  $\mathbf{X}_1$  for all  $k \geq 2$ . Thus Condition (B-4)(vi) gives that

$$\sup_{\theta \in \Theta} \left| \ln \frac{g^\theta(\psi^\theta \langle Y_{1:k-1} \rangle(x_1); Y_k)}{g^\theta(\psi^\theta \langle Y_{-\infty:k-1} \rangle; Y_k)} \right| \leq A_k(1) \times A_k(2) \times A_k(3) \times A_k(4), \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.},$$

where

$$\begin{aligned} A_k(1) &= \sup_{\theta \in \Theta} H(d_X(\psi^\theta \langle Y_{1:k-1} \rangle(x_1), \psi^\theta \langle Y_{-\infty:k-1} \rangle)) \\ A_k(2) &= \sup_{\theta \in \Theta} e^{C d_X(x_1, \psi^\theta \langle Y_{-\infty:k-1} \rangle)} \\ A_k(3) &= \sup_{\theta \in \Theta} e^{C d_X(x_1, \psi^\theta \langle Y_{1:k-1} \rangle(x_1))} \\ A_k(4) &= \bar{\phi}(Y_k). \end{aligned}$$

By (4.64) and (B-4)(vii), we have

$$A_k(1) = O_{k \rightarrow \infty}(\varrho^k), \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.}$$

With (4.57), this yields (4.39) in the case where  $C = 0$ . For  $C > 0$ , we further observe that, by (4.59) and (4.53), we have, for all  $\theta_* \in \Theta$  and  $k \in \mathbb{Z}_+$ ,

$$\tilde{\mathbb{E}}^{\theta_*} [\ln^+ A_k(2)] \leq \tilde{\mathbb{E}}^{\theta_*} \left[ C \sum_{m \geq 0} \varrho^m \bar{\phi}(Y_{-m+k-1}) \right] = \frac{C \pi_2^{\theta_*}(\bar{\phi})}{1 - \varrho} < \infty.$$

Then [Lemma 4.6.7](#) implies that,  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,  $A_k(2) = O(\eta^{-k})$  for any  $\eta \in (0, 1)$ . The same property applies similarly to  $A_k(3)$  by using [\(4.58\)](#) in place of [\(4.59\)](#). This yields [\(4.39\)](#) in the case where  $C > 0$ , which concludes the proof.  $\square$

## Chapter 5

# General-order Observation-driven Models: Ergodicity, Consistency and Asymptotic Normality of the MLE

**Summary.** In this chapter, we allow the state equation of the hidden process of the observation-driven model discussed in [Chapter 4](#) to depend also on other lagged variables of the hidden and the observation processes, resulting in a GARCH( $p, q$ )-type observation-driven model. In this generalized class, stationary and ergodic solutions as well as consistency of the maximum likelihood estimator (MLE) are derived under sufficient and easy-to-check conditions. The asymptotic normality of the MLE is also treated. We illustrate these results with specific observation-driven times series, including the GARCH( $p, q$ ), the log-linear Poisson GARCH of order ( $p, q$ ), and the NBIN-GARCH( $p, q$ ) models. An empirical study is also provided.

### 5.1 Introduction

Since the introduction by [Cox \[1981\]](#), observation-driven models (ODMs) have been receiving renewed interest in recent years. These models are widely applied in various fields ranging from economics (see [Pindyck and Rubinfeld \[1998\]](#)), environmental study (see [Bhaskaran et al. \[2013\]](#)), epidemiology and public health study (see [Zeger \[1988\]](#), [Davis et al. \[1999\]](#), [Ferland et al. \[2006\]](#)), finance (see [Liesenfeld and Richard \[2003\]](#), [Rydberg and Shephard \[2003\]](#), [Fokianos and Tjøstheim \[2011\]](#), [Francq and Zakoian \[2011\]](#)) and population dynamics (see [Ives et al. \[2003\]](#)). The celebrated GARCH(1, 1) model

introduced in [Bollerslev \[1986\]](#) as well as most of the models derived from this one are typical examples of ODMs; see [Bollerslev \[2008\]](#) for a list of some of them. A list of contributions on this class of models specifically dealing with discrete data includes [Streett \[2000\]](#), [Davis et al. \[2003\]](#), [Heinen \[2003\]](#), [Ferland et al. \[2006\]](#), [Fokianos et al. \[2009\]](#), [Franke \[2010\]](#), [Woodard et al. \[2010\]](#), [Fokianos and Tjøstheim \[2011\]](#), [Henderson et al. \[2011\]](#), [Neumann \[2011\]](#), [Davis and Liu \[2012\]](#), [Doukhan et al. \[2012\]](#), [Douc et al. \[2013\]](#), [Fokianos et al. \[2013\]](#), [Christou and Fokianos \[2014, 2015\]](#) and [Douc et al. \[2015\]](#).

ODMs have the nice feature that the computations of the associated (conditional) likelihood and its derivatives are easy, the parameter estimation is hence relatively simple, and the prediction, which is a prime objective in many time series applications, is straightforward. However, it turns out that the asymptotic properties of the maximum likelihood estimator (MLE) for this class can be cumbersome to establish, except when they can be derived using computations specific to the studied model (the GARCH(1,1) case being one of the most celebrated example). The literature concerning the asymptotic theory of the MLE when the observed variable has Poisson distribution includes [Fokianos et al. \[2009\]](#), [Fokianos and Tjøstheim \[2011, 2012\]](#) and [Wang et al. \[2014\]](#). For a more general case where the model belongs to the class of one-parameter exponential ODMs, such as the Bernoulli, the exponential, the negative binomial (with known frequency parameter) and the Poisson autoregressive models, the consistency and the asymptotic normality of the MLE have been derived in [Davis and Liu \[2012\]](#). However, the one-parameter exponential family is inadequate to deal with models such as multi-parametric, mixture or multivariate ODMs (the negative binomial with all unknown parameters and mixture Poisson ODMs are examples of this case). A more general consistency result, yet not the asymptotic normality, has been obtained recently in [Douc et al. \[2013\]](#). This general result allows the observed process to admit various forms of distribution and to take values in any Borel space, and allows the hidden process to assume values in any locally compact Polish space endowed with the associated Borel  $\sigma$ -field. This result has later been extended and refined in [Douc et al. \[2015\]](#). However, most of the results obtained so far have been derived only under the framework of GARCH(1,1)-type or first-order ODMs, yet less is known for the GARCH( $p, q$ )-type discrete ODMs, as highlighted as a remaining unsolved problem in [Tjøstheim \[2015\]](#).

In this contribution, we consider among others (see [Streett \[2000\]](#), [Heinen \[2003\]](#)) a general class of ODMs that is capable to account for several lagged variables of both hidden and observation processes. Namely, we develop theory and inference for the class of general-order ODMs parallel to the

GARCH( $p, q$ ) family, under the assumption of well-specified models. For the development on the GARCH( $p, q$ ) model, see for example [Francq et al. \[2004\]](#), [Francq and Zakoian \[2011\]](#) and for multivariate case, see [Comte and Lieberman \[2003\]](#). We extend the approaches introduced in [Douc et al. \[2013\]](#) and later used in [Douc et al. \[2015\]](#) to establish the consistency of the MLE and the ergodicity for the models in this general context. These results can in principle be obtained by embedding the studied model into the corresponding first-order one and then applying the results obtained therein to the embedded model. By appropriate adjustments, the consistency of the MLE for the general-order ODM can be derived. Yet for ergodicity, the generalized result seems not trivial; it turns out that more general assumptions are required, compared to the usual first-order ODMs. In either [Douc et al. \[2013\]](#) or [Douc et al. \[2015\]](#), the ergodicity of the usual first-order ODMs is obtained by showing the uniqueness of invariant probability measure for the hidden process, which is by its own right a Markov chain. In these papers, showing the existence of the invariant probability measure for the hidden process relies on the Foster-Lyapunov-type assumption; however, for general ODMs or embedded ones, this assumption may be replaced by the *iterative* Foster-Lyapunov-type assumption instead (see Assumption [\(AG-4\)](#) of [Section 5.3.4](#)). The asymptotic normality of the MLE is also investigated in this general setting. In this study, we restrict our consideration to the special yet an important case where the current hidden state variable is expressed as a linear function with respect to its own past variables. This class is rich enough to cover most of the familiar ODMs, such as the GARCH( $p, q$ ), the log-linear Poisson autoregression of order ( $p, q$ ) and the NBIN-GARCH( $p, q$ ) models, and other instances such as multi-parametric, mixture or multivariate ODMs. To establish the asymptotic normality, we follow the classical approach by first approximating the score function by the stationary version of it and then developing Taylor expansion of the stationary score function around the true parameter. By appropriate assumptions, the central limit theorem for Martingale difference applies and the stationary score function can then be shown to be asymptotically Gaussian. Then assuming invertibility of the asymptotic Fisher information matrix, the asymptotic normality follows. All the results are presented under sufficient and easy-to-check conditions. As demonstration, they are then applied to the GARCH( $p, q$ ), the log-linear Poisson autoregression of order ( $p, q$ ) and the NBIN-GARCH( $p, q$ ) models. Finally, we provide an empirical study suggesting that in some circumstance higher-order model may fit the data better than the first-order one if the class of ODMs is used.

This chapter is structured as follows. Specific definitions and notation are introduced in [Section 5.2](#). In [Section 5.3](#) presents the main results of

consistency and asymptotic normality of the MLE as well as ergodic solution for the model. We apply these results to some specific models in [Section 5.4](#). Numerical experiments is given in [Section 5.5](#). Finally, [Section 5.6](#) contains the postponed proofs.

## 5.2 Definitions and Notation

Let  $(\mathsf{X}, d_{\mathsf{X}})$  be a complete and separable metric space endowed with the associated Borel  $\sigma$ -field  $\mathcal{X}$  and  $(\mathsf{Y}, \mathcal{Y})$  be a Borel space. Let  $(\Theta, \Delta)$ , the set of parameters, be a compact metric space equipped with the metric  $\Delta$ ,  $\{G^\theta : \theta \in \Theta\}$  be a family of probability kernels on  $\mathsf{X} \times \mathsf{Y}$  and  $p, q$  be two positive integers. The observation-driven time series model can be formally defined as follows.

**Definition 5.2.1.** A time series  $\{Y_k : k \geq -q + 1\}$  valued in  $\mathsf{Y}$  is said to be distributed according to an *observation-driven model of order  $(p, q)$*  (hereafter,  $\text{ODM}(p, q)$ ) with parameter  $\theta \in \Theta$  if there exist a family of measurable functions  $\{(x_{1:p}, y_{1:q}) \mapsto \psi_{y_{1:q}}^\theta(x_{1:p}) : \theta \in \Theta\}$  from  $(\mathsf{X}^p \times \mathsf{Y}^q, \mathcal{X}^{\otimes p} \otimes \mathcal{Y}^{\otimes q})$  to  $(\mathsf{X}, \mathcal{X})$  and a process  $\{X_k : k \geq -p + 1\}$  on  $(\mathsf{X}, \mathcal{X})$  such that for all  $k \in \mathbb{Z}_+$ ,

$$\begin{aligned} X_{k+1} &= \psi_{Y_{k-q+1:k}}^\theta(X_{k-p+1:k}), \\ Y_{k+1} \mid \mathcal{F}_k &\sim G^\theta(X_{k+1}; \cdot), \end{aligned} \tag{5.1}$$

where  $\mathcal{F}_k = \sigma(X_{-p+1:k+1}, Y_{-q+1:k})$  and  $u_{\ell:m} := (u_\ell, \dots, u_m)$  for  $\ell \leq m$ . Moreover, we say that the model is dominated by some  $\sigma$ -finite measure  $\nu$  on  $(\mathsf{Y}, \mathcal{Y})$  if for all  $x \in \mathsf{X}$ , the probability kernel  $G^\theta(x; \cdot)$  is dominated by  $\nu$ . In this case we denote by  $g^\theta(x; \cdot)$  its Radon–Nikodym derivative,  $g^\theta(x; y) = \frac{dG^\theta(x; \cdot)}{d\nu}(y)$ , and we always assume that for all  $(x, y) \in \mathsf{X} \times \mathsf{Y}$  and for all  $\theta \in \Theta$ ,

$$g^\theta(x; y) > 0.$$

One of the most popular examples of this class is the general GARCH( $p, q$ ) model introduced by [Bollerslev \[1986\]](#), where  $\mathsf{X} = (0, \infty)$ ,  $\mathsf{Y} = \mathbb{R}$ ,  $G^\theta(x; \cdot)$  is the centered Gaussian distribution of variance  $x$ , the deterministic function  $\psi_{y_{1:q}}^\theta(x_{1:p}) = \omega + \sum_{i=1}^p a_i x_i + \sum_{i=1}^q b_i y_i^2$  and  $\theta = (\omega, a_{1:p}, b_{1:q})$  with  $\omega > 0$  and  $a_{1:p}, b_{1:q} \geq 0$ . This model was then extensively studied by, for example, [Bougerol and Picard \[1992\]](#), [Francq et al. \[2004\]](#), [Francq and Zakoian \[2009\]](#), [Lindner \[2009\]](#), [Francq and Zakoian \[2011\]](#) and the references therein. For other GARCH examples of this class, see [Bollerslev \[2008\]](#).

**Remark 5.2.2.** When  $p = q = 1$ , then the  $\text{ODM}(p, q)$  defined by (5.1) collapses to the first-order ODM considered in [Douc et al. \[2013\]](#) and [Douc](#)



et al. [2015]. Note also that setting  $r := \max(p, q)$  and defining  $\varphi^\theta : \mathbf{X}^r \times \mathbf{Y}^r \rightarrow \mathbf{X}$  by, for all  $(x_{-r+1:0}, y_{-r+1:0}) \in \mathbf{X}^r \times \mathbf{Y}^r$ ,

$$\varphi_{y_{-r+1:0}}^\theta(x_{-r+1:0}) := \psi_{y_{-q+1:0}}^\theta(x_{-p+1:0}),$$

then the  $\text{ODM}(p, q)$  can be generally embedded in an  $\text{ODM}(r, r)$ . Thus without loss of generality, we can always assume that  $p = q$ .

The inference about the model parameter is performed by relying on the conditional likelihood of the observations  $(Y_1, \dots, Y_n)$  given  $(X_{-p+1:0}, Y_{-q+1:0}) = (x_{-p+1:0}, y_{-q+1:0}) := z$  for an arbitrary  $z \in \mathbf{X}^p \times \mathbf{Y}^q$ . The corresponding conditional density function with respect to  $\nu^{\otimes n}$  is, under parameter  $\theta \in \Theta$ , for all  $z = (x_{-p+1:0}, y_{-q+1:0}) \in \mathbf{X}^p \times \mathbf{Y}^q$ ,

$$y_{1:n} \mapsto \prod_{k=0}^{n-1} g^\theta(\psi^\theta\langle y_{-q+1:k} \rangle(x_{-p+1:0}); y_{k+1}), \quad (5.2)$$

where, for any vector  $y_{0:k}$ , the function  $\psi^\theta\langle y_{0:k} \rangle : \mathbf{X}^p \times \mathbf{Y}^q \rightarrow \mathbf{X}$  is successively defined by, for all  $k \in \mathbb{Z}_+$ ,

$$\begin{aligned} & \psi^\theta\langle y_{-q+1:k} \rangle(x_{-p+1:0}) \\ & := \psi_{y_{-q+1:k}}^\theta(\psi^\theta\langle y_{-q+1:k-p} \rangle(x_{-p+1:0}), \dots, \psi^\theta\langle y_{-q+1:k-1} \rangle(x_{-p+1:0})) \end{aligned} \quad (5.3)$$

with the convention, for  $1 \leq j \leq p$ ,

$$\psi^\theta\langle y_{-q+1:-j} \rangle(x_{-p+1:0}) := x_{-j+1}.$$

Then, the corresponding (conditional) maximum likelihood estimator (MLE)  $\hat{\theta}_{z,n}$  of the parameter  $\theta$ , with  $z = (x_{-p+1:0}, y_{-q+1:0})$ , is defined by

$$\hat{\theta}_{z,n} \in \operatorname{argmax}_{\theta \in \Theta} \mathbf{L}_{z,n}^\theta, \quad (5.4)$$

where

$$\mathbf{L}_{z,n}^\theta := n^{-1} \sum_{k=0}^{n-1} \ln g^\theta(\psi^\theta\langle y_{-q+1:0}, Y_{1:k} \rangle(x_{-p+1:0}); Y_{k+1}). \quad (5.5)$$

In this contribution, we investigate the convergence of  $\hat{\theta}_{z,n}$  as  $n \rightarrow \infty$  for some (well-chosen) value of  $z$  under the assumption that the model is well specified and the observations are in a steady state. That is, we assume that the observations  $\{Y_k : k \in \mathbb{Z}_+\}$  are distributed according to  $\tilde{\mathbb{P}}^{\theta_\star}$  with some  $\theta_\star \in \Theta$ , where, for all  $\theta \in \Theta$ ,  $\tilde{\mathbb{P}}^\theta$  denotes the stationary distribution of the observation-driven time series corresponding to the parameter  $\theta$ . However,

whether such a distribution is well defined is not always obvious. Let us now detail how this probability distribution can be obtained.

For simplicity, we assume that  $p = q$  and that  $p \geq 2$  and let  $Z = \mathbf{X}^p \times \mathbf{Y}^{p-1}$  and  $\mathcal{Z} = \mathcal{X}^{\otimes p} \otimes \mathcal{Y}^{\otimes(p-1)}$ . Let  $\Psi^\theta : Z \times \mathbf{Y} \rightarrow Z$  be defined by, for all  $z = (z_{1:2p-1}) \in Z$  and  $y \in \mathbf{Y}$ ,

$$\Psi_y^\theta(z) := \left( z_2, \dots, z_p, \psi_{z_{p+1:2p-1}, y}^\theta(z_{1:p}), z_{p+2}, \dots, z_{2p-1}, y \right) \quad (5.6)$$

The function  $\Psi^\theta$  is indeed measurable on  $(Z \times \mathbf{Y}, \mathcal{Z} \otimes \mathcal{Y})$ . Moreover, for any vector  $y_{1:k} \in \mathbf{Y}^k$ , we define  $\Psi^\theta \langle y_{1:k} \rangle$  as a  $Z \rightarrow Z$  function obtained by the successive composition of  $\Psi_{y_1}^\theta, \Psi_{y_2}^\theta, \dots$ , and  $\Psi_{y_k}^\theta$ ,

$$\Psi^\theta \langle y_{1:k} \rangle = \Psi_{y_k}^\theta \circ \Psi_{y_{k-1}}^\theta \circ \dots \circ \Psi_{y_1}^\theta. \quad (5.7)$$

with the convention  $\Psi^\theta \langle y_{1:0} \rangle(z) = z$ . By letting  $Z_k = (X_{k-p+1:k}, Y_{k-p+1:k-1})$  and observing that  $\mathcal{F}_k = \sigma(Z_{0:k+1}, Y_{0:k})$ , Model (5.1) can be replaced by: for all  $k \in \mathbb{Z}_+$ ,

$$\begin{aligned} Z_{k+1} &= \Psi_{Y_k}^\theta(Z_k), \\ Y_{k+1} \mid \mathcal{F}_k^{Z, Y} &\sim H^\theta(Z_{k+1}; \cdot), \end{aligned} \quad (5.8)$$

where, for all  $z = (z_{1:2p-1}) \in Z$ ,

$$H^\theta(z; \cdot) := G^\theta(\Pi_p(z); \cdot) \quad (5.9)$$

and, for all  $j \in \{1, \dots, 2p-1\}$ ,  $\Pi_j(z) = z_j$ . By this representation, the ODM( $p, p$ ) is thus imbedded in an ODM( $1, 1$ ). This in principle allows us to apply the same results obtained for the class of ODMs( $1, 1$ ) to the broader class of ODMs( $p, p$ ). As an ODM( $1, 1$ ), the bivariate process  $\{(Z_k, Y_k) : k \in \mathbb{Z}_+\}$  is a Markov chain on the space  $(Z \times \mathbf{Y}, \mathcal{Z} \otimes \mathcal{Y})$  with transition kernel  $K^\theta$  satisfying, for all  $(z, y) \in Z \times \mathbf{Y}$ ,  $A \in \mathcal{Z}$  and  $B \in \mathcal{Y}$ ,

$$K^\theta((z, y); A \times B) = \iint \mathbb{1}_{A \times B}(z', y') \delta_{\Psi_y^\theta(z)}(dz') G^\theta(\Pi_p(z); dy'). \quad (5.10)$$

Note also that, by itself, the process  $\{Z_k : k \in \mathbb{Z}_+\}$  is a Markov chain on  $(Z, \mathcal{Z})$  with transition kernel  $R^\theta$  satisfying, for all  $z \in Z$  and  $A \in \mathcal{Z}$ ,

$$R^\theta(z; A) = \int \mathbb{1}_A(\Psi_y^\theta(z)) H^\theta(z; dy) = \int \mathbb{1}_A(\Psi_y^\theta(z)) G^\theta(\Pi_p(z); dy). \quad (5.11)$$

Nevertheless, to apply known results to this embedded ODM(1, 1), some generalizations are needed. First we assume that  $(Y, d_Y)$  is a locally compact, complete and separable space equipped with the metric  $d_Y$ . Then, since  $(X, d_X)$  is a locally compact, complete and separable space, this implies that the space  $(Z, d_Z)$  is also locally compact, complete and separable with the metric  $d_Z$  appropriately defined as a function of  $d_X$  and  $d_Y$  so that  $\mathcal{Z}$  is the associated Borel  $\sigma$ -field. Throughout this chapter, let us assume the following ergodicity assumption.

- (AG-1) For all  $\theta \in \Theta$ , the transition kernel  $K^\theta$  of the complete chain admits a unique stationary distribution  $\pi^\theta$  on  $Z \times Y$ .

With this assumption, we can now define  $\tilde{\mathbb{P}}^\theta$ . The following notation and definitions will be used throughout the chapter.

**Definition 5.2.3.** Under Assumption (AG-1), we denote by  $\pi_1^\theta$  and  $\pi_2^\theta$  the marginal distributions of  $\pi^\theta$  on  $Z$  and  $Y$ , respectively, and by  $\mathbb{P}^\theta$  and  $\tilde{\mathbb{P}}^\theta$  the probability distributions defined respectively as follows.

- a)  $\mathbb{P}^\theta$  denotes the extension of  $\mathbb{P}_{\pi^\theta}^\theta$  on the whole line  $(Z \times Y)^{\mathbb{Z}}$ .
- b)  $\tilde{\mathbb{P}}^\theta$  is the corresponding projection on the component  $Y^{\mathbb{Z}}$ .

We also use the symbols  $\mathbb{E}^\theta$  and  $\tilde{\mathbb{E}}^\theta$  to denote the expectations corresponding to  $\mathbb{P}^\theta$  and  $\tilde{\mathbb{P}}^\theta$ , respectively. Moreover, for all  $\theta, \theta' \in \Theta$ , we write  $\theta \sim \theta'$  if and only if  $\tilde{\mathbb{P}}^\theta = \tilde{\mathbb{P}}^{\theta'}$ . This defines an equivalence relation on the parameter set  $\Theta$  and the corresponding equivalence class of  $\theta$  is denoted by  $[\theta] := \{\theta' \in \Theta : \theta \sim \theta'\}$ .

The equivalence relationship  $\sim$  was introduced by [Leroux \[1992\]](#) as an alternative to the classical identifiability condition.

For any probability distribution  $\xi$  on  $Z \times Y$ , we denote by  $\mathbb{P}_\xi^\theta$  the distribution of the Markov chain  $\{(Z_k, Y_k) : k \in \mathbb{Z}_+\}$  with kernel  $K^\theta$  and initial probability measure  $\xi$ .

## 5.3 Main Results

### 5.3.1 Preliminaries

In well-specified setting, a general result on the consistency of the MLE for a class of first-order ODMs has been obtained in [Douc et al. \[2013\]](#). This result has been later extended in [Douc et al. \[2015\]](#) to a wider class of first-order ODMs, but the latter has been derived under a handy set of assumptions which appear to be quite direct and easy when checking. The

approach used to establish the convergence of the MLE  $\hat{\theta}_{z,n}$  in these references and in our contribution is briefly described as follows.

First, we establish that, as the number of observations  $n \rightarrow \infty$ , the normalized log-likelihood  $\mathsf{L}_{z,n}^\theta$  defined in (5.5), for some well-chosen  $z \in \mathbb{Z}$ , can be approximated by

$$n^{-1} \sum_{k=1}^n \ln p^\theta(Y_k | Y_{-\infty:k-1}),$$

where  $p^\theta(\cdot|\cdot)$  is a  $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s. finite real-valued measurable function defined on  $(\mathcal{Y}^{\mathbb{Z}}, \mathcal{Y}^{\otimes \mathbb{Z}})$ . To define  $p^\theta(\cdot|\cdot)$ , we usually set, for all  $y_{-\infty:1} \in \mathcal{Y}^{\mathbb{Z}^-}$ , whenever the following limit is well defined,

$$p^\theta(y_1 | y_{-\infty:0}) = \begin{cases} \lim_{m \rightarrow \infty} g^\theta(\Pi_p(\Psi^\theta \langle y_{-m:0} \rangle(z)); y_1) & \text{if the limit exists,} \\ \infty & \text{otherwise.} \end{cases} \quad (5.12)$$

By (AG-1), the observed process  $\{Y_k : k \in \mathbb{Z}\}$  is ergodic under  $\tilde{\mathbb{P}}^{\theta_\star}$  and provided that

$$\tilde{\mathbb{E}}^{\theta_\star} [\ln^+ p^\theta(Y_1 | Y_{-\infty:0})] < \infty,$$

it then follows that

$$\lim_{n \rightarrow \infty} \mathsf{L}_{z,n}^\theta = \tilde{\mathbb{E}}^{\theta_\star} [\ln p^\theta(Y_1 | Y_{-\infty:0})], \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

Finally, we show that with probability tending to one, the MLE  $\hat{\theta}_{z,n}$  eventually lies in a neighborhood of the set

$$\Theta_\star = \operatorname{argmax}_{\theta \in \Theta} \tilde{\mathbb{E}}^{\theta_\star} [\ln p^\theta(Y_1 | Y_{-\infty:0})], \quad (5.13)$$

which only depends on  $\theta_\star$ , establishing that

$$\lim_{n \rightarrow \infty} \Delta(\hat{\theta}_{z,n}, \Theta_\star) = 0, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}, \quad (5.14)$$

where  $\Delta$  is the metric endowing the parameter space  $\Theta$ .

In Douc et al. [2015], their easy-to-check conditions also induce that for all  $\theta, \theta_\star \in \Theta$ , there exists a  $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s. finite measurable function  $\Psi^\theta \langle \cdot \rangle : \mathcal{Z}^{\mathbb{Z}^-} \rightarrow \mathcal{Z}$  such that for some appropriate value  $z \in \mathcal{Z}$ ,

$$\lim_{m \rightarrow \infty} \Psi^\theta \langle Y_{-m:0} \rangle(z) = \Psi^\theta \langle Y_{-\infty:0} \rangle, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.} \quad (5.15)$$

$$\mathcal{Z}_1 = \Psi^{\theta_\star} \langle Y_{-\infty:0} \rangle, \quad \mathbb{P}^{\theta_\star}\text{-a.s.} \quad (5.16)$$

and that

$$p^\theta(Y_1 | Y_{-\infty:0}) = g^\theta(\Pi_p(\Psi^\theta \langle Y_{-\infty:0} \rangle); Y_1), \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.} \quad (5.17)$$

In addition, it is shown that for all  $\theta, \theta_\star \in \Theta$ ,

- (i) If  $\theta \neq \theta_*$ ,  $y \mapsto p^\theta(y|Y_{-\infty:0})$  is a density function  $\tilde{\mathbb{P}}^{\theta_*}$ -a.s.
- (ii) Under  $\tilde{\mathbb{P}}^{\theta_*}$ , the function  $y \mapsto p^{\theta_*}(y|Y_{-\infty:0})$  is the conditional density function of  $Y_1$  given  $Y_{-\infty:0}$ .

Under these same conditions, Douc et al. [2014] further showed that the maximizing set  $\Theta_*$  defined in (5.13) indeed reduces to the equivalence class of  $\theta_*$  through the equivalence relation defined in Definition 5.2.3. Then, the convergence in (5.14) yields a so-called *equivalence-class consistency*:

$$\lim_{n \rightarrow \infty} \Delta(\hat{\theta}_{z_1, n}, [\theta_*]) = 0, \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.}, \quad (5.18)$$

for some well-chosen  $z_1 \in \mathbf{Z}$ . From (5.18) the (strong) consistency of the MLE  $\hat{\theta}_{z, n}$  will immediately follow if one can show that  $[\theta_*]$  reduces to the singleton  $\{\theta_*\}$ , which is often referred to as solving *identifiability* problem. In this contribution, this issue is also investigated. We will show that if moreover the probability density kernel  $G^\theta$  and the measurable function  $\Psi^\theta \langle Y_{-\infty:0} \rangle$  defined in (5.15) and (5.16) satisfy some certain conditions (similar to the conditions in [Douc et al., 2013, Proposition 21]), then the strong consistency holds.

### 5.3.2 Convergence of the MLE

We always assume in this part that Assumption (AG-1) holds throughout. Note that every ODM of any order  $p \geq 1$  can be embedded in a first-order ODM. Therefore, the approach used to derive the convergence of the MLE in the class of first-order models can be applied to the class of higher-order ones up to some necessary adaptations. The following is a list of additional assumptions sufficient for obtaining the convergence of the MLE for a class of higher-order ODMs.

(AG-2) There exists a function  $\bar{V} : \mathbf{Z} \rightarrow \mathbb{R}_+$  such that, for all  $\theta \in \Theta$ ,  $\pi_1^\theta(\bar{V}) < \infty$ .

**Remark 5.3.1.** Assumption (AG-2) is usually obtained as a byproduct of the proof of Assumption (AG-1); see Section 5.3.4. It is here stated as an assumption for convenience.

The following set of conditions can readily be checked on  $g^\theta$ ,  $\psi^\theta$  and  $\Psi^\theta$ .

(BG-1) For all  $y \in \mathbf{Y}$ , the function  $(\theta, x) \mapsto g^\theta(x; y)$  is continuous on  $\Theta \times \mathbf{X}$ .

(BG-2) For all  $y_{1:p} \in \mathbf{Y}$ , the function  $(\theta, x_{1:p}) \mapsto \psi_{y_{1:p}}^\theta(x_{1:p})$  is continuous on  $\Theta \times \mathbf{X}^p$ .

The function  $\bar{V}$  appearing in (BG-3)(viii) below is the same one as in Assumption (AG-2). Moreover, in this condition and in what follows, we write

$f \lesssim V$  for a real-valued function  $f$  and a nonnegative function  $V$  defined on the same space  $Z$ , whenever there exists a positive constant  $c$  such that  $|f(z)| \leq cV(z)$  for all  $z \in Z$ .

(BG-3) There exist  $z_1 \in Z$ , a set  $Z_1 \subseteq Z$  such that  $\Pi_p(Z_1)$  is closed,  $\varrho \in (0, 1)$ ,  $C \geq 0$  and measurable functions  $\bar{\Psi} : Z \rightarrow \mathbb{R}_+$ ,  $H : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  and  $\bar{\phi} : Y \rightarrow \mathbb{R}_+$  such that the following assertions hold.

(i) For all  $\theta \in \Theta$  and  $(z, y) \in Z \times Y$ ,  $\Pi_p(\Psi_y^\theta(z)) \in Z_1$ .

(ii)  $\sup_{(\theta, z, y) \in \Theta \times Z_1 \times Y} g^\theta(\Pi_p(z); y) < \infty$ .

(iii) For all  $\theta \in \Theta$ ,  $n \in \mathbb{Z}_+$ ,  $z \in Z$ , and  $y_{1:n} \in Y^n$ ,

$$d_Z(\Psi^\theta\langle y_{1:n} \rangle(z_1), \Psi^\theta\langle y_{1:n} \rangle(z)) \leq \varrho^n \bar{\Psi}(z), \quad (5.19)$$

(iv)  $\bar{\Psi}$  is locally bounded.

(v) For all  $\theta \in \Theta$  and  $y \in Y$ ,  $\bar{\Psi}(\Psi_y^\theta(z_1)) \leq \bar{\phi}(y)$ .

(vi) For all  $\theta \in \Theta$  and  $(z, z', y) \in Z_1 \times Z_1 \times Y$ ,

$$\left| \ln \frac{g^\theta(\Pi_p(z); y)}{g^\theta(\Pi_p(z'); y)} \right| \leq H(d_Z(z, z')) e^{C(d_Z(z_1, z) \vee d_Z(z_1, z'))} \bar{\phi}(y), \quad (5.20)$$

(vii)  $H(u) = O(u)$  as  $u \rightarrow 0$ .

(viii) If  $C = 0$ , then, for all  $\theta \in \Theta$ ,

$$G^\theta \ln^+ \bar{\phi} \lesssim \bar{V}; \quad (5.21)$$

otherwise, for all  $\theta \in \Theta$ ,

$$G^\theta \bar{\phi} \lesssim \bar{V}. \quad (5.22)$$

Let us now state the equivalence-class consistency of the MLE for the higher-order ODMs. The proof can be adapted and easily follows from [Douc et al., 2015, Theorem 1] and is thus omitted here.

**Theorem 5.3.2.** *Assume that (AG-1), (AG-2), (BG-1), (BG-2) and (BG-3) hold. Then, letting  $z_1 \in Z$  as in (BG-3), the convergence (5.18) of the MLE holds.*

The *strong* consistency of the MLE follows from the following theorem.

**Theorem 5.3.3.** *Assume that (AG-1), (AG-2), (BG-1), (BG-2) and (BG-3) hold. Suppose in addition that*

(a) *for all  $\theta = (\vartheta, r), \theta' = (\vartheta', r') \in \Theta$  and  $x, x' \in X$ ,*

$$G^\theta(x; \cdot) = G^{\theta'}(x'; \cdot) \quad \text{implies} \quad r = r' \quad \text{and} \quad x = x',$$

(b) for all  $\theta = (\vartheta, r)$ ,  $\theta_\star = (\vartheta_\star, r_\star) \in \Theta$ ,  $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.,

$$\Pi_p(\Psi^{(\vartheta, r)}\langle Y_{-\infty:0} \rangle) = \Pi_p(\Psi^{(\vartheta_\star, r_\star)}\langle Y_{-\infty:0} \rangle) \quad \text{implies} \quad \vartheta = \vartheta_\star,$$

where  $\Psi^\theta\langle Y_{-\infty:0} \rangle$  is defined by (5.15) and (5.16). Then, letting  $z_1 \in Z$  as in (BG-3),

$$\lim_{n \rightarrow \infty} \Delta(\hat{\theta}_{z_1, n}, \theta_\star) = 0, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.} \quad (5.23)$$

*Proof.* From Theorem 5.3.2, we have  $\Theta_\star = [\theta_\star]$ , where  $\Theta_\star$  is given in (5.13). Now let  $\theta = (\vartheta, r) \in [\theta_\star]$  and write  $\theta_\star = (\vartheta_\star, r_\star)$ . The Kulback-Leibler divergence implies,  $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.,

$$G^{(\vartheta, r)}(\Pi_p(\Psi^{(\vartheta, r)}\langle Y_{-\infty:0} \rangle); \cdot) = G^{(\vartheta_\star, r_\star)}(\Pi_p(\Psi^{(\vartheta_\star, r_\star)}\langle Y_{-\infty:0} \rangle); \cdot).$$

Then from (a), we obtain

$$r = r_\star$$

and

$$\Pi_p(\Psi^{(\vartheta, r)}\langle Y_{-\infty:0} \rangle) = \Pi_p(\Psi^{(\vartheta_\star, r_\star)}\langle Y_{-\infty:0} \rangle), \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

From (b), we also have  $\vartheta = \vartheta_\star$ . Thus the proof follows.  $\square$

### 5.3.3 Asymptotic Normality of the MLE

In this section, we treat a special but an important case of ODMs where the space  $Z$  is a subset of a finite-dimensional vector space;  $\Theta$  is a compact subset of  $\mathbb{R}^d$ , for some positive integer  $d$ ; and the function  $\Psi_y^\theta(z)$  defined in (5.6) admits the following form: for all  $(z, y, \theta) \in Z \times Y \times \Theta$ ,

$$\Psi_y^\theta(z) = a(\theta, y)z + b(\theta, y), \quad (5.24)$$

where  $a(\theta, y)$  and  $b(\theta, y)$  are two matrices with appropriate dimensions, and for all  $\theta \in \Theta$ ,  $y \mapsto a(\theta, y)$  and  $y \mapsto b(\theta, y)$  are measurable maps on  $(Y, \mathcal{Y})$ . The underlying model is assumed to be well-specified and  $\theta_\star \in \Theta$  denotes the true parameter. As in Section 5.3.2, Assumption (AG-1) is supposed to hold throughout this section. Assumption (CG-1) below guarantees the existence of  $\Psi^\theta\langle \cdot \rangle$  satisfying (5.15) and (5.16), and its differentiability on  $\Theta$ . In what follows, we denote by  $|\cdot|$  a vector norm on  $\mathbb{R}^d$  and by  $\|\cdot\|$  a matrix norm.

(CG-1) We have the following.

- (i) For all  $y \in Y$ ,  $\theta \mapsto a(\theta, y)$  and  $\theta \mapsto b(\theta, y)$  are twice continuously differentiable on  $\Theta$ .

- (ii) There exist constants  $C > 0$  and  $\rho \in (0, 1)$  such that for all  $\theta \in \Theta$  and  $y_{1:n} \in \mathbf{Y}$ ,

$$\left\| \prod_{\ell=1}^n a(\theta, y_\ell) \right\| \leq C\rho^n.$$

- (iii)  $\tilde{\mathbb{E}}^{\theta^*}[\sup_{\theta \in \Theta} \|b(\theta, Y_1)\|] < \infty$ .

- (iv) For all  $i \in \{1, \dots, d\}$ ,

$$\tilde{\mathbb{E}}^{\theta^*} \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial a(\theta, Y_1)}{\partial \theta_i} \right\| \right] + \tilde{\mathbb{E}}^{\theta^*} \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial b(\theta, Y_1)}{\partial \theta_i} \right\| \right] < \infty.$$

- (v) For all  $i, j \in \{1, \dots, d\}$ ,

$$\tilde{\mathbb{E}}^{\theta^*} \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial^2 a(\theta, Y_1)}{\partial \theta_i \partial \theta_j} \right\| \right] + \tilde{\mathbb{E}}^{\theta^*} \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial^2 b(\theta, Y_1)}{\partial \theta_i \partial \theta_j} \right\| \right] < \infty.$$

**Lemma 5.3.4.** *Assume (AG-1) and (CG-1). Then for all  $\theta \in \Theta$ , there exists a  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s. finite measurable function  $\Psi^\theta \langle \cdot \rangle : \mathbf{Z}^{\mathbb{Z}^-} \rightarrow \mathbf{Z}$  satisfying (5.15) and (5.16) for all  $z \in \mathbf{Z}$ . The function  $\theta \mapsto \Psi^\theta \langle Y_{-\infty:0} \rangle$  is  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s. twice continuously differentiable on  $\Theta$  and uniformly on  $\Theta$ , we have  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,*

$$\frac{\partial \Psi^\theta \langle Y_{-\infty:0} \rangle}{\partial \theta} = \lim_{m \rightarrow \infty} \frac{\partial \Psi^\theta \langle Y_{-m:0} \rangle(z)}{\partial \theta} \quad (5.25)$$

$$\frac{\partial^2 \Psi^\theta \langle Y_{-\infty:0} \rangle}{\partial \theta^T \partial \theta} = \lim_{m \rightarrow \infty} \frac{\partial^2 \Psi^\theta \langle Y_{-m:0} \rangle(z)}{\partial \theta^T \partial \theta}. \quad (5.26)$$

Moreover,

$$\tilde{\mathbb{E}}^{\theta^*} \left[ \sup_{\theta \in \Theta} \left\| \Psi^\theta \langle Y_{-\infty:0} \rangle \right\| \right] < \infty, \quad (5.27)$$

$$\tilde{\mathbb{E}}^{\theta^*} \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial \Psi^\theta \langle Y_{-\infty:0} \rangle}{\partial \theta} \right\| \right] < \infty, \quad (5.28)$$

$$\tilde{\mathbb{E}}^{\theta^*} \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial^2 \Psi^\theta \langle Y_{-\infty:0} \rangle}{\partial \theta^T \partial \theta} \right\| \right] < \infty. \quad (5.29)$$

*Proof.* The proof is postponed to [Section 5.6.1](#) for convenience.  $\square$

- (CG-2) For all  $y \in \mathbf{Y}$ , the function  $(\theta, x) \mapsto g^\theta(x; y)$  twice continuously differentiable on  $\Theta \times \mathbf{X}$ .

For stating further assumptions, the following notation may be needed. For all  $k \in \mathbb{Z}_+^*$ ,  $\theta \in \Theta$  and  $z \in \mathbf{Z}$ , let

$$\ell_{z,k}^\theta := \ln g^\theta \left( \Pi_p \left( \Psi^\theta \langle Y_{1:k-1} \rangle(z) \right); Y_k \right).$$



Then for all  $n \in \mathbb{Z}_+^*$ , it follows that  $L_{z,n}^\theta$  defined by (5.5) can be rewritten as

$$L_{z,n}^\theta = n^{-1} \sum_{k=1}^n \ell_{z,k}^\theta \quad (5.30)$$

For all  $k \in \mathbb{Z}_+^*$ ,  $\theta \in \Theta$  and  $z \in \mathbb{Z}$ , let

$$\ell_k^\theta := \ln g^\theta \left( \Pi_p \left( \Psi^\theta \langle Y_{-\infty:k-1} \rangle \right); Y_k \right),$$

which is  $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s. well defined by Lemma 5.3.4. Note that from (5.17), we have for all  $k \in \mathbb{Z}_+^*$ ,  $\theta \in \Theta$ ,  $\ell_k^\theta = \ln p^\theta(Y_k | Y_{-\infty:k-1})$ . Now for all  $n \in \mathbb{Z}_+^*$  and  $\theta \in \Theta$ , define

$$L_n^\theta := n^{-1} \sum_{k=1}^n \ell_k^\theta. \quad (5.31)$$

**Remark 5.3.5.** Note that by Lemma 5.3.4, (AG-1), (CG-1) and (CG-2) imply that the functions  $\theta \mapsto \ell_{z,k}^\theta$  and  $\theta \mapsto \ell_k^\theta$  are twice continuously differentiable on some neighborhood  $\mathcal{V}(\theta_\star)$  of  $\theta_\star$ ,  $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.

Assumptions (AG-1), (CG-1) and (CG-2) together with the following set of assumptions, (CG-3)–(CG-7), are sufficient to yield the asymptotic normality of  $\hat{\theta}_{z,n}$  for arbitrary  $z \in \mathbb{Z}$ .

- (CG-3) The true parameter  $\theta_\star$  lies within the interior of  $\Theta$ .
- (CG-4) The strong consistency holds, that is,  $\lim_{n \rightarrow \infty} \hat{\theta}_{z,n} = \theta_\star$ ,  $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.
- (CG-5) There exist a constant  $\alpha > 1/2$  and a  $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s. finite random variable  $C > 0$  such that for all  $k \geq 1$ ,

$$\sup_{\theta \in \mathcal{V}(\theta_\star)} \left| \frac{\partial \ell_{z,k}^\theta}{\partial \theta} - \frac{\partial \ell_k^\theta}{\partial \theta} \right| \leq C k^{-\alpha}, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.},$$

where  $\mathcal{V}(\theta_\star)$  is a neighborhood of  $\theta_\star$ .

- (CG-6) There exists a neighborhood  $\mathcal{V}(\theta_\star)$  of  $\theta_\star$  such that

- (i)  $\tilde{\mathbb{E}}^{\theta_\star} \left[ \sup_{\theta \in \mathcal{V}(\theta_\star)} \left| \frac{\partial \ell_1^\theta}{\partial \theta} \right|^2 \right] < \infty$ .
- (ii)  $\tilde{\mathbb{E}}^{\theta_\star} \left[ \sup_{\theta \in \mathcal{V}(\theta_\star)} \left\| \frac{\partial^2 \ell_1^\theta}{\partial \theta^T \partial \theta} \right\| \right] < \infty$ .
- (iii)  $\tilde{\mathbb{E}}^{\theta_\star} \left[ \sup_{\theta \in \mathcal{V}(\theta_\star)} \left| \frac{\partial}{\partial \theta} g^\theta \left( \Pi_p \left( \Psi^\theta \langle Y_{-\infty:0} \rangle \right); Y_1 \right) \right| \right] < \infty$ .
- (iv)  $\tilde{\mathbb{E}}^{\theta_\star} \left[ \sup_{\theta \in \mathcal{V}(\theta_\star)} \left\| \frac{\partial^2}{\partial \theta^T \partial \theta} g^\theta \left( \Pi_p \left( \Psi^\theta \langle Y_{-\infty:0} \rangle \right); Y_1 \right) \right\| \right] < \infty$ .

- (CG-7) The matrix  $\Sigma_\star = -\tilde{\mathbb{E}}^{\theta_\star} \left[ \frac{\partial^2 \ell_1^{\theta_\star}}{\partial \theta^T \partial \theta} \right]$  is invertible.

Note that if Assumptions (AG-1), (CG-1) and (CG-2) hold, then by Remark 5.3.5,  $\theta \mapsto \mathbf{L}_{z,n}^\theta$  and  $\theta \mapsto \mathbf{L}_n^\theta$  respectively given by (5.30) and (5.31) are twice continuously differentiable on some neighborhood  $\mathcal{V}(\theta_*)$  of  $\theta_*$ ,  $\tilde{\mathbb{P}}^{\theta_*}$ -a.s. If in addition Assumption (CG-5) holds, the following fact is obtained.

**Lemma 5.3.6.** *Assume (AG-1), (CG-1), (CG-2) and (CG-5), then*

$$\lim_{n \rightarrow \infty} \sup_{\theta \in \mathcal{V}(\theta_*)} n^{1/2} \left| \frac{\partial \mathbf{L}_{z,n}^\theta}{\partial \theta} - \frac{\partial \mathbf{L}_n^\theta}{\partial \theta} \right| = 0, \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.},$$

*Proof.* We have from (CG-5),  $\tilde{\mathbb{P}}^{\theta_*}$ -a.s.,

$$\sup_{\theta \in \mathcal{V}(\theta_*)} n^{1/2} \left| \frac{\partial \mathbf{L}_{z,n}^\theta}{\partial \theta} - \frac{\partial \mathbf{L}_n^\theta}{\partial \theta} \right| \leq \frac{1}{n^{1/2}} \sum_{k=1}^n \sup_{\theta \in \mathcal{V}(\theta_*)} \left| \frac{\partial \ell_{z,k}^\theta}{\partial \theta} - \frac{\partial \ell_k^\theta}{\partial \theta} \right| \leq \frac{C}{n^{1/2}} \sum_{k=1}^n \frac{1}{k^\alpha}.$$

To complete, observe that the rightmost term of the above inequalities converges to 0 as  $n \rightarrow \infty$  whenever  $\alpha > 1/2$ .  $\square$

**Theorem 5.3.7.** *Assume that (AG-1) and (CG-1)–(CG-7) hold. Then, for any  $z \in \mathbf{Z}$ ,*

$$\sqrt{n}(\hat{\theta}_{z,n} - \theta_*) \rightsquigarrow_{\theta_*} \mathcal{N}(0, \Sigma_\star^{-1}), \quad (5.32)$$

where the symbol  $\rightsquigarrow_{\theta_*}$  means weak convergence under  $\tilde{\mathbb{P}}^{\theta_*}$ , and  $\mathcal{N}(0, \Sigma_\star^{-1})$  stands for the centered Gaussian distribution with covariance matrix  $\Sigma_\star^{-1}$ .

*Proof of Theorem 5.3.7.* From Lemma 5.3.6, we have

$$\lim_{n \rightarrow \infty} n^{1/2} \sup_{\theta \in \mathcal{V}(\theta_*)} \left| \frac{\partial \mathbf{L}_{z,n}^\theta}{\partial \theta} - \frac{\partial \mathbf{L}_n^\theta}{\partial \theta} \right| = 0, \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.}$$

Since for sufficiently large  $n$ ,  $\hat{\theta}_{z,n} \in \mathcal{V}(\theta_*)$  and  $\frac{\partial \mathbf{L}_{z,n}^{\hat{\theta}_{z,n}}}{\partial \theta} = 0$ , then

$$n^{1/2} \frac{\partial \mathbf{L}_n^{\hat{\theta}_{z,n}}}{\partial \theta} = n^{1/2} \frac{\partial \mathbf{L}_{z,n}^{\hat{\theta}_{z,n}}}{\partial \theta} + o_P(1) = o_P(1).$$

Note that we also have,

$$n^{1/2} \frac{\partial \mathbf{L}_n^{\hat{\theta}_{z,n}}}{\partial \theta} = n^{1/2} \frac{\partial \mathbf{L}_n^{\theta_*}}{\partial \theta} + \left( \frac{\partial^2 \mathbf{L}_n^{\theta_*}}{\partial \theta^T \partial \theta} + \varepsilon_n \right) n^{1/2} (\hat{\theta}_{z,n} - \theta_*),$$

where

$$\varepsilon_n = \int_0^1 \left( \frac{\partial^2 \mathbf{L}_n^{\theta_* + t(\hat{\theta}_{z,n} - \theta_*)}}{\partial \theta^T \partial \theta} - \frac{\partial^2 \mathbf{L}_n^{\theta_*}}{\partial \theta^T \partial \theta} \right) dt. \quad (5.33)$$

Thus,

$$n^{1/2} \frac{\partial \mathbf{L}_n^{\theta_*}}{\partial \theta} = - \left( \frac{\partial^2 \mathbf{L}_n^{\theta_*}}{\partial \theta^T \partial \theta} + \varepsilon_n \right) n^{1/2} (\hat{\theta}_{z,n} - \theta_*) + o_P(1).$$

Note that from (CG-6)(ii) and by ergodicity, Birkhoff ergodic theorem applies and we have

$$\lim_{n \rightarrow \infty} \frac{\partial^2 \mathbf{L}_n^{\theta_*}}{\partial \theta^T \partial \theta} = \lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n \frac{\partial^2 \ell_k^{\theta_*}}{\partial \theta^T \partial \theta} = -\Sigma_*, \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.}$$

Thus to complete the proof, we only need to show that

- (a)  $n^{1/2} \frac{\partial \mathbf{L}_n^{\theta_*}}{\partial \theta} \rightsquigarrow_{\theta_*} \mathcal{N}(0, \Sigma_*)$  and
- (b)  $\varepsilon_n = o_P(1)$ .

Let us now we show (a). Note that we can write

$$n^{1/2} \frac{\partial \mathbf{L}_n^{\theta_*}}{\partial \theta} = n^{-1/2} \sum_{k=1}^n \frac{\partial \ell_k^{\theta_*}}{\partial \theta}.$$

Let  $\xi_k = \frac{\partial \ell_k^{\theta_*}}{\partial \theta}$ ,  $k \in \mathbb{Z}_+$ . From (CG-6)(iii), it readily follows that

$$\begin{aligned} \tilde{\mathbb{E}}^{\theta_*} [\xi_k | Y_{-\infty:k-1}] &= \int_{\mathcal{Y}} \frac{\partial g^{\theta_*}(\Pi_p(\Psi^{\theta_*} \langle Y_{-\infty:k-1} \rangle); y)}{\partial \theta} \nu(dy) \\ &= \frac{\partial}{\partial \theta} \int_{\mathcal{Y}} g^{\theta}(\Pi_p(\Psi^{\theta} \langle Y_{-\infty:k-1} \rangle); y) \nu(dy) \Big|_{\theta=\theta_*} \\ &= \frac{\partial}{\partial \theta} 1 \Big|_{\theta=\theta_*} = 0, \end{aligned}$$

so that the sequence  $\{\xi_k : k \in \mathbb{Z}_+\}$  is a stationary and ergodic Martingale difference with respect to  $\mathcal{F}_{-\infty:k} = \sigma(Y_{-\infty:k})$  in  $L_2(\tilde{\mathbb{P}}^{\theta_*})$ . By (CG-6) and (CG-7), its covariance matrix is equal to  $\Sigma_*$ . To see this, observe that  $\tilde{\mathbb{P}}^{\theta_*}$ -a.s.,

$$\begin{aligned} &\tilde{\mathbb{E}}^{\theta_*} [\xi_k \xi_k^T | \mathcal{F}_{-\infty:k-1}] \\ &= \int_{\mathcal{Y}} \frac{1}{g^{\theta_*}(\psi_k^{\theta_*}; y)} \left( \frac{\partial g^{\theta_*}(\psi_k^{\theta_*}; y)}{\partial \theta} \right) \left( \frac{\partial g^{\theta_*}(\psi_k^{\theta_*}; y)}{\partial \theta} \right)^T \nu(dy) \\ &= \int_{\mathcal{Y}} \left( \frac{\partial^2 g^{\theta_*}(\psi_k^{\theta_*}; y)}{\partial \theta^T \partial \theta} - \frac{\partial^2 \ln g^{\theta_*}(\psi_k^{\theta_*}; y)}{\partial \theta^T \partial \theta} g^{\theta_*}(\psi_k^{\theta_*}; y) \right) \nu(dy), \end{aligned}$$

with  $\psi_k^{\theta_\star} := \Pi_p(\Psi^{\theta_\star} \langle Y_{-\infty:k-1} \rangle)$ . By (CG-6)(iv), we have

$$\int_{\mathcal{Y}} \frac{\partial^2 g^{\theta_\star}(\psi_k^{\theta_\star}; y)}{\partial \theta^T \partial \theta} = 0, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

Since  $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.,

$$\tilde{\mathbb{E}}^{\theta_\star} \left[ \frac{\partial^2 \ell_k^{\theta_\star}}{\partial \theta^T \partial \theta} \middle| \mathcal{F}_{-\infty:k-1} \right] = \int_{\mathcal{Y}} \frac{\partial^2 \ln g^{\theta_\star}(\psi_k^{\theta_\star}; y)}{\partial \theta^T \partial \theta} g^{\theta_\star}(\psi_k^{\theta_\star}; y) \nu(dy),$$

it then follows that

$$\tilde{\mathbb{E}}^{\theta_\star} [\xi_k \xi_k^T \middle| \mathcal{F}_{-\infty:k-1}] = -\tilde{\mathbb{E}}^{\theta_\star} \left[ \frac{\partial^2 \ell_k^{\theta_\star}}{\partial \theta^T \partial \theta} \middle| \mathcal{F}_{-\infty:k-1} \right], \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

Taking the expectation on both sides, then using tower property and using (CG-7), the claim thus follows. By ergodicity and from (CG-6)(i), we have

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n \tilde{\mathbb{E}}^{\theta_\star} [\xi_k \xi_k^T \middle| \mathcal{F}_{-\infty:k-1}] = \Sigma_\star, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

Moreover, for  $\epsilon > 0$ ,

$$\begin{aligned} & \lim_{n \rightarrow \infty} n^{-1} \left\| \sum_{k=1}^n \tilde{\mathbb{E}}^{\theta_\star} [\xi_k \xi_k^T \mathbb{1}(|\xi_k| > \epsilon \sqrt{n}) \middle| \mathcal{F}_{-\infty:k-1}] \right\| \\ & \leq \lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n \tilde{\mathbb{E}}^{\theta_\star} [|\xi_k|^2 \mathbb{1}(|\xi_k| > \epsilon \sqrt{n}) \middle| \mathcal{F}_{-\infty:k-1}] \\ & \leq \lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n \tilde{\mathbb{E}}^{\theta_\star} [|\xi_k|^2 \mathbb{1}(|\xi_k| > M) \middle| \mathcal{F}_{-\infty:k-1}] \\ & = \tilde{\mathbb{E}}^{\theta_\star} [|\xi_1|^2 \mathbb{1}(|\xi_1| > M)], \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}, \end{aligned}$$

where  $M > 0$ . Since  $M$  is arbitrary, letting  $M \rightarrow \infty$ , we have

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n \tilde{\mathbb{E}}^{\theta_\star} [\xi_k \xi_k^T \mathbb{1}(|\xi_k| > \epsilon \sqrt{n}) \middle| \mathcal{F}_{-\infty:k-1}] = 0, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.}$$

Thus by martingale central limit theorem, see, for instance, [Hall and Heyde \[1980\]](#),

$$n^{1/2} \frac{\partial \mathcal{L}_n^{\theta_\star}}{\partial \theta} = n^{-1/2} \sum_{k=1}^n \xi_k \rightsquigarrow_{\theta_\star} \mathcal{N}(0, \Sigma_\star).$$

Now it remains to show (b). Let  $\rho > 0$  such that  $\mathbf{B}(\theta_*, \rho) \subset \mathcal{V}(\theta_*)$ . Since  $\tilde{\mathbb{P}}^{\theta_*}$ -a.s.,  $\hat{\theta}_{z,n} \rightarrow \theta_*$  as  $n \rightarrow \infty$ , then there exists a  $\tilde{\mathbb{P}}^{\theta_*}$ -a.s. finite integer  $N_0$  such that for all  $n \geq N_0$ ,  $\tilde{\mathbb{P}}^{\theta_*}$ -a.s.,  $\theta_* + t(\hat{\theta}_{z,n} - \theta_*) \in \mathbf{B}(\theta_*, \rho)$ . Then from (5.33),  $\tilde{\mathbb{P}}^{\theta_*}$ -a.s., for all  $n \geq N_0$ ,

$$\begin{aligned} \|\varepsilon_n\| &\leq \int_0^1 \left\| \frac{\partial^2 \mathbf{L}_n^{\theta_* + t(\hat{\theta}_{z,n} - \theta_*)}}{\partial \theta^T \partial \theta} - \frac{\partial^2 \mathbf{L}_n^{\theta_*}}{\partial \theta^T \partial \theta} \right\| dt \\ &\leq \sup_{\theta \in \mathbf{B}(\theta_*, \rho)} \left\| \frac{\partial^2 \mathbf{L}_n^\theta}{\partial \theta^T \partial \theta} - \frac{\partial^2 \mathbf{L}_n^{\theta_*}}{\partial \theta^T \partial \theta} \right\| \\ &\leq n^{-1} \sum_{k=1}^n \sup_{\theta \in \mathbf{B}(\theta_*, \rho)} \left\| \frac{\partial^2 \ell_k^\theta}{\partial \theta^T \partial \theta} - \frac{\partial^2 \ell_k^{\theta_*}}{\partial \theta^T \partial \theta} \right\|. \end{aligned}$$

By (CG-6)(ii),  $\mathbf{B}(\theta_*, \rho) \subset \mathcal{V}(\theta_*)$  and Birkhoff ergodic theorem, we get that

$$\limsup_{n \rightarrow \infty} \|\varepsilon_n\| \leq \tilde{\mathbb{E}}^{\theta_*} \left[ \sup_{\theta \in \mathbf{B}(\theta_*, \rho)} \left\| \frac{\partial^2 \ell_1^\theta}{\partial \theta^T \partial \theta} - \frac{\partial^2 \ell_1^{\theta_*}}{\partial \theta^T \partial \theta} \right\| \right], \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.}$$

Here  $\rho$  is an arbitrary positive number, provided that  $\mathbf{B}(\theta_*, \rho) \subset \mathcal{V}(\theta_*)$ . By dominated convergence and Remark 5.3.5, the right-hand side of the previous display can be made arbitrarily close to zero by letting  $\rho$  tend to zero. Hence we get (b), and the proof is complete.  $\square$

**Remark 5.3.8.** From the proof of Theorem 5.3.7, it is also shown that

$$\Sigma_* = \tilde{\mathbb{E}}^{\theta_*} \left[ \left( \frac{\partial \ell_1^{\theta_*}}{\partial \theta} \right) \left( \frac{\partial \ell_1^{\theta_*}}{\partial \theta} \right)^T \right].$$

Thus, to obtain (CG-7), that is showing that  $\Sigma_*$  is invertible, it is sufficient to show that for all  $\mathbf{a} \in \mathbb{R}^d$ ,

$$\mathbf{a}^T \frac{\partial \ell_1^{\theta_*}}{\partial \theta} = 0, \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.}, \implies \mathbf{a} = 0. \quad (5.34)$$

### 5.3.4 Ergodicity

In this section, we provide conditions that yield stationarity and ergodicity of the Markov chain  $\{(Z_k, Y_k) : k \in \mathbb{Z}_+\}$ . We will set  $\theta$  to be an arbitrary value in  $\Theta$  and since this is a “for all  $\theta$  (...)” condition, to save space and alleviate the notational burden, we will drop the superscript  $\theta$  from, for example,  $G^\theta$ ,  $R^\theta$  and  $\psi^\theta$  and respectively write  $G$ ,  $R$  and  $\psi$ , instead.

Ergodicity of Markov chains are usually studied using  $\varphi$ -irreducibility. This approach is well known to be quite efficient when dealing with fully dominated models; see [Meyn and Tweedie \[2009\]](#). It is not at all the same picture for integer-valued observation-driven models, where other tools need to be invoked; see [Fokianos and Tjøstheim \[2011\]](#), [Douc et al. \[2013, 2015\]](#). Our result here is obtained in the same fashion as in [Douc et al. \[2015\]](#), which is inspired by the approach in [Douc et al. \[2013\]](#). The following is the list of assumptions to obtain our main result on ergodicity. We should mention that some of our assumptions (for instance, Assumption [\(AG-4\)](#)) are more general than those derived in [Douc et al. \[2013\]](#) and [Douc et al. \[2015\]](#).

- (AG-3) The measurable space  $(Z, d_Z)$  is a locally compact, complete and separable metric space and its associated  $\sigma$ -field  $\mathcal{Z}$  is the Borel  $\sigma$ -field.
- (AG-4) There exists a positive integer  $q$  such that the Markov kernel  $R^q$  is weak Feller, that is, for any continuous and bounded function  $f$  defined on  $X$ ,  $R^q f$  is continuous and bounded on  $X$ . Moreover, there exist  $(\lambda, \beta) \in (0, 1) \times \mathbb{R}_+$  and a measurable function  $V : Z \rightarrow \mathbb{R}_+$  such that  $R^q V \leq \lambda V + \beta$  and  $\{V \leq M\}$  is a compact set for any  $M > 0$ .
- (AG-5) The Markov kernel  $R$  admits a reachable point, that is, there exists  $z_0 \in Z$  such that, for any  $z \in Z$  and any neighborhood  $\mathcal{N}$  of  $z_0$ ,  $R^m(z; \mathcal{N}) > 0$  for at least one positive integer  $m$ .
- (AG-6) There exists a positive integer  $\ell$  such that

$$\sup_{\substack{(z, z', y_{1:\ell}) \in Z^2 \times Y^\ell \\ z \neq z'}} \frac{d_Z(\Psi_{\langle y_{1:\ell} \rangle}(z), \Psi_{\langle y_{1:\ell} \rangle}(z'))}{d_Z(z, z')} < 1, \quad (5.35)$$

and

$$\sup_{\substack{(z, z', y) \in Z^2 \times Y \\ z \neq z'}} \frac{d_Z(\Psi_y(z), \Psi_y(z'))}{d_Z(z, z')} < \infty. \quad (5.36)$$

- (AG-7) There exist a measurable function  $\alpha$  from  $Z^2$  to  $[0, 1]$ , a measurable function  $\phi : Z^2 \rightarrow X = \Pi_p(Z)$  and a measurable function  $W : Z^2 \rightarrow [1, \infty)$  such that the following assertions hold.

- (i) For all  $(z, z') \in Z^2$  and  $y \in Y$ ,

$$\min \{g(\Pi_p(z); y), g(\Pi_p(z'); y)\} \geq \alpha(z, z') g(\phi(z, z'); y). \quad (5.37)$$

- (ii) For all  $z \in Z$ ,  $W(z, \cdot)$  is finitely bounded in a neighborhood of  $z$ , that is, there exists  $\gamma_z > 0$  such that  $\sup_{z' \in B(z, \gamma_z)} W(z, z') < \infty$ .

- (iii) For all  $(z, z') \in \mathbf{Z}^2$ ,  $1 - \alpha(z, z') \leq d_{\mathbf{Z}}(z, z')W(z, z')$ .  
 (iv) There exist an integer  $m > 0$  and a real number  $D > 0$  such that

$$\sup_{(z, z') \in \mathbf{Z}^2} (J_m((z, z'); W) - W(z, z')) < \infty$$

and

$$\sup_{(z, z') \in \mathbf{Z}^2} (J_1((z, z'); W) - DW(z, z')) < \infty,$$

where for any positive integer  $n$ ,  $(z, z') \in \mathbf{Z}^2$  and measurable function  $f : \mathbf{Z}^2 \rightarrow \mathbb{R}$ ,

$$J_n((z, z'); f) := \int f(\Psi\langle y_{1:n} \rangle(z), \Psi\langle y_{1:n} \rangle(z')) \times \prod_{k=1}^n G(\phi(\Psi\langle y_{1:k-1} \rangle(z), \Psi\langle y_{1:k-1} \rangle(z')); dy_k).$$

**Remark 5.3.9.** Assumption **(AG-3)** holds if the measurable metric spaces  $(\mathbf{X}, d_{\mathbf{X}})$  and  $(\mathbf{Y}, d_{\mathbf{Y}})$  are locally compact, complete and separable and their associated  $\sigma$ -field  $\mathcal{X}$  and  $\mathcal{Y}$  are the Borel  $\sigma$ -fields. In this case, the metric  $d_{\mathbf{Z}}$  can be appropriately defined as a function of  $d_{\mathbf{X}}$  and  $d_{\mathbf{Y}}$ .

**Remark 5.3.10.** Note that if the Markov kernel  $R$  is weak Feller, then the Markov kernel  $R^q$  is also weak Feller.

**Lemma 5.3.11.** *Assume **(AG-3)** and **(AG-4)**. Then  $R$  admits an invariant probability distribution  $\tilde{\pi}$ ; moreover,  $\tilde{\pi}V < \infty$ .*

*Proof.* From Assumptions **(AG-3)** and **(AG-4)** and by **Remark 5.3.10**, the transition kernel  $R^q$  admits an invariant probability distribution denoted by  $\pi_q$ . Let  $\tilde{\pi}$  be defined by, for all  $A \in \mathcal{Z}$ ,

$$\tilde{\pi}(A) = \frac{1}{q} \sum_{k=1}^q \pi_q R^k(A).$$

Obviously, we have  $\tilde{\pi}R = \tilde{\pi}$ , which shows that  $R$  admits an invariant probability distribution  $\tilde{\pi}$ . Now let  $M > 0$ . Then we have for all  $n \in \mathbb{Z}_+$ ,

$$\begin{aligned} \tilde{\pi}(V \wedge M) &= \tilde{\pi}R^{nq}(V \wedge M) \\ &\leq \tilde{\pi}R^{nq}(V) \wedge M \\ &\leq \lambda^n \tilde{\pi}(V) \wedge M + \frac{\beta}{1 - \lambda} \wedge M. \end{aligned}$$

Letting  $n \rightarrow \infty$ , we then obtain

$$\tilde{\pi}(V \wedge M) \leq \frac{\beta}{1-\lambda} \wedge M.$$

Finally, since  $M$  is arbitrary, by letting  $M \rightarrow \infty$  and by monotone convergence theorem, we get  $\tilde{\pi}V < \infty$  as desired.  $\square$

We can now state the main ergodicity result.

**Theorem 5.3.12.** *Conditions (AG-3), (AG-4), (AG-5), (AG-6) and (AG-7) imply that  $K$  admits a unique stationary distribution  $\pi$  on  $Z \times Y$ . Moreover,  $\pi_1 \bar{V} < \infty$  for every  $\bar{V} : X \rightarrow \mathbb{R}_+$  such that  $\bar{V} \lesssim V$ .*

For convenience, we postponed the proof of [Theorem 5.3.12](#) to [Section 5.6.2](#). The first conclusion of [Theorem 5.3.12](#) can directly be applied for all  $\theta \in \Theta$  to check [\(AG-1\)](#).

Assumptions [\(AG-4\)](#) and [\(AG-5\)](#) have to be checked directly on the Markov kernel  $R$  defined by [\(5.11\)](#). To this end, it can be useful to define, for any given  $z \in Z$ , the distribution

$$\bar{\mathbb{P}}_z := \mathbb{P}_{\delta_z \otimes G(\Pi_p(z); \cdot)} \quad (5.38)$$

on  $(Z \times Y)^{\mathbb{Z}_+}$ , where  $\mathbb{P}_\xi$  is defined for any distribution  $\xi$  on  $Z \times Y$  as in [Definition 5.2.3](#). Then the first component process  $\{Z_k : k \in \mathbb{Z}_+\}$  associated to  $\bar{\mathbb{P}}_z$  is a Markov chain with Markov kernel  $R$  and initial distribution  $\delta_z$ .

The following lemma provides a general way for constructing the instrumental functions  $\alpha$  and  $\phi$  that appear in [\(AG-7\)](#). The proof can be easily adapted from [[Douc et al., 2015](#), Lemma 1] and is thus omitted.

**Lemma 5.3.13.** *Suppose that  $\Pi_p(Z) = X = \mathbb{C}^S$  for some measurable space  $(S, \mathcal{S})$  and  $\mathbb{C} \subseteq \mathbb{R}$ . Thus for all  $x \in X$ , we write  $x = (x_s)_{s \in S}$ , where  $x_s \in \mathbb{C}$  for all  $s \in S$ . Suppose moreover that for all  $x = (x_s)_{s \in S} \in X$ , we can express the conditional density  $g(x; \cdot)$  as a mixture of densities of the form  $j(x_s)h(x_s; \cdot)$  over  $s \in S$ . This means that for all  $t \in \mathbb{C}$ ,  $y \mapsto j(t)h(t; y)$  is a density with respect to  $\nu$  and there exists a probability measure  $\mu$  on  $(S, \mathcal{S})$  such that*

$$g(x; y) = \int_S j(x_s)h(x_s; y)\mu(ds), \quad y \in Y. \quad (5.39)$$

We moreover assume that  $h$  takes nonnegative values and that one of the two following assumptions holds.

- (H'-1) For all  $y \in Y$ , the function  $h(\cdot; y) : t \mapsto h(t; y)$  is nondecreasing.
- (H'-2) For all  $y \in Y$ , the function  $h(\cdot; y) : t \mapsto h(t; y)$  is nonincreasing.



For all  $(x, x') \in \mathbf{X}^2$ , we denote  $x \wedge x' := (\min\{x_s, x'_s\})_{s \in \mathcal{S}}$  and  $x \vee x' := (\max\{x_s, x'_s\})_{s \in \mathcal{S}}$  and we define  $\alpha(z, z')$  and  $\phi(z, z')$  by: for all  $(z, z') \in \mathbf{Z}^2$ , letting  $x = \Pi_p(z)$  and  $x' = \Pi_p(z')$ ,

$$\begin{cases} \alpha(z, z') = \inf_{s \in \mathcal{S}} \left\{ \frac{j(x_s \vee x'_s)}{j(x_s \wedge x'_s)} \right\} & \text{and } \phi(z, z') = x \wedge x' & \text{under (H'-1);} \\ \alpha(z, z') = \inf_{s \in \mathcal{S}} \left\{ \frac{j(x_s \wedge x'_s)}{j(x_s \vee x'_s)} \right\} & \text{and } \phi(z, z') = x \vee x' & \text{under (H'-2).} \end{cases}$$

Then  $\alpha$  and  $\phi$  defined above satisfy (AG-7)(i).

## 5.4 Examples

In this section, we apply our main results derived above to three models of interest, namely, the classical GARCH( $p, p$ ), the log-linear Poisson GARCH( $p, p$ ) and the NBIN-GARCH( $p, p$ ) models. For the celebrated GARCH( $p, p$ ) model, however, we illustrate only the ergodicity result and for the asymptotic result of the MLE for this model, we refer to [Francq et al. \[2004\]](#). To the best of our knowledge, the stationarity and ergodicity as well as the asymptotic properties of the MLE for the general log-linear Poisson GARCH( $p, p$ ) and NBIN-GARCH( $p, p$ ) models have not been derived so far. In the sequel,  $\mathbf{A}$ ,  $\mathbf{b}$  and  $\boldsymbol{\omega}$  always denote, respectively, the matrix

$$\begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 \\ a_p & a_{p-1} & a_{p-2} & a_{p-3} & \cdots & a_1 & b_p & b_{p-1} & \cdots & b_3 & b_2 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix},$$

the vector  $b_1 \boldsymbol{\delta}_p + \boldsymbol{\delta}_{2p-1}$ , with for all  $j \in \{1, \dots, 2p-1\}$ ,  $\boldsymbol{\delta}_j = (\delta_{ij})_{i \in \{1, \dots, 2p-1\}}$ ,  $\delta_{ij}$  being the Kronecker delta, and the vector  $\boldsymbol{\omega} \boldsymbol{\delta}_p$ . We also denote by  $\mathbf{A}_*$ ,  $\mathbf{b}_*$  and  $\boldsymbol{\omega}_*$  the corresponding values of  $\mathbf{A}$ ,  $\mathbf{b}$  and  $\boldsymbol{\omega}$  at  $(\omega, a_{1:p}, b_{1:p}) = (\omega^*, a_{1:p}^*, b_{1:p}^*)$ , respectively. For any matrix  $\mathbf{M}$ , we denote its transpose by  $\mathbf{M}^T$ . For any real numbers  $a$  and  $b$ , we denote their maximum value by  $a \vee b$  or  $\max\{a, b\}$  and their minimum value by  $a \wedge b$  or  $\min\{a, b\}$ .

### 5.4.1 GARCH( $p, p$ ) Model

**Example 5.4.1** (Definition of GARCH( $p, p$ )). Consider GARCH( $p, p$ ),  $p \in \mathbb{Z}_+^*$ , defined as follows: for all  $k \in \mathbb{Z}_+$ ,

$$\begin{aligned} X_{k+1} &= \omega + \sum_{i=1}^p a_i X_{k-i+1} + \sum_{i=1}^p b_i Y_{k-i+1}^2, \\ Y_{k+1} | \mathcal{F}_k &= X_{k+1}^{1/2} \varepsilon_{k+1}, \end{aligned} \quad (5.40)$$

where

- the parameter space

$$\Theta \subset \{\theta = (\omega, a_{1:p}, b_{1:p}) \in \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^p : \omega > 0, a_{1:p} \geq 0, b_{1:p} \geq 0\},$$

- $\mathcal{F}_k = \sigma(X_{-p+1:k+1}, Y_{-p+1:k})$ ,
- The process  $\{\varepsilon_k : k \in \mathbb{Z}_+\}$  are i.i.d. random sequence with zero mean and unit variance on  $\mathbb{R}$  and are independent of  $\mathcal{F}_k$ . For each  $k \in \mathbb{Z}_+$ , the distribution of  $\varepsilon_k$  admits a density  $\gamma$  with respect to the Lebesgue measure on  $\mathbb{R}$ . The density  $\gamma$  is assumed to be positive, to be continuous over  $\mathbb{R}$  and to satisfy that  $\gamma(u) + \gamma(-u)$  is monotone on  $u \in \mathbb{R}_+$ . We further assume that  $\epsilon > 0$ ,  $\mathbb{P}(\varepsilon_0^2 < \epsilon) > 0$ .
- $\mathbf{X} = (0, \infty)$ ,  $\mathcal{X} = \mathcal{B}(\mathbf{X})$ ,  $\mathbf{Y} = \mathbb{R}$ ,  $\mathcal{Y} = \mathcal{B}(\mathbf{Y})$  and the metrics on  $\mathbf{X}$  and  $\mathbf{Y}$  are any metrics on  $\mathbb{R}$ .

Note that, by setting for all  $k$ ,  $U_k = Y_k^2$  (then  $\mathbf{U} = \mathbb{R}_+$ ,  $\mathcal{U} = \mathcal{B}(\mathbf{U})$  and  $d_{\mathbf{U}}$  is any metric on  $\mathbb{R}$ ), Model (5.40) can be embedded in Model (5.41) below:

$$\begin{aligned} X_{k+1} &= \psi_{U_{k-p+1:k}}^\theta(X_{k-p+1:k}), \\ U_{k+1} | \mathcal{F}_k &= X_{k+1} \varepsilon_{k+1}^2, \end{aligned} \quad (5.41)$$

where for all  $(x_{1:p}, u_{1:p}) \in \mathbf{X}^p \times \mathbf{U}^p$ ,

$$\psi_{u_{k-p+1:k}}^\theta(x_{k-p+1:k}) := \omega + \sum_{i=1}^p a_i x_{p-i+1} + \sum_{i=1}^p b_i u_{p-i+1}. \quad (5.42)$$

The density  $g^\theta$  and the probability law  $G^\theta$  of  $U_k$  conditional on  $X_k = x$  are respectively given by, for all  $u \in \mathbf{U}$  and  $A \in \mathcal{U}$ ,

$$g(x; u) = \frac{1}{2\sqrt{xu}} \left[ \gamma\left(\sqrt{x^{-1}u}\right) + \gamma\left(-\sqrt{x^{-1}u}\right) \right] > 0, \quad (5.43)$$

and

$$G^\theta(x; A) = \int_{\mathbf{U}} \mathbb{1}_A(u) g^\theta(x; u) du, \quad (5.44)$$

with  $g^\theta(x; 0) = 1$  for all  $x \in \mathbf{X}$ . In view of [Definition 5.2.1](#), Model [\(5.41\)](#) is an ODM( $p, p$ ) and is dominated by the Lebesgue measure on  $\mathbb{R}$ . Now let  $\mathbf{Z} = \mathbf{X}^p \times \mathbf{U}^{p-1}$  and  $\mathcal{Z} = \mathcal{X}^{\otimes p} \otimes \mathcal{U}^{\otimes(p-1)}$  and define the measurable function  $\Psi^\theta$  similarly as in [\(5.6\)](#). The function  $\Psi^\theta$  then has the form: for all  $(z, u) \in \mathbf{Z} \times \mathbf{U}$ ,

$$\Psi_u^\theta(z) = \boldsymbol{\omega} + \mathbf{A}z + u\mathbf{b}. \quad (5.45)$$

Moreover, denoting  $Z_k = (X_{k-p+1:k}, U_{k-p+1:k-1})$  for all  $k$ , the process  $\{Z_k : k \in \mathbb{Z}_+\}$  is a Markov chain on  $(\mathbf{Z}, \mathcal{Z})$  whose transition kernel  $R^\theta$  is given by [\(5.11\)](#) with  $G^\theta$  and  $\Psi^\theta$  defined in [\(5.44\)](#) and [\(5.45\)](#), respectively.

Let  $\mathbf{I}_n$  be the identity matrix of order  $n$  and denote the characteristic polynomial of the matrix  $\mathbf{A} + \mathbf{b}\boldsymbol{\delta}_p^T$  by  $P(\lambda) = \det(\lambda\mathbf{I}_{2p-1} - (\mathbf{A} + \mathbf{b}\boldsymbol{\delta}_p^T))$ . For any square matrix  $\mathbf{M}$ , denote the spectral radius of  $\mathbf{M}$  by  $|\lambda|_{\max}(\mathbf{M})$ . Straightforward computation shows that

$$P(\lambda) = \lambda^{p-1} \left( \lambda^p - \sum_{k=1}^p (a_k + b_k) \lambda^{p-k} \right). \quad (5.46)$$

The following lemma is immediate.

**Lemma 5.4.2.** *Let  $a_{1:p} \geq 0$ ,  $b_{1:p} \geq 0$ . Then  $\sum_{i=1}^p (a_i + b_i) < 1$  if and only if*

$$|\lambda|_{\max}(\mathbf{A} + \mathbf{b}\boldsymbol{\delta}_p^T) < 1.$$

*Proof.* The proof is postponed to [Section 5.6.3](#). □

**Remark 5.4.3.** It is well known that for any nonnegative matrices  $\mathbf{M} = (m_{ij})$  and  $\mathbf{N} = (n_{ij})$  such that  $\mathbf{M} \geq \mathbf{N}$ , that is for all  $i, j$ ,  $m_{ij} \geq n_{ij}$ , we have  $|\lambda|_{\max}(\mathbf{M}) \geq |\lambda|_{\max}(\mathbf{N})$ . Thus for our matrices  $\mathbf{A}$  and  $\mathbf{A} + \mathbf{b}\boldsymbol{\delta}_p^T$ , we have  $|\lambda|_{\max}(\mathbf{A}) \leq |\lambda|_{\max}(\mathbf{A} + \mathbf{b}\boldsymbol{\delta}_p^T)$ .

**Proposition 5.4.4.** *Suppose that  $\theta \in \Theta$  satisfies  $|\lambda|_{\max}(\mathbf{A} + \mathbf{b}\boldsymbol{\delta}_p^T) < 1$ . Then the conclusion of [Theorem 5.3.12](#) holds with  $\bar{V}(z) = \mathbf{1}^T z$  for all  $z \in \mathbf{Z}$ .*

**Remark 5.4.5.** If  $a_i \geq 0$ ,  $b_i \geq 0$ , for all  $i \in \{1, \dots, p\}$  and  $\sum_{i=1}^p (a_i + b_i) < 1$ , then from [Lemma 5.4.2](#), [Proposition 5.4.4](#) holds with  $\bar{V}(z) = \mathbf{1}^T z$  for all  $z \in \mathbf{Z}$ .

*Proof.* Now set  $\bar{V}(z) = V(z) = \mathbf{1}^T z$ ,  $z \in \mathbf{Z}$ . From [Theorem 5.3.12](#), we need to show that [\(AG-3\)](#), [\(AG-4\)](#), [\(AG-5\)](#), [\(AG-6\)](#) and [\(AG-7\)](#) hold. For convenience, we drop the superscript  $\theta$  from for example  $R^\theta$ ,  $\Psi^\theta$ , etc. Assumption [\(AG-3\)](#) holds with any metric  $d_{\mathbf{Z}}$  associated to a norm on the finite dimensional space  $\mathbf{Z}$ . (The precise choice of this metric is postponed to the verification

of (AG-6).) The fact that  $R$  is weak Feller follows from the continuity of the density  $\gamma$  and  $\Pi_p(\cdot)$ . Moreover, we have, for all  $z \in \mathbf{Z}$ ,

$$\begin{aligned} RV(z) &= \int V(\boldsymbol{\omega} + \mathbf{A}z + u\mathbf{b}) G(\Pi_p(z); du) \\ &= \mathbf{1}^T \boldsymbol{\omega} + \mathbf{1}^T (\mathbf{A} + \mathbf{b}\boldsymbol{\delta}_p^T) z. \end{aligned}$$

By iteration, we obtain, for all  $k \in \mathbb{Z}_+^*$ ,

$$R^k V(z) = \mathbf{1}^T \boldsymbol{\omega} + \cdots + \mathbf{1}^T (\mathbf{A} + \mathbf{b}\boldsymbol{\delta}_p^T)^{k-1} \boldsymbol{\omega} + \mathbf{1}^T (\mathbf{A} + \mathbf{b}\boldsymbol{\delta}_p^T)^k z.$$

Since  $|\lambda|_{\max}(\mathbf{A} + \mathbf{b}\boldsymbol{\delta}_p^T) < 1$ , then there exist  $(q, \rho) \in \mathbb{Z}_+^* \times (0, 1)$  such that

$$R^q V(z) \leq \mathbf{1}^T \boldsymbol{\omega} + \cdots + \mathbf{1}^T (\mathbf{A} + \mathbf{b}\boldsymbol{\delta}_p^T)^{q-1} \boldsymbol{\omega} + \rho^q \mathbf{1}^T z.$$

Hence (AG-4) holds with  $\beta = \mathbf{1}^T \boldsymbol{\omega} + \cdots + \mathbf{1}^T (\mathbf{A} + \mathbf{b}\boldsymbol{\delta}_p^T)^{q-1} \boldsymbol{\omega} \geq 0$  and  $\lambda = \rho^q < 1$ . We now prove (AG-5). We have for all  $n \geq 1$ ,  $t_{0:n-1} \in \mathbf{U}^n$  and  $z \in \mathbf{Z}$ ,

$$\Psi^\theta \langle u_{0:n-1} \rangle (z) = \mathbf{A}^n z + \sum_{j=0}^{n-1} \mathbf{A}^j (\boldsymbol{\omega} + u_{n-1-j} \mathbf{b}). \quad (5.47)$$

Let us use the norm

$$\|\mathbf{M}\| = \bigvee_{j=1}^{2p-1} \sum_{i=1}^{2p-1} |\mathbf{M}_{i,j}| = \sup_{|z| \leq 1} |\mathbf{M}z|$$

on  $(2p-1) \times (2p-1)$  matrices, where  $|z| = \sum_{i=1}^{2p-1} |z_i|$  is a norm of  $z = z_{1:2p-1} \in \mathbf{Z}$ . Note that since  $|\lambda|_{\max}(\mathbf{A} + \mathbf{b}\boldsymbol{\delta}_p^T) < 1$ , there exists  $\delta \in (0, 1)$  and  $c > 0$  such that, for any  $k \geq 1$ ,

$$\left\| (\mathbf{A} + \mathbf{b}\boldsymbol{\delta}_p^T)^k \right\| \leq c \delta^k. \quad (5.48)$$

Using that  $\mathbf{A}$ ,  $\mathbf{b}$ ,  $\boldsymbol{\delta}_p$  all have nonnegative entries, we have

$$\|\mathbf{A}^k\| \leq \left\| (\mathbf{A} + \mathbf{b}\boldsymbol{\delta}_p^T)^k \right\|. \quad (5.49)$$

Hence  $(\mathbf{I}_{2p-1} - \mathbf{A})^{-1} = \mathbf{I}_{2p-1} + \sum_{k \geq 1} \mathbf{A}^k$  is well defined and we set  $z_\infty = (\mathbf{I}_{2p-1} - \mathbf{A})^{-1} \boldsymbol{\omega}$  so that, with (5.65), we have

$$\Psi^\theta \langle u_{0:n-1} \rangle (z) - z_\infty = \mathbf{A}^n z + \sum_{j \geq n} \mathbf{A}^j \boldsymbol{\omega} + \sum_{j=0}^{n-1} u_{n-1-j} \mathbf{A}^j \mathbf{b}.$$

Then, using definition (5.38), we get that,  $\bar{\mathbb{P}}_z$ -a.s., for all  $n \geq 1$ ,

$$\begin{aligned} |Z_n - z_\infty| &= |\Psi \langle U_{0:n-1} \rangle(z) - z_\infty| \\ &\leq |\mathbf{A}^n(z - z_\infty)| + \sum_{j \geq n} |\mathbf{A}^j \boldsymbol{\omega}| + \left( \max_{0 \leq j \leq n-1} U_j \right) \sum_{j=0}^{n-1} |\mathbf{A}^j \mathbf{b}|. \end{aligned}$$

With (5.48) and (5.49), this implies

$$\bar{\mathbb{P}}_z \left( |Z_n - z_\infty| \leq c \left[ \delta^n \left( |z - z_\infty| + \frac{|\boldsymbol{\omega}|}{1 - \delta} \right) + \frac{|\mathbf{b}|}{1 - \delta} \max_{0 \leq j \leq n-1} U_j \right] \right) = 1.$$

To obtain (AG-5), it is sufficient to observe that, since  $g$  takes positive values in (5.43), for any positive  $\epsilon$ ,  $z \in \mathbf{Z}$  and any  $n \geq 1$ ,

$$\bar{\mathbb{P}}_z \left( \max_{0 \leq j \leq n-1} U_j < \epsilon \right) > 0.$$

Next we prove (AG-6). We have

$$\Psi_u(z) - \Psi_u(z') = \mathbf{A}(z - z').$$

Since (5.48) and (5.49) imply that  $|\lambda|_{\max}(\mathbf{A}) < 1$ , there exists a vector norm which makes  $\mathbf{A}$  strictly contracting. Choosing the metric  $d_Z$  on  $\mathbf{Z}$  (embedded in  $\mathbb{R}^{2p-1}$ ) as the one derived from this norm, we get (AG-6). To show (AG-7), we rely on Lemma 5.3.13. Let us set  $\mathbf{C} = (0, \infty) = \mathbf{X}$  and  $\mathbf{S} = \{1\}$ , then the probability measure  $\mu$  boils down the Dirac mass on  $\mathbf{S}$ . For all  $(x, u) \in \mathbf{X} \times \mathbf{U}$ , we define:

- (a) if  $\gamma(u) + \gamma(-u)$  is nonincreasing in  $u \in \mathbb{R}_+$ ,  
 $j(x) = \sqrt{x^{-1}}$  and  $h(x; u) = \frac{1}{2\sqrt{u}} \left[ \gamma \left( \sqrt{x^{-1}u} \right) + \gamma \left( -\sqrt{x^{-1}u} \right) \right]$ ; and
- (b) if  $\gamma(u) + \gamma(-u)$  is nondecreasing in  $u \in \mathbb{R}_+$ ,  
 $j(x) = \sqrt{x}$  and  $h(x; u) = \frac{1}{2x\sqrt{u}} \left[ \gamma \left( \sqrt{x^{-1}u} \right) + \gamma \left( -\sqrt{x^{-1}u} \right) \right]$ .

Clearly, (5.39) holds. In Case (a),  $h$  satisfies (H'-1) and in Case (b),  $h$  satisfies (H'-2). Hence, by Lemma 5.3.13, we obtain  $\alpha$  and  $\phi$ , which are given, respectively, as follows. For all  $z, z' \in \mathbf{Z}$ , in all cases, define

$$\alpha(z, z') = \left( \frac{\Pi_p(z) \wedge \Pi_p(z')}{\Pi_p(z) \vee \Pi_p(z')} \right)^{\frac{1}{2}}$$

and  $\phi(z, z') = \Pi_p(z) \wedge \Pi_p(z')$  in Case (a) and  $\phi(z, z') = \Pi_p(z) \vee \Pi_p(z')$  in Case (b). These so-defined functions  $\alpha$  and  $\phi$  indeed satisfy (AG-7)(i). For

any  $z, z' \in \mathbf{Z}$ , we have

$$\begin{aligned} 1 - \alpha(z, z') &= 1 - \left( \frac{\Pi_p(z) \wedge \Pi_p(z')}{\Pi_p(z) \vee \Pi_p(z')} \right)^{\frac{1}{2}} \\ &\leq \Pi_p(z)^{-1} \wedge \Pi_p(z')^{-1} |\Pi_p(z) - \Pi_p(z')| \\ &\leq W(z, z') d_{\mathbf{Z}}(z, z'), \end{aligned}$$

where  $d_{\mathbf{Z}}$  is the metric previously defined and  $W$  is defined by  $W(z, z') = 1 \vee (c_{d_{\mathbf{Z}}}(\Pi_p(z)^{-1} \wedge \Pi_p(z')^{-1}))$  with  $c_{d_{\mathbf{Z}}} > 0$  is conveniently chosen (such a constant exists since  $d_{\mathbf{Z}}$  is the metric associated to a norm and  $\mathbf{Z}$  has finite dimension). Then (AG-7)(ii) and (AG-7)(iii) hold and, since for all  $u \in \mathbf{U}$  and  $z \in \mathbf{Z}$ ,  $\Pi_p(\Psi_u(z)) \geq \omega$ , then  $W(\Psi_u(z), \Psi_u(z'))$  is uniformly bounded over  $(z, z', u) \in \mathbf{Z} \times \mathbf{Z} \times \mathbf{U}$  and (AG-7)(iv) holds. This completes the proof.  $\square$

### 5.4.2 Log-linear Poisson GARCH( $p, p$ ) Model

**Example 5.4.6** (Definition of the log linear Poisson GARCH( $p, p$ )). Consider the log-linear Poisson GARCH of order  $(p, p)$ ,  $p \in \mathbb{Z}_+^*$ , which is defined as follows: for all  $k \in \mathbb{Z}_+$ ,

$$\begin{aligned} X_{k+1} &= \omega + \sum_{i=1}^p a_i X_{k-i+1} + \sum_{i=1}^p b_i \ln(1 + Y_{k-i+1}), \\ Y_{k+1} | \mathcal{F}_k &\sim \mathcal{P}(e^{X_{k+1}}), \end{aligned} \tag{5.50}$$

where

- the parameter space

$$\theta = (\omega, a_{1:p}, b_{1:p}) \in \Theta \subset \mathbb{R}^{2p+1},$$

- $\mathcal{F}_k = \sigma(X_{-p+1:k+1}, Y_{-p+1:k})$ ,
- $\mathbf{X} = \mathbb{R}$ ,  $\mathcal{X} = \mathcal{B}(\mathbf{X})$ ,  $\mathbf{Y} = \mathbb{Z}_+$ ,  $\mathcal{Y} = \mathcal{B}(\mathbf{Y})$  and the metrics on  $\mathbf{X}$  and  $\mathbf{Y}$  are the same usual metric on  $\mathbb{R}$ .

By setting for all  $k$ ,  $U_k = \ln(1 + Y_k)$  (hence  $\mathbf{U} = \ln(1 + \mathbb{Z}_+)$ ,  $\mathcal{U} = \mathcal{B}(\mathbf{U})$  and  $d_{\mathbf{U}}$  is the usual metric on  $\mathbb{R}$ ), then Model (5.50) can be embedded in Model (5.51) below:

$$\begin{aligned} X_{k+1} &= \psi_{U_{k-p+1:k}}^{\theta}(X_{k-p+1:k}), \\ U_{k+1} | \mathcal{F}_k &\sim G^{\theta}(X_{k+1}; \cdot), \end{aligned} \tag{5.51}$$

where for all  $(x_{1:p}, u_{1:p}) \in \mathbf{X}^p \times \mathbf{U}^p$ ,

$$\psi_{u_{k-p+1:k}}^\theta(x_{k-p+1:k}) := \omega + \sum_{i=1}^p a_i x_{p-i+1} + \sum_{i=1}^p b_i u_{p-i+1} \quad (5.52)$$

and  $G^\theta(x; \cdot)$  is the conditional law of  $U_k$  given  $X_k = x$  defined by

$$G^\theta(x; A) = \sum_{u \in A} e^{-e^x} \frac{e^{x(e^u - 1)}}{(e^u - 1)!}, \quad A \in \mathcal{U}. \quad (5.53)$$

The density  $g^\theta$  of  $U_k$  conditionally on  $X_k = x$  with respect to the counting measure on  $\mathbf{U}$  is given by, for all  $u \in \mathbf{U}$ ,

$$g^\theta(x; u) = e^{-e^x} \frac{e^{x(e^u - 1)}}{(e^u - 1)!} > 0. \quad (5.54)$$

By **Definition 5.2.1**, Model (5.51) is an ODM( $p, p$ ) and is dominated by the counting measure on  $\mathbf{U}$ . Now let  $\mathbf{Z} = \mathbf{X}^p \times \mathbf{U}^{p-1}$  and  $\mathcal{Z} = \mathcal{X}^{\otimes p} \otimes \mathcal{U}^{\otimes(p-1)}$  and define the measurable function  $\Psi^\theta$  similarly as in (5.6). The function  $\Psi^\theta$  then take a simple linear form

$$\Psi_u^\theta(z) = \boldsymbol{\omega} + \mathbf{A}z + u\mathbf{b}, \quad (z, u) \in \mathbf{Z} \times \mathbf{U}. \quad (5.55)$$

As in **Example 5.4.1**, denoting  $Z_k = (X_{k-p+1:k}, U_{k-p+1:k-1})$  for all  $k$ , the process  $\{Z_k : k \in \mathbb{Z}_+\}$  is a Markov chain on  $(\mathbf{Z}, \mathcal{Z})$  whose transition kernel  $R^\theta$  is given by (5.11) with  $G^\theta$  and  $\Psi^\theta$  defined in (5.53) and (5.55), respectively.

In this section, we will use the vector and the matrix norms defined respectively on  $\mathbb{R}^{2p-1}$  and the space of  $(2p-1) \times (2p-1)$  matrices as follows:

$$\begin{aligned} |z|_\infty &= \bigvee_{i=1}^{2p-1} |\delta_i^T z| = \bigvee_{i=1}^{2p-1} |z_i|, \\ \|\mathbf{M}\|_\infty &= \bigvee_{i=1}^{2p-1} \sum_{j=1}^{2p-1} |\mathbf{M}_{i,j}| = \sup_{|z|_\infty \leq 1} |\mathbf{M}z|_\infty. \end{aligned}$$

For notational convenience, we will also denote  $\mathbf{A}_1 = \mathbf{A}$  and  $\mathbf{A}_2 = \mathbf{A} + \mathbf{b}\boldsymbol{\delta}_p^T$ .

**Proposition 5.4.7.** *Suppose that  $\theta \in \Theta$  satisfies*

$$\bigvee_{(i_1, \dots, i_p) \in \{1,2\}^p} \left\| \prod_{\ell=1}^p \mathbf{A}_{i_\ell} \right\|_\infty < 1. \quad (5.56)$$

*Then we have the following.*

(i) *Theorem 5.3.12* holds with the function  $\bar{V} : \mathbf{Z} \rightarrow \mathbb{R}_+$  defined by, for all  $z \in \mathbf{Z}$ ,

$$\bar{V}(z) = e^{\tau|z|_\infty}, \quad \tau > 0. \quad (5.57)$$

(ii) *Theorem 5.3.2* holds with any  $z_1 \in \mathbf{Z}$ . If the true parameter  $\theta_\star = (\omega^\star, a_{1:p}^\star, b_{1:p}^\star)$  moreover satisfies, for all  $k \in \mathbb{Z}_+$ ,

$$(\mathbf{A}^k - \mathbf{A}_\star^k)\mathbf{b}_\star = 0 \quad \text{implies} \quad \mathbf{A} = \mathbf{A}_\star, \quad (5.58)$$

then *Theorem 5.3.3* also holds.

(iii) If the true parameter  $\theta_\star$  lies in the interior of  $\Theta$  and satisfies (5.58), then *Theorem 5.3.7* holds.

**Remark 5.4.8.** When  $p = 1$ , then  $\mathbf{A}_1 = \mathbf{A}$  and  $\mathbf{A}_2 = \mathbf{A} + \mathbf{b}\delta_p^T$  correspond to  $a_1$  and  $a_1 + b_1$ , respectively. Condition (5.56) then reduces to

$$|a_1| \vee |a_1 + b_1| < 1.$$

This condition is weaker than the one derived in Douc et al. [2013] where  $|b_1| < 1$  is also imposed.

**Remark 5.4.9.** To check (5.58), it suffices to check that there exists  $k \in \mathbb{Z}_+$  such that the determinant

$$\det(\mathbf{A}_\star^k \mathbf{b}_\star, \mathbf{A}_\star^{k+1} \mathbf{b}_\star, \dots, \mathbf{A}_\star^{k+2p-2} \mathbf{b}_\star) \neq 0.$$

**Remark 5.4.10.** When  $p = 1$ , (5.58) holds if and only if  $b^\star \neq 0$ .

Before proving *Proposition 5.4.7*, let us show the following fact.

**Lemma 5.4.11.** *If  $Y \sim \mathcal{P}(\zeta)$  with  $\zeta > 0$ , then for any  $\vartheta \in \mathbb{R}$ , there exist constants  $c_1 = c_1(\vartheta), c_2 = c_2(\vartheta) > 0$  such that*

$$\mathbb{E}[(1 + Y)^\vartheta] \leq c_1 + c_2 \zeta^\vartheta \mathbb{1}\{\vartheta > 0\}. \quad (5.59)$$

*Proof.* The proof is postponed to *Section 5.6.4*. □

*Proof. Proof of Proposition 5.4.7–(i).* Now set  $V(z) = \bar{V}(z) = e^{\tau|z|_\infty}$ ,  $z \in \mathbf{Z}$ . From *Theorem 5.3.12*, we need to show that (AG-3), (AG-4), (AG-5), (AG-6) and (AG-7) hold. Assumption (AG-3) holds with the metric  $d_Z$  defined by, for all  $(z, z') \in \mathbf{Z}^2$ ,

$$d_Z(z, z') = \bigvee_{i=1}^{2p-1} |z_i - z'_i| = |z - z'|_\infty. \quad (5.60)$$



The fact that  $R$  is weak Feller follows from the continuity of the density  $g$  with respect to the  $x$  component in (5.54) and the continuity of  $\Pi_p(\cdot)$ . We will show that, with  $V$  defined in (5.57), then we have  $R^p V \leq \lambda V + \beta$  for some  $(\lambda, \beta) \in (0, 1) \times \mathbb{R}_+$ . Note that Lemma 5.4.11 implies that, for all  $\vartheta \in \mathbb{R}$  and  $\zeta > 0$ ,

$$\int e^{\vartheta u} G(\zeta; du) \leq c_1 + c_2 e^{\vartheta \zeta} \mathbb{1}\{\vartheta > 0\}, \quad (5.61)$$

where  $c_1$  and  $c_2$  are some positive constants. Note further that for all  $z \in \mathbb{Z}$  and  $\tau > 0$ , we have

$$V(z) = e^{\tau|z|^\infty} \leq \sum_{\sigma \in \{-1, 1\}} \sum_{k=1}^{2p-1} e^{\sigma \tau z_k} \leq 2(2p-1)e^{\tau|z|^\infty} \quad (5.62)$$

Without loss of generality, let us assume for the moment that  $\tau = 1$ . Now, using (5.61) and (5.62) and noting that, for all  $z \in \mathbb{Z}$ ,  $z_k = \delta_k^T z$ ,  $\Pi_p(z) = \delta_p^T z$  and  $\mathbb{1}\{\sigma \delta_k^T \mathbf{b} > 0\} \leq 1$ , we have

$$\begin{aligned} RV(z) &\leq \sum_{\sigma \in \{-1, 1\}} \sum_{k=1}^{2p-1} \int e^{\sigma \delta_k^T (\omega + \mathbf{A}z + u\mathbf{b})} G(\delta_p^T z; du) \\ &\leq \sum_{\sigma \in \{-1, 1\}} \sum_{k=1}^{2p-1} e^{\sigma \delta_k^T (\omega + \mathbf{A}z)} \int e^{\sigma \delta_k^T \mathbf{b}u} G(\delta_p^T z; du) \\ &\leq \sum_{\sigma \in \{-1, 1\}} \sum_{k=1}^{2p-1} e^{\sigma \delta_k^T (\omega + \mathbf{A}z)} \left( \gamma_1 + \gamma_2 e^{\sigma \delta_k^T \mathbf{b} \delta_p^T z} \right), \end{aligned}$$

where  $\gamma_1$  and  $\gamma_2$  are some positive constants. This further yields that, with  $\mathbf{A}_1 = \mathbf{A}$  and  $\mathbf{A}_2 = \mathbf{A} + \mathbf{b} \delta_p^T$ ,

$$\begin{aligned} RV(z) &\leq \sum_{\sigma \in \{-1, 1\}} \sum_{k=1}^{2p-1} \left( \beta_1 e^{\sigma \delta_k^T \mathbf{A}z} + \beta_2 e^{\sigma \delta_k^T (\mathbf{A} + \mathbf{b} \delta_p^T)z} \right) \\ &\leq \sum_{\sigma \in \{-1, 1\}} \sum_{k=1}^{2p-1} \left( \beta_1 e^{\sigma \delta_k^T \mathbf{A}_1 z} + \beta_2 e^{\sigma \delta_k^T \mathbf{A}_2 z} \right), \end{aligned}$$

where  $\beta_1$  and  $\beta_2$  are some positive constants. By iteration,  $R^{m+1}V = R(R^m V)$ , we obtain that, for all  $m \in \mathbb{Z}_+^*$ ,

$$\begin{aligned}
 R^m V(z) &\leq \sum_{\sigma \in \{-1,1\}} \sum_{k=1}^{2p-1} \sum_{(i_1, \dots, i_p) \in \{1,2\}^m} \beta_{i_1, \dots, i_m} e^{\sigma \delta_k^T (\prod_{\ell=1}^m \mathbf{A}_{i_\ell}) z} \\
 &\leq \sum_{(i_1, \dots, i_m) \in \{1,2\}^m} \beta_{i_1, \dots, i_m} \sum_{\sigma \in \{-1,1\}} \sum_{k=1}^{2p-1} e^{\sigma \delta_k^T (\prod_{\ell=1}^m \mathbf{A}_{i_\ell}) z} \\
 &\leq \sum_{(i_1, \dots, i_m) \in \{1,2\}^m} \beta_{i_1, \dots, i_m} \sum_{\sigma \in \{-1,1\}} \sum_{k=1}^{2p-1} e^{\|\prod_{\ell=1}^m \mathbf{A}_{i_\ell}\|_\infty |z|_\infty} \\
 &\leq 2(2p-1) \sum_{(i_1, \dots, i_m) \in \{1,2\}^m} \beta_{i_1, \dots, i_m} e^{\|\prod_{\ell=1}^m \mathbf{A}_{i_\ell}\|_\infty |z|_\infty} \\
 &\leq \left( 2(2p-1) \sum_{(i_1, \dots, i_m) \in \{1,2\}^m} \beta_{i_1, \dots, i_m} \right) e^{V_{(i_1, \dots, i_m) \in \{1,2\}^m} \|\prod_{\ell=1}^m \mathbf{A}_{i_\ell}\|_\infty |z|_\infty},
 \end{aligned} \tag{5.63}$$

where, for all  $(i_1, \dots, i_m) \in \{1, 2\}^m$ ,  $\beta_{i_1, \dots, i_m}$  are positive constants. Thus, for all  $z \in \mathbb{Z}$ , with  $m = p$ , we get

$$R^p V(z) \leq \gamma_p e^{\eta_p |z|_\infty},$$

where

$$\gamma_p := 2(2p-1) \sum_{(i_1, \dots, i_p) \in \{1,2\}^p} \beta_{i_1, \dots, i_p} > 0$$

and

$$\eta_p := \bigvee_{(i_1, \dots, i_p) \in \{1,2\}^p} \left\| \prod_{\ell=1}^p \mathbf{A}_{i_\ell} \right\|_\infty < 1.$$

Since

$$\lim_{|z|_\infty \rightarrow \infty} \frac{e^{\eta_p |z|_\infty}}{V(z)} = 0$$

and for all  $M > 0$ ,

$$\sup_{|z|_\infty \leq M} R^p V(z) < \infty,$$

then there exists a pair  $(\lambda, \beta) \in (0, 1) \times \mathbb{R}_+$  such that  $R^p V \leq \lambda V + \beta$ , and hence (AG-4) holds. Note that when  $m = p - 1$  and  $\tau = 1$ , we have  $e^{|\Pi_p(z)|} \leq \gamma_{p-1} R^{p-1} V(z)$  for some  $\gamma_{p-1} > 0$ . We now turn to prove (AG-5). Note that from (5.56), there exist  $\delta \in (0, 1)$  and  $c > 0$  such that, for any  $k \geq 1$ ,

$$\|\mathbf{A}^k\|_\infty \leq c \delta^k. \tag{5.64}$$

Hence,  $(\mathbf{I}_{2p-1} - \mathbf{A})^{-1} = \mathbf{I}_{2p-1} + \sum_{k \geq 1} \mathbf{A}^k$  is well defined. Now, set  $z_\infty = (\mathbf{I}_{2p-1} - \mathbf{A})^{-1} \boldsymbol{\omega}$ . Let  $\mathcal{N}_\infty$  be an open neighborhood of  $z_\infty$  and let  $z \in \mathbf{Z}$ . Define recursively the sequence  $\{z_k : k \in \mathbb{Z}_+\}$  with  $z_0 = z$  and for all  $k \geq 1$ ,  $z_k = \boldsymbol{\omega} + \mathbf{A}z_{k-1}$ . From (5.64), this so-defined sequence admits a unique limiting point  $\lim_{n \rightarrow \infty} z_n = z_\infty$ . Thus, there exists some  $n \in \mathbb{Z}_+^*$  such that for all  $k \geq n$ ,  $z_k \in \mathcal{N}_\infty$ . For such  $n$ , we have

$$\begin{aligned} R^n(z; \mathcal{N}_\infty) &= \bar{\mathbb{P}}_z(Z_n \in \mathcal{N}_\infty) \geq \bar{\mathbb{P}}_z(Z_k = z_k \text{ for all } k = 1, \dots, n) \\ &= \bar{\mathbb{P}}_z(U_0 = \dots = U_{n-1} = 0) > 0. \end{aligned}$$

Next we prove (AG-6). We have for all  $n \geq 1$ ,  $u_{0:n-1} \in \mathbf{U}^n$  and  $z \in \mathbf{Z}$ ,

$$\Psi \langle u_{0:n-1} \rangle (z) = \mathbf{A}^n z + \sum_{j=0}^{n-1} \mathbf{A}^j (\boldsymbol{\omega} + u_{n-1-j} \mathbf{b}). \quad (5.65)$$

Then for all  $(z, z') \in \mathbf{Z}^2$ ,  $u_{0:p-1} \in \mathbf{U}^p$ ,

$$\begin{aligned} d_{\mathbf{Z}}(\Psi_{u_0}(z), \Psi_{u_0}(z')) &= |\mathbf{A}(z - z')|_\infty \\ &\leq \|\mathbf{A}\|_\infty |z - z'|_\infty \\ &\leq D d_{\mathbf{Z}}(z, z'), \end{aligned}$$

and

$$\begin{aligned} d_{\mathbf{Z}}(\Psi \langle u_{0:p-1} \rangle (z), \Psi \langle u_{0:p-1} \rangle (z')) &= |\mathbf{A}^p(z - z')|_\infty \\ &\leq \|\mathbf{A}^p\|_\infty |z - z'|_\infty \\ &\leq \rho d_{\mathbf{Z}}(z, z'), \end{aligned}$$

where  $D = \|\mathbf{A}\|_\infty > 0$  and  $\rho = \|\mathbf{A}^p\|_\infty < 1$  from (5.56), showing (AG-6). To show (AG-7), we rely on Lemma 5.3.13. Let us set  $\mathbf{C} = \mathbb{R} = \mathbf{X}$  and  $\mathbf{S} = \{1\}$ , then the probability measure  $\mu$  boils down the Dirac mass on  $\mathbf{S}$ . For all  $(x, u) \in \mathbf{X} \times \mathbf{U}$ , let  $j(x) = e^{-e^x}$  and  $h(x; u) = \frac{e^{x(e^u-1)}}{(e^u-1)!}$ . Indeed,  $h$  satisfies (H'-1). Hence, from Lemma 5.3.13, the instrumental functions  $\alpha$  and  $\phi$  are obtained by setting: for all  $z, z' \in \mathbf{Z}$ ,

$$\alpha(z, z') = \frac{e^{-e^{\Pi_p(z) \vee \Pi_p(z')}}}{e^{-e^{\Pi_p(z) \wedge \Pi_p(z')}}} = e^{-|e^{\Pi_p(z)} - e^{\Pi_p(z')}|}$$

and  $\phi(z, z') = \Pi_p(z) \wedge \Pi_p(z')$ . These so-defined functions  $\alpha$  and  $\phi$  indeed satisfy (AG-7)(i). For  $(z, z') \in \mathbf{Z}^2$ , we have

$$\begin{aligned} 1 - \alpha(z, z') &= 1 - e^{-|e^{\Pi_p(z)} - e^{\Pi_p(z')}|} \leq |e^{\Pi_p(z)} - e^{\Pi_p(z')}| \\ &\leq e^{|\Pi_p(z) \vee \Pi_p(z')|} |\Pi_p(z) - \Pi_p(z')| \\ &\leq W(z, z') d_{\mathbf{Z}}(z, z'), \end{aligned}$$

where  $W$  is defined by, for all  $z, z' \in \mathbf{Z}$ ,

$$W(z, z') := \gamma_{p-1} (R^{p-1}V_1(z) + R^{p-1}V_1(z'))$$

and  $V_1(z) = e^{|z|^\infty}$ . Obviously, this so-defined function  $W$  satisfies (AG-7)(ii) and (AG-7)(iii). To complete the proof of Proposition 5.4.7–(i), it remains to show (AG-7)(iv). Note that (AG-4) holds with  $V(z) = e^{\tau|z|^\infty}$  for any  $\tau > 0$ , thus it also holds with  $V_1$ . For all  $z, z' \in \mathbf{Z}$ ,

$$\begin{aligned} J_1((z, z'); W) &= \int W(\Psi_u(z), \Psi_u(z'))G(\phi(z, z'); du) \\ &= \gamma_{p-1} \int R^{p-1}V_1(\Psi_u(z))G(\phi(z, z'); du) + \gamma_{p-1} \int V_1(\Psi_u(z'))G(\phi(z, z'); du). \end{aligned}$$

From (5.63), with  $m = p - 1$ , we have

$$\begin{aligned} &\int R^{p-1}V_1(\Psi_u(z))G(\phi(z, z'); du) \\ &\leq \sum_{\sigma \in \{-1, 1\}} \sum_{k=1}^{2p-1} \sum_{(i_1, \dots, i_{p-1}) \in \{1, 2\}^{p-1}} \beta_{i_1, \dots, i_{p-1}} \times \\ &\quad \int e^{\sigma \delta_k^T (\prod_{\ell=1}^{p-1} \mathbf{A}_{i_\ell}) (\omega + \mathbf{A}z + u\mathbf{b})} G(\phi(z, z'); du). \quad (5.66) \end{aligned}$$

Note that by (5.61), there exist constants  $c'_1$  and  $c'_2$  such that

$$\begin{aligned} &\int e^{\sigma \delta_k^T (\prod_{\ell=1}^{p-1} \mathbf{A}_{i_\ell}) \mathbf{b}u} G(\phi(z, z'); du) \\ &\leq c'_1 + c'_2 e^{\sigma \delta_k^T (\prod_{\ell=1}^{p-1} \mathbf{A}_{i_\ell}) \mathbf{b}(\delta_p^T z \wedge \delta_p^T z')} \mathbb{1} \left\{ \sigma \delta_k^T \left( \prod_{\ell=1}^{p-1} \mathbf{A}_{i_\ell} \right) \mathbf{b} > 0 \right\} \\ &\leq c'_1 + c'_2 e^{\sigma \delta_k^T (\prod_{\ell=1}^{p-1} \mathbf{A}_{i_\ell}) \mathbf{b} \delta_p^T z}, \end{aligned}$$

where we have used  $\delta_p^T z \wedge \delta_p^T z' \leq \delta_p^T z$  and  $\mathbb{1} \left\{ \sigma \delta_k^T \left( \prod_{\ell=1}^{p-1} \mathbf{A}_{i_\ell} \right) \mathbf{b} > 0 \right\} \leq 1$ . Plugging this inequality into (5.66) and noting that  $\mathbf{A}_1 = \mathbf{A}$  and  $\mathbf{A}_2 = \mathbf{A} + \mathbf{b} \delta_p^T$ , then there exist positive constants  $\gamma_{i_1, \dots, i_p}$ ,  $(i_1, \dots, i_p) \in \{1, 2\}^p$ , such that

$$\begin{aligned} &\int R^{p-1}V_1(\Psi_u(z))G(\phi(z, z'); du) \\ &\leq \sum_{\sigma \in \{-1, 1\}} \sum_{k=1}^{2p-1} \sum_{(i_1, \dots, i_p) \in \{1, 2\}^p} \gamma_{i_1, \dots, i_p} e^{\sigma \delta_k^T (\prod_{\ell=1}^p \mathbf{A}_{i_\ell}) z} \\ &\leq \left( 2(2p-1) \sum_{(i_1, \dots, i_p) \in \{1, 2\}^p} \gamma_{i_1, \dots, i_p} \right) e^{\eta_p |z|^\infty}, \end{aligned}$$

where  $\eta_p = \bigvee_{(i_1, \dots, i_p) \in \{1, 2\}^p} \|\prod_{\ell=1}^p \mathbf{A}_{i_\ell}\|_\infty < 1$ . Similarly, we have

$$\int R^{p-1} V_1(\Psi_u(z')) G(\phi(z, z'); du) \leq \left( 2(2p-1) \sum_{(i_1, \dots, i_p) \in \{1, 2\}^p} \gamma'_{i_1, \dots, i_p} \right) e^{\eta_p |z'|_\infty}$$

for some positive constants  $\gamma'_{i_1, \dots, i_p}$ ,  $(i_1, \dots, i_p) \in \{1, 2\}^p$ . Thus, there exists  $M > 0$  such that for all  $z, z' \in \mathbf{Z}$ ,

$$J_1((z, z'); W) = M \left( e^{\eta_p |z|_\infty} + e^{\eta_p |z'|_\infty} \right).$$

Assumption **(AG-7)(iv)** follows by observing that

$$\lim_{|z|_\infty \vee |z'|_\infty \rightarrow \infty} \frac{J_1((z, z'); W)}{W(z, z')} = 0.$$

This completes the proof of **Proposition 5.4.7–(i)**.

**Proof of Proposition 5.4.7–(ii)**. As a consequence, we have **(AG-1)**. Assumption **(AG-2)** directly follows. Thus, to show the first part of **Proposition 5.4.7–(ii)**, we only need to show that Assumptions **(BG-1)**, **(BG-2)** and **(BG-3)** are satisfied. The second part then follows if we can moreover show that **(a)** and **(b)** in **Theorem 5.3.3** hold.

Clearly, **(BG-1)** and **(BG-2)** hold by the definitions of  $\psi^\theta$  and  $g^\theta$  given by (5.52) and (5.54), respectively. It remains to show **(BG-3)**. Since  $\Theta$  is compact, then

$$|\omega| \leq \bar{\omega}, |a_i| \leq \bar{a}, |b_i| \leq \bar{b}, \quad \bigvee_{i_1, \dots, i_p \in \{1, 2\}^p} \left\| \prod_{\ell=1}^p \mathbf{A}_{i_\ell} \right\|_\infty \leq \bar{\rho}, \|\mathbf{A}\|_\infty \leq L, \quad (5.67)$$

for some  $(\bar{\omega}, \bar{a}, \bar{b}, \bar{\rho}) \in (0, \infty)^3 \times (0, 1)$  and  $L > 0$ . We set  $\mathbf{Z}_1 = \mathbf{Z}$ . And so **(BG-3)(i)** holds. Moreover, for all  $(\theta, z, u) \in \Theta \times \mathbf{Z}_1 \times \mathbf{U}$ ,  $g^\theta(\Pi_p(z); y) \leq 1$ . Thus, Condition **(BG-3)(ii)** holds. Now let  $z_1 \in \mathbf{Z}$ . Using (5.64), (5.65) and (5.67), we get that there exist constants  $\bar{c} > 0$  and  $\bar{\delta} \in (0, 1)$  such that for all  $z \in \mathbf{Z}$ ,  $u_{1:n} \in \mathbf{U}^n$  and  $\theta \in \Theta$ ,

$$\begin{aligned} |\Psi^\theta \langle u_{1:n} \rangle(z_1) - \Psi^\theta \langle u_{1:n} \rangle(z)|_\infty &= |\mathbf{A}^n(z_1 - z)|_\infty \\ &\leq \bar{c} \bar{\delta}^n |z_1 - z|_\infty. \end{aligned}$$

Thus by (5.60), we get **(BG-3)(iii)** with

$$\bar{\Psi}(z) = c |z_1 - z|_\infty.$$

Hence (BG-3)(iv) holds and since

$$|z_1 - \Psi_u^\theta(z_1)| \leq (L+1)|z_1|_\infty + \bar{\omega} + (1 \vee \bar{b})u,$$

we also get (BG-3)(v) provided that

$$\bar{\phi}(u) \geq (L+1)|z_1|_\infty + \bar{\omega} + (1 \vee \bar{b})u. \quad (5.68)$$

It is straightforward to show that for all  $\theta \in \Theta$ ,  $(x, x') \in \Pi_p(\mathbf{Z}_1) \times \Pi_p(\mathbf{Z}_1)$  and  $u \in \mathbf{U}$ ,

$$\begin{aligned} |\ln g^\theta(x; u) - \ln g^\theta(x'; u)| &\leq |x - x'| e^{|x| \vee |x'|} e^u \\ &\leq |x - x'| e^{|x - \Pi_p(z_1)| \vee |x' - \Pi_p(z_1)|} e^{u + |\Pi_p(z_1)|}. \end{aligned}$$

Thus for all  $\theta \in \Theta$ ,  $(z, z') \in \mathbf{Z}_1 \times \mathbf{Z}_1$  and  $u \in \mathbf{U}$ ,

$$|\ln g^\theta(x; u) - \ln g^\theta(x'; u)| \leq |z - z'|_\infty e^{|z - z_1|_\infty \vee |z' - z_1|_\infty} e^{u + |\Pi_p(z_1)|}.$$

We thus obtain (BG-3)(v), (BG-3)(vi) and (BG-3)(vii) by setting  $C = 1$ ,

$$H(s) = s, \quad s \in \mathbb{R}_+,$$

and

$$\bar{\phi}(u) = (L+1)|z_1|_\infty + \bar{\omega} + (1 \vee \bar{b})u + e^{u + |\Pi_p(z_1)|}.$$

In addition, for all  $\theta \in \Theta$ ,  $x \in \mathbf{X}$  and  $\tau > 0$ , we have

$$\int e^{\tau u} G^\theta(x; du) \leq 1 + 2^\tau e^{\tau x}.$$

Hence, by letting  $\bar{V} = V$  and using that for all  $\theta \in \Theta$ ,  $\pi_1^\theta V < \infty$ , with the above definitions, we obtain (BG-3)(viii). This establishes the equivalence-class consistency of the MLE for the log-linear Poisson GARCH( $p, p$ ).

To conclude the strong consistency of the MLE, it remains to check (a) and (b) in Theorem 5.3.3. Condition (a) clearly holds by the identifiability of the Poisson distribution. Note that as a consequence of the earlier proof, we have, for all  $\theta \in \Theta$ ,  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\Psi^\theta \langle U_{-\infty:0} \rangle = \sum_{k=0}^{\infty} \mathbf{A}^k(\boldsymbol{\omega} + U_{-k} \mathbf{b}).$$

By stationarity, for all  $\theta \in \Theta$  and  $t \in \mathbb{Z}$ , we have,  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\Psi^\theta \langle U_{-\infty:t} \rangle = \sum_{k=0}^{\infty} \mathbf{A}^k(\boldsymbol{\omega} + U_{-k+t} \mathbf{b}).$$

Moreover, if

$$\Psi^\theta \langle U_{-\infty:0} \rangle = \Psi^{\theta_\star} \langle U_{-\infty:0} \rangle, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.},$$

then by stationarity, for all  $t \in \mathbb{Z}$ ,

$$\Psi^\theta \langle U_{-\infty:t} \rangle = \Psi^{\theta_\star} \langle U_{-\infty:t} \rangle, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.},$$

and thus

$$\sum_{k=0}^{\infty} \mathbf{A}^k(\boldsymbol{\omega} + U_{-k+t} \mathbf{b}) = \sum_{k=0}^{\infty} \mathbf{A}_\star^k(\boldsymbol{\omega}_\star + U_{-k+t} \mathbf{b}_\star), \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.} \quad (5.69)$$

This implies

$$(\mathbf{b} - \mathbf{b}_\star) U_t = \sum_{k=1}^{\infty} \mathbf{A}_\star^k(\boldsymbol{\omega}_\star + U_{-k+t} \mathbf{b}_\star) - \sum_{k=1}^{\infty} \mathbf{A}^k(\boldsymbol{\omega} + U_{-k+t} \mathbf{b}), \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.} \quad (5.70)$$

Conditionally on  $\sigma(U_{-\infty:t-1})$ ,  $(e^{U_t} - 1)$  is a Poisson variable with positive intensity. Thus the left-hand side of (5.70) is  $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s. constant only if  $\mathbf{b} = \mathbf{b}_\star$ , implying  $b_1^\star = b_1$ . Then, (5.69) reduces to

$$\sum_{k=1}^{\infty} \mathbf{A}^k(\boldsymbol{\omega} + U_{-k+t} \mathbf{b}) = \sum_{k=1}^{\infty} \mathbf{A}_\star^k(\boldsymbol{\omega}_\star + U_{-k+t} \mathbf{b}_\star), \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.} \quad (5.71)$$

Thus by repeating the same argument as to obtain  $\mathbf{b} = \mathbf{b}_\star$ , (5.71) and so on yield that, for all integer  $k \geq 1$ ,

$$(\mathbf{A}^k - \mathbf{A}_\star^k) \mathbf{b}_\star = 0.$$

And by (5.58), we have  $\mathbf{A} = \mathbf{A}_\star$ , yielding that  $(a_{1:p}, b_{2:p}) = (a_{1:p}^\star, b_{2:p}^\star)$ . It is immediate from (5.69) that  $\boldsymbol{\omega} = \boldsymbol{\omega}_\star$ , and this is equivalent to having that  $\omega = \omega^\star$ . The strong consistency therefore follows.

**Proof of Proposition 5.4.7–(iii).** By Theorem 5.3.7, it suffices to show that (CG-1)–(CG-7) hold. From the definition of  $\Psi_u^\theta(z)$ , we have  $a(\theta, u) = \mathbf{A}$  and  $b(\theta, u) = \boldsymbol{\omega} + u \mathbf{b}$ , both admitting linear forms in  $\theta$ . Thus from compactness of  $\Theta$  and (5.56) and by noting that for all  $\tau > 0$ ,

$$\tilde{\mathbb{E}}^{\theta_\star} [e^{\tau U_1}] < \infty, \quad (5.72)$$

it follows that (CG-1) holds. Assumptions (CG-2) and (CG-3) clearly hold. Assumption (CG-4) is immediate as the conclusion of Proposition 5.4.7-(ii) above. We have

$$\begin{aligned}\ell_{z,k}^\theta &= \ln g^\theta(\Pi_p(\Psi^\theta\langle U_{1:k-1}\rangle(z)); U_k) \\ &= (e^{U_k} - 1)\delta_p^T \Psi^\theta\langle U_{1:k-1}\rangle(z) - e^{\delta_p^T \Psi^\theta\langle U_{1:k-1}\rangle(z)} - \ln(e^{U_k} - 1)!,\end{aligned}$$

$$\begin{aligned}\ell_k^\theta &= \ln g^\theta(\Pi_p(\Psi^\theta\langle U_{-\infty:k-1}\rangle(z)); U_k) \\ &= (e^{U_k} - 1)\delta_p^T \Psi^\theta\langle U_{-\infty:k-1}\rangle - e^{\delta_p^T \Psi^\theta\langle U_{-\infty:k-1}\rangle} - \ln(e^{U_k} - 1)!. \end{aligned}$$

Then from Lemma 5.3.4, for all  $i$ ,  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\frac{\partial \ell_{z,k}^\theta}{\partial \theta_i} = \left( e^{U_k} - 1 - e^{\delta_p^T \Psi^\theta\langle U_{1:k-1}\rangle(z)} \right) \delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta\langle U_{1:k-1}\rangle(z)$$

$$\frac{\partial \ell_k^\theta}{\partial \theta_i} = \left( e^{U_k} - 1 - e^{\delta_p^T \Psi^\theta\langle U_{-\infty:k-1}\rangle} \right) \delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta\langle U_{-\infty:k-1}\rangle$$

Thus for all  $i$ ,  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\begin{aligned}\left| \frac{\partial \ell_{z,k}^\theta}{\partial \theta_i} - \frac{\partial \ell_k^\theta}{\partial \theta_i} \right| &\leq \left| \delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta\langle U_{-\infty:k-1}\rangle - \delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta\langle U_{1:k-1}\rangle(z) \right| (e^{U_k} - 1) \\ &+ \left| e^{\delta_p^T \Psi^\theta\langle U_{-\infty:k-1}\rangle} - e^{\delta_p^T \Psi^\theta\langle U_{1:k-1}\rangle(z)} \right| \left| \delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta\langle U_{-\infty:k-1}\rangle \right| \\ &+ e^{|\delta_p^T \Psi^\theta\langle U_{1:k-1}\rangle(z)|} \left| \delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta\langle U_{-\infty:k-1}\rangle - \delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta\langle U_{1:k-1}\rangle(z) \right|\end{aligned}$$

Note that for all  $z \in \mathbf{Z}$  and  $k \geq 1$ , we have,  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\Psi^\theta\langle U_{-\infty:k-1}\rangle - \Psi^\theta\langle U_{1:k-1}\rangle(z) = \mathbf{A}^{k-1} (\Psi^\theta\langle U_{-\infty:0}\rangle - z),$$

and for all  $i$ ,

$$\begin{aligned}\frac{\partial}{\partial \theta_i} \Psi^\theta\langle U_{-\infty:k-1}\rangle - \frac{\partial}{\partial \theta_i} \Psi^\theta\langle U_{1:k-1}\rangle(z) \\ = \frac{\partial(\mathbf{A}^{k-1})}{\partial \theta_i} (\Psi^\theta\langle U_{-\infty:0}\rangle - z) + \mathbf{A}^{k-1} \frac{\partial}{\partial \theta_i} \Psi^\theta\langle U_{-\infty:0}\rangle.\end{aligned}$$



By similar arguments as in the proof of [Lemma 5.3.4](#) (see [Section 5.6.1](#)) and noting that  $\sup_{\theta \in \Theta} |\Psi^\theta \langle U_{-\infty:0} \rangle|_\infty$  and  $\sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \theta} \Psi^\theta \langle U_{-\infty:0} \rangle \right|_\infty$  are finite  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s., there exist a constant  $\rho_1 \in (0, 1)$  and a  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s. finite random variable  $\tilde{M}_1 = \tilde{M}_1(z)$  such that for all  $k \geq 1$ ,  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\left| \Psi^\theta \langle U_{-\infty:k-1} \rangle - \Psi^\theta \langle U_{1:k-1} \rangle(z) \right|_\infty \leq \tilde{M}_1 \rho_1^k,$$

and for all  $i$ ,  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\left| \frac{\partial}{\partial \theta_i} \Psi^\theta \langle U_{-\infty:k-1} \rangle - \frac{\partial}{\partial \theta_i} \Psi^\theta \langle U_{1:k-1} \rangle(z) \right|_\infty \leq \tilde{M}_1 \rho_1^k.$$

From [Lemma 5.3.4](#), we also have that, for all  $i$ ,

$$\tilde{\mathbb{E}}^{\theta^*} \left[ \sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \theta_i} \Psi^\theta \langle U_{-\infty:0} \rangle \right|_\infty \right] < \infty. \quad (5.73)$$

Recall that for all  $t, t' \in \mathbb{R}$ ,  $|e^t - e^{t'}| \leq e^{|\vee|t|} |t - t'|$ . Thus, we deduce that

$$\begin{aligned} & \left| \frac{\partial \ell_{z,k}^\theta}{\partial \theta_i} - \frac{\partial \ell_k^\theta}{\partial \theta_i} \right| \\ & \leq e^{\tilde{M}_1} \rho_1^k \tilde{M}_1 \left[ e^{U_k} - 1 + e^{|\Psi^\theta \langle U_{-\infty:k-1} \rangle|_\infty} \left( 1 + \left| \frac{\partial}{\partial \theta_i} \Psi^\theta \langle U_{-\infty:k-1} \rangle \right|_\infty \right) \right]. \end{aligned}$$

On the other hand,  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s., we have

$$e^{|\Psi^\theta \langle U_{-\infty:0} \rangle|_\infty} \leq e^{\sup_{\theta \in \Theta} |\Psi^\theta \langle U_{-\infty:0} \rangle|_\infty} \leq e^{\sum_{k=0}^\infty \sup_{\theta \in \Theta} |\mathbf{A}^k(\boldsymbol{\omega} + U_{-k} \mathbf{b})|_\infty}.$$

From [\(5.64\)](#) and [\(5.67\)](#), there exist constants  $c_1, c_2 > 0$  and  $\bar{\delta} \in (0, 1)$  such that,  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$e^{\sup_{\theta \in \Theta} |\Psi^\theta \langle U_{-\infty:0} \rangle|_\infty} \leq c_1 e^{c_2 \sum_{k=0}^\infty \bar{\delta}^k U_{-k}} \leq c_1 (1 - \bar{\delta}) \sum_{k=0}^\infty \bar{\delta}^k e^{c_2 (1 - \bar{\delta})^{-1} U_{-k}},$$

and from [\(5.72\)](#), it follows that for any positive  $t > 0$ ,

$$\tilde{\mathbb{E}}^{\theta^*} \left[ e^{t \sup_{\theta \in \Theta} |\Psi^\theta \langle U_{-\infty:0} \rangle|_\infty} \right] < \infty. \quad (5.74)$$

Thus, by stationarity and from [\(5.73\)](#) and [\(5.74\)](#), [Lemma 5.6.6](#) assures that there exist a constant  $\rho_2 \in (0, 1)$  and a  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s. finite random variable  $\tilde{M}_2 = \tilde{M}_2(z)$  such that for all  $i$  and for all  $k \geq 1$ ,  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\left| \frac{\partial \ell_{z,k}^\theta}{\partial \theta_i} - \frac{\partial \ell_k^\theta}{\partial \theta_i} \right| \leq \rho_2^k \tilde{M}_2.$$

Therefore, Assumption (CG-5) holds. We now check (CG-6). We have

$$\begin{aligned}\ell_1^\theta &= \ln g^\theta(\Pi_p(\Psi^\theta\langle U_{-\infty:0}\rangle); U_1) \\ &= (e^{U_1} - 1)\delta_p^T \Psi^\theta\langle U_{-\infty:0}\rangle - e^{\delta_p^T \Psi^\theta\langle U_{-\infty:0}\rangle} - \ln(e^{U_1} - 1)!\end{aligned}$$

Then for all  $i$  and  $j$ , we have

$$\frac{\partial \ell_1^\theta}{\partial \theta_i} = \left( e^{U_1} - 1 - e^{\delta_p^T \Psi^\theta\langle U_{-\infty:0}\rangle} \right) \delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta\langle U_{-\infty:0}\rangle$$

and from Lemma 5.3.4,  $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.,

$$\begin{aligned}\frac{\partial^2 \ell_1^\theta}{\partial \theta_i \partial \theta_j} &= \left( e^{U_1} - 1 - e^{\delta_p^T \Psi^\theta\langle U_{-\infty:0}\rangle} \right) \delta_p^T \frac{\partial^2}{\partial \theta_i \partial \theta_j} \Psi^\theta\langle U_{-\infty:0}\rangle \\ &\quad + e^{\delta_p^T \Psi^\theta\langle U_{-\infty:0}\rangle} \left( \delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta\langle U_{-\infty:0}\rangle \right) \left( \delta_p^T \frac{\partial}{\partial \theta_j} \Psi^\theta\langle U_{-\infty:0}\rangle \right).\end{aligned}$$

Using similar arguments as the proof of Lemma 5.3.4 together with (5.72), we can show that for all  $i$  and  $j$ , we have

$$\tilde{\mathbb{E}}^{\theta_\star} \left[ \sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \theta_i} \Psi^\theta\langle U_{-\infty:0}\rangle \right|_\infty^2 \right] < \infty$$

and

$$\tilde{\mathbb{E}}^{\theta_\star} \left[ \sup_{\theta \in \Theta} \left| \frac{\partial^2}{\partial \theta_i \partial \theta_j} \Psi^\theta\langle U_{-\infty:0}\rangle \right|_\infty \right] < \infty.$$

By (5.74) and (5.72), we can directly show that, for all  $i$  and  $j$ ,

$$\tilde{\mathbb{E}}^{\theta_\star} \left[ \sup_{\theta \in \Theta} \left| \frac{\partial \ell_1^\theta}{\partial \theta_i} \right|^2 \right] < \infty$$

and

$$\tilde{\mathbb{E}}^{\theta_\star} \left[ \sup_{\theta \in \Theta} \left| \frac{\partial^2 \ell_1^\theta}{\partial \theta_i \partial \theta_j} \right| \right] < \infty.$$

Thus, (CG-6)-(i) and (CG-6)-(ii) hold. Note that for all  $i, j$ , we have,  $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.,

$$\frac{\partial}{\partial \theta_i} g^\theta(\Pi_p(\Psi^\theta\langle U_{-\infty:0}\rangle); U_1) = g^\theta(\Pi_p(\Psi^\theta\langle U_{-\infty:0}\rangle); U_1) \frac{\partial \ell_1^\theta}{\partial \theta_i}$$

and

$$\begin{aligned} & \frac{\partial^2}{\partial\theta_i\partial\theta_j} g^\theta (\Pi_p (\Psi^\theta \langle Y_{-\infty:0} \rangle); U_1) \\ &= g^\theta (\Pi_p (\Psi^\theta \langle U_{-\infty:0} \rangle); U_1) \frac{\partial^2 \ell_1^\theta}{\partial\theta_i\partial\theta_j} + \frac{\partial}{\partial\theta_i} g^\theta (\Pi_p (\Psi^\theta \langle U_{-\infty:0} \rangle); Y_1) \frac{\partial \ell_1^\theta}{\partial\theta_j} \end{aligned}$$

Then, (CG-6)-(iii) follows by observing that for all  $i$ ,  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\left| \frac{\partial}{\partial\theta_i} g^\theta (\Pi_p (\Psi^\theta \langle U_{-\infty:0} \rangle); U_1) \right| = g^\theta (\Pi_p (\Psi^\theta \langle U_{-\infty:0} \rangle); U_1) \left| \frac{\partial \ell_1^\theta}{\partial\theta_i} \right| \leq \sup_{\theta \in \Theta} \left| \frac{\partial \ell_1^\theta}{\partial\theta_i} \right|,$$

and

$$\tilde{\mathbb{E}}^{\theta^*} \left[ \sup_{\theta \in \Theta} \left| \frac{\partial \ell_1^\theta}{\partial\theta_i} \right| \right] \leq \left( \tilde{\mathbb{E}}^{\theta^*} \left[ \sup_{\theta \in \Theta} \left| \frac{\partial \ell_1^\theta}{\partial\theta_i} \right|^2 \right] \right)^{1/2} < \infty.$$

Similarly, (CG-6)-(iv) holds. To complete the proof of asymptotic normality, it remains to check (CG-7). By Remark 5.3.8, it is sufficient to show that (5.34) holds. We establish this fact by following the argument provided by Francq et al. [2004]. Now let  $\boldsymbol{\alpha} = (\alpha_0, \dots, \alpha_{2p}) \in \mathbb{R}^{2p+1}$  and for convenience, write  $\theta = (\theta_0, \dots, \theta_{2p}) = (\omega, a_{1:p}, b_{1:p})$ . We have,  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\boldsymbol{\alpha}^T \frac{\partial \ell_1^{\theta^*}}{\partial \theta} = \sum_{i=0}^{2p} \alpha_i \left( e^{U_1} - 1 - e^{\boldsymbol{\delta}_p^T \Psi^{\theta^*} \langle U_{-\infty:0} \rangle} \right) \boldsymbol{\delta}_p^T \frac{\partial \Psi^{\theta^*} \langle U_{-\infty:0} \rangle}{\partial \theta_i}.$$

Since  $(e^{U_1} - 1)$  conditionally on  $\sigma(U_{-\infty:0})$  is a Poisson variable with positive intensity, then  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,  $\boldsymbol{\alpha}^T \frac{\partial \ell_1^{\theta^*}}{\partial \theta} = 0$  implies

$$\sum_{i=0}^{2p} \alpha_i \boldsymbol{\delta}_p^T \frac{\partial \Psi^{\theta^*} \langle U_{-\infty:0} \rangle}{\partial \theta_i} = 0, \quad \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.}$$

By stationarity, we have for all  $t \in \mathbb{Z}$ ,

$$\sum_{i=0}^{2p} \alpha_i \boldsymbol{\delta}_p^T \frac{\partial \Psi^{\theta^*} \langle U_{-\infty:t} \rangle}{\partial \theta_i} = 0, \quad \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.} \quad (5.75)$$

In view of (5.51) and by noting that  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,  $X_t^\theta = \boldsymbol{\delta}_p^T \Psi^\theta \langle U_{-\infty:t} \rangle$  for all  $\theta \in \Theta$  and  $t \in \mathbb{Z}$ , we have from (5.75) that for all  $t \in \mathbb{Z}$ ,

$$0 = \boldsymbol{\alpha}^T \frac{\partial X_t^{\theta^*}}{\partial \theta} = \boldsymbol{\alpha}^T \begin{pmatrix} 1 \\ X_{t-1}^{\theta^*} \\ \vdots \\ X_{t-p}^{\theta^*} \\ U_{t-1} \\ \vdots \\ U_{t-p} \end{pmatrix} + \sum_{i=1}^p a_i \boldsymbol{\alpha}^T \frac{\partial X_{t-i}^{\theta^*}}{\partial \theta} = \boldsymbol{\alpha}^T \begin{pmatrix} 1 \\ X_{t-1}^{\theta^*} \\ \vdots \\ X_{t-p}^{\theta^*} \\ U_{t-1} \\ \vdots \\ U_{t-p} \end{pmatrix}, \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.}$$

Since  $(e^{U_{t-1}} - 1)$  conditionally on  $\sigma(U_{-\infty:t-2})$  is a Poisson variable with positive intensity, we then have that  $\alpha_{p+1} = 0$ . By the same reason, it can be shown that  $\alpha_{p+2:p+2+i} = 0$  if  $\alpha_{2:2+i} = 0$ , for  $i \in \{0, \dots, p-2\}$ . Thus, if  $\boldsymbol{\alpha} \neq 0$  entails an ODM( $p-1, p-1$ ) representation, which is impossible since the model is identifiable. Hence,  $\boldsymbol{\alpha} = 0$  and this completes the proof.  $\square$

Our next example is the NBIN-GARCH( $p, p$ ),  $p \in \mathbb{Z}_+^*$ , which is defined as follows.

### 5.4.3 NBIN-GARCH( $p, p$ ) Model

**Example 5.4.12** (Definition of NBIN-GARCH( $p, p$ )). Consider a statistical model recursively defined by: for all  $k \in \mathbb{Z}_+$ ,

$$\begin{aligned} X_{k+1} &= \omega + \sum_{i=1}^p a_i X_{k-i+1} + \sum_{i=1}^p b_i Y_{k-i+1}, \\ Y_{k+1} | \mathcal{F}_k &\sim \mathcal{NB} \left( r, \frac{X_{k+1}}{1 + X_{k+1}} \right), \end{aligned} \quad (5.76)$$

where

– the parameter space

$$\Theta \subset \{\theta = (\omega, a_{1:p}, b_{1:p}, r) \in \mathbb{R}_+^* \times \mathbb{R}_+^p \times \mathbb{R}_+^p \times \mathbb{R}_+^*\},$$

–  $\mathcal{F}_k = \sigma(X_{-p+1:k+1}, Y_{-p+1:k})$ ,

–  $\mathbf{X} = (0, \infty)$ ,  $\mathcal{X} = \mathcal{B}(\mathbf{X})$ ,  $\mathbf{Y} = \mathbb{Z}_+$ ,  $\mathcal{Y} = \mathcal{B}(\mathbf{Y})$  and the metrics on  $\mathbf{X}$  and  $\mathbf{Y}$  are any metrics on  $\mathbb{R}$  and  $\mathbb{Z}$ , respectively.

Denoting for all  $(x_{1:p}, y_{1:p}) \in \mathbf{X}^p \times \mathbf{Y}^p$ ,

$$\psi_{y_{1:p}}^\theta(x_{1:p}) := \omega + \sum_{i=1}^p a_i x_{p-i+1} + \sum_{i=1}^p b_i y_{p-i+1}, \quad (5.77)$$

Model (5.76) can be expressed as the standard form of Model (5.78) below:

$$\begin{aligned} X_{k+1} &= \psi_{Y_{k-p+1:k}}^\theta(X_{k-p+1:k}), \\ Y_{k+1} | \mathcal{F}_k &\sim \mathcal{NB}\left(r, \frac{X_{k+1}}{1 + X_{k+1}}\right). \end{aligned} \quad (5.78)$$

The density  $g^\theta$  and the probability law  $G^\theta$  of  $Y_k$  conditional on  $X_k = x$  are given by, for all  $y \in \mathcal{Y}$  and  $A \in \mathcal{Y}$ ,

$$g^\theta(x; y) = \frac{\Gamma(r + y)}{y! \Gamma(r)} \left(\frac{1}{1 + x}\right)^r \left(\frac{x}{1 + x}\right)^y, \quad (5.79)$$

and

$$G^\theta(x; A) = \sum_{y \in A} \mathbb{1}_A(y) g^\theta(x; y). \quad (5.80)$$

According to [Definition 5.2.1](#), Model (5.76) is an ODM( $p, p$ ) and is dominated by the counting measure on  $\mathbb{Z}$ . Now, let  $\mathbf{Z} = \mathbf{X}^p \times \mathbf{Y}^{p-1}$  and  $\mathcal{Z} = \mathcal{X}^{\otimes p} \otimes \mathcal{Y}^{\otimes (p-1)}$  and define the measurable function  $\Psi^\theta$  similarly as in (5.6), which is given by, for all  $(z, y) \in \mathbf{Z} \times \mathcal{Y}$ ,

$$\Psi_y^\theta(z) = \boldsymbol{\omega} + \mathbf{A}z + y\mathbf{b}. \quad (5.81)$$

The process  $\{Z_k : k \in \mathbb{Z}_+\}$ , where  $Z_k = (X_{k-p+1:k}, Y_{k-p+1:k-1})$ , is a Markov chain on  $(\mathbf{Z}, \mathcal{Z})$  with transition kernel  $R^\theta$  given by (5.11) with  $\Psi^\theta$  and  $G^\theta$  defined in (5.80) and (5.81), respectively. Let us now state our ergodicity, consistency and asymptotic normality results for the NBIN-GARCH( $p, p$ ) model.

**Proposition 5.4.13.** *Suppose that  $\theta \in \Theta$  satisfies*

$$|\lambda|_{\max}(\mathbf{A} + r\mathbf{b}\boldsymbol{\delta}_p^T) < 1. \quad (5.82)$$

- (i) *Then the conclusion of [Theorem 5.3.12](#) holds with  $\bar{V}(z) = \mathbf{1}^T z$  for all  $z \in \mathbf{Z}$ .*
- (ii) *[Theorem 5.3.2](#) holds with any  $z_1 \in \mathbf{Z}$ . If the true parameter  $\theta_\star = (\boldsymbol{\omega}^\star, \mathbf{a}_{1:p}^\star, \mathbf{b}_{1:p}^\star)$  moreover satisfies: for all  $k \in \mathbb{Z}_+$ ,*

$$(\mathbf{A}^k - \mathbf{A}_\star^k)\mathbf{b}_\star = 0 \quad \text{implies} \quad \mathbf{A} = \mathbf{A}_\star, \quad (5.83)$$

*then [Theorem 5.3.3](#) also holds.*

**Remark 5.4.14.** From [Lemma 5.4.2](#), Condition (5.82) is equivalent to  $\sum_{i=1}^p (a_i + rb_i) < 1$ .

*Proof. Proof of Proposition 5.4.13–(i).* In this part, for convenience, let us drop the super script  $\theta$  from for example  $G^\theta$ ,  $R^\theta$ , etc. Now set  $V = \bar{V}$ . From Theorem 5.3.12, we need to show that (AG-3), (AG-4), (AG-5), (AG-6) and (AG-7) hold. Assumption (AG-3) holds with any metric  $d_Z$  associated to a norm on the finite dimensional space  $Z$  so that  $\mathcal{Z}$  is the associated Borel  $\sigma$ -field (The precise choice of this metric is postponed to the verification of (AG-6)). The fact that  $R$  is weak Feller follows by observing that for fixed  $r > 0$ , as  $p \rightarrow p'$ ,  $\mathcal{NB}(r, p)$  converges weakly to  $\mathcal{NB}(r, p')$ . Moreover, we have, for all  $z \in Z$ ,

$$\begin{aligned} RV(z) &= \int V(\boldsymbol{\omega} + \mathbf{A}z + y\mathbf{b}) G(\Pi_p(z); dy) \\ &= \mathbf{1}^T \boldsymbol{\omega} + \mathbf{1}^T (\mathbf{A} + r\mathbf{b}\boldsymbol{\delta}_p^T) z. \end{aligned}$$

By iteration, we obtain that for all  $k \in \mathbb{Z}_+^*$ ,

$$R^k V(z) = \mathbf{1}^T \boldsymbol{\omega} + \cdots + \mathbf{1}^T (\mathbf{A} + r\mathbf{b}\boldsymbol{\delta}_p^T)^{k-1} \boldsymbol{\omega} + \mathbf{1}^T (\mathbf{A} + r\mathbf{b}\boldsymbol{\delta}_p^T)^k z.$$

Since  $|\lambda|_{\max}(\mathbf{A} + r\mathbf{b}\boldsymbol{\delta}_p^T) < 1$ , then there exists a pair  $(q, \rho) \in \mathbb{Z}_+^* \times (0, 1)$  such that

$$R^q V(z) \leq \mathbf{1}^T \boldsymbol{\omega} + \cdots + \mathbf{1}^T (\mathbf{A} + r\mathbf{b}\boldsymbol{\delta}_p^T)^{q-1} \boldsymbol{\omega} + \rho^q \mathbf{1}^T z.$$

Hence (AG-4) holds with some  $q \geq 1$ ,  $\lambda = \rho^q < 1$  and  $\beta = \mathbf{1}^T \boldsymbol{\omega} + \cdots + \mathbf{1}^T (\mathbf{A} + r\mathbf{b}\boldsymbol{\delta}_p^T)^{q-1} \boldsymbol{\omega} \geq 0$ . We now show (AG-5). Let us use the norm

$$\|\mathbf{M}\| = \bigvee_{j=1}^{2p-1} \sum_{i=1}^{2p-1} |\mathbf{M}_{i,j}| = \sup_{|z| \leq 1} |\mathbf{M}z|$$

on  $(2p-1) \times (2p-1)$  matrices, where  $|z| = \sum_{i=1}^{2p-1} |z_i|$  is a norm of  $z = z_{1:2p-1} \in Z$ . Note that from (5.82), there exists  $\delta \in (0, 1)$  and  $c > 0$  such that, for any  $k \geq 1$ ,

$$\|\mathbf{A}^k\| \leq c \delta^k. \quad (5.84)$$

Hence  $(\mathbf{I}_{2p-1} - \mathbf{A})^{-1} = \mathbf{I}_{2p-1} + \sum_{k \geq 1} \mathbf{A}^k$  is well defined and we set  $z_\infty = (\mathbf{I}_{2p-1} - \mathbf{A})^{-1} \boldsymbol{\omega}$ . Let  $\mathcal{N}_\infty$  be an open neighborhood of  $z_\infty$  and let  $z \in Z$ . Define recursively the sequence  $z_0 = z$  and for all  $k \geq 1$ ,  $z_k = \boldsymbol{\omega} + \mathbf{A}z_{k-1}$ . From (5.84), this so-defined sequence admits a unique limiting point  $\lim_{n \rightarrow \infty} z_n = z_\infty$ . Thus, there exists some  $n \in \mathbb{Z}_+^*$  such that for all  $k \geq n$ ,  $z_k \in \mathcal{N}_\infty$ . For such  $n$ , we have

$$\begin{aligned} R^n(z, \mathcal{N}_\infty) &= \bar{\mathbb{P}}_z(Z_n \in \mathcal{N}_\infty) \geq \bar{\mathbb{P}}_z(Z_k = z_k \text{ for all } k = 1, \dots, n) \\ &= \bar{\mathbb{P}}_z(Y_0 = \cdots = Y_{n-1} = 0) > 0. \end{aligned}$$

Next we prove (AG-6). We have

$$\Psi_y(z) - \Psi_y(z') = \mathbf{A}(z - z') .$$

Since (5.82) implies that  $|\lambda|_{\max}(\mathbf{A}) < 1$ , there exists a vector norm  $|\cdot|_{\mathbf{Z}}$ , which makes  $\mathbf{A}$  strictly contracting. Choosing the metric  $d_{\mathbf{Z}}$  on  $\mathbf{Z}$  (embedded in  $\mathbb{R}^{2p-1}$ ) as the one derived from this norm, we get (AG-6). To show (AG-7), we rely on Lemma 5.3.13. Let us set  $\mathbf{C} = (0, \infty) = \mathbf{X}$  and  $\mathbf{S} = \{1\}$ , then the probability measure  $\mu$  boils down to the Dirac mass on  $\mathbf{S}$ . For all  $(x, y) \in \mathbf{X} \times \mathbf{Y}$ , let  $j(x) = (1+x)^{-r}$  and  $h(x; y) = \frac{\Gamma(r+y)}{y! \Gamma(r)} \left(\frac{x}{1+x}\right)^y$ . Indeed,  $h$  satisfies (H'-1). Hence, from Lemma 5.3.13 we obtain  $\alpha$  and  $\phi$  by setting: for all  $z, z' \in \mathbf{Z}$ ,

$$\alpha(z, z') = \left( \frac{1 + \Pi_p(z) \wedge \Pi_p(z')}{1 + \Pi_p(z) \vee \Pi_p(z')} \right)^r$$

and

$$\phi(z, z') = \Pi_p(z) \wedge \Pi_p(z') .$$

These so-defined functions  $\alpha$  and  $\phi$  indeed satisfy (AG-7)(i). For any given  $r > 0$ , for all  $(z, z') \in \mathbf{Z}^2$ , we have

$$\begin{aligned} 1 - \alpha(z, z') &= 1 - \left( \frac{1 + \Pi_p(z) \wedge \Pi_p(z')}{1 + \Pi_p(z) \vee \Pi_p(z')} \right)^r \\ &\leq (1 \vee r) |\Pi_p(z) - \Pi_p(z')| \\ &\leq W(z, z') d_{\mathbf{Z}}(z, z') , \end{aligned}$$

where  $d_{\mathbf{Z}}$  is the metric previously defined and  $W$  is defined by  $W(z, z') = c_{d_{\mathbf{Z}}}(1 \vee r)$  with  $c_{d_{\mathbf{Z}}} > 0$  is conveniently chosen (such a constant exists since  $d_{\mathbf{Z}}$  is the metric associated to a norm and  $\mathbf{Z}$  is a subset of finite dimensional space). Thus, (AG-7)(ii) and (AG-7)(iii) hold and, since for all  $y \in \mathbf{Y}$  and  $z \in \mathbf{Z}$ ,  $\Pi_p(\Psi_y(z)) \geq \omega$ ,  $W(\Psi_y(z), \Psi_y(z'))$  is uniformly bounded over  $(z, z', y) \in \mathbf{Z} \times \mathbf{Z} \times \mathbf{Y}$  and (AG-7)(iv) holds. This completes the proof.

**Proof of Proposition 5.4.13–(ii).** Note that (AG-2) immediately follows. Thus, to show the first part of Proposition 5.4.7–(ii), we only need to show that Assumptions (BG-1), (BG-2) and (BG-3) are satisfied. The second part then follows if we can moreover show that (a) and (b) in Theorem 5.3.3 hold.

Clearly, (BG-1) and (BG-2) hold by the definitions of  $\psi^\theta$  and  $g^\theta$  given by (5.77) and (5.79), respectively. It remains to show (BG-3). Since  $\Theta$  is compact, then

$$\underline{\omega} \leq \omega \leq \bar{\omega}, 0 \leq a_i \leq \bar{a}, 0 \leq b_i \leq \bar{b}, |\lambda|_{\max}(\mathbf{A} + r \mathbf{b} \delta_p^T) \leq \bar{\rho}, \underline{r} \leq r \leq \bar{r} \quad (5.85)$$

for some  $(\underline{\omega}, \bar{\omega}, \bar{a}, \bar{b}, \underline{r}, \bar{r}, \bar{\rho}) \in (0, \infty)^6 \times (0, 1)$ .

We set  $Z_1 = Z$  so that **(BG-3)(i)** holds. Moreover, for all  $(\theta, z, y) \in \Theta \times Z_1 \times Y$ ,  $g^\theta(\Pi_p(z); y) \leq 1$ . Thus, Condition **(BG-3)(ii)** holds. Note that we have, for all  $n \in \mathbb{Z}_+^*$  and  $z \in Z$ ,

$$\Psi^\theta \langle y_{0:n-1} \rangle (z) = \mathbf{A}^n z + \sum_{j=0}^{n-1} \mathbf{A}^j (\boldsymbol{\omega} + y_{n-1-j} \mathbf{b}). \quad (5.86)$$

Now let  $|\cdot|$  be the Taxicab norm on  $Z$  and let  $z_1 \in Z$ . Using **(5.84)**, **(5.85)** and **(5.86)**, we get that there exist constants  $\bar{c} > 0$  and  $\bar{\delta} \in (0, 1)$  such that for all  $z \in Z$ ,  $y_{1:n} \in Y^n$  and  $\theta \in \Theta$ ,

$$\begin{aligned} d_Z (\Psi^\theta \langle y_{1:n} \rangle (z_1), \Psi^\theta \langle y_{1:n} \rangle (z)) &= |\mathbf{A}^n (z_1 - z)|_Z \\ &\leq \bar{c} \bar{\delta}^n |z_1 - z|. \end{aligned}$$

Thus, we get **(BG-3)(iii)** with

$$\bar{\Psi}(z) = \bar{c} |z_1 - z|.$$

Hence **(BG-3)(iv)** holds and since

$$|z_1 - \Psi_y^\theta(z_1)| \leq (L + 1) |z_1| + \bar{\omega} + (1 + \bar{b})y,$$

for some  $L > 0$ , we also get **(BG-3)(v)** provided that for all  $y \in Y$ ,

$$\bar{\phi}(y) \geq (L + 1) |z_1| + \bar{\omega} + (1 + \bar{b})y. \quad (5.87)$$

For all  $\theta \in \Theta$ ,  $(x, x') \in \Pi_p(Z_1) \times \Pi_p(Z_1)$  and  $y \in Y$ ,

$$\begin{aligned} |\ln g^\theta(x; y) - \ln g^\theta(x'; y)| &= |(r + y) [\ln(1 + x') - \ln(1 + x)] + y [\ln x - \ln x']| \\ &\leq [(r + y)(1 + \underline{\omega})^{-1} + y \underline{\omega}^{-1}] |x - x'| \\ &\leq [\bar{r} + y(1 + \underline{\omega}^{-1})] |x - x'|. \end{aligned}$$

We thus obtain **(BG-3)(v)**, **(BG-3)(vi)** and **(BG-3)(vii)** by setting  $C = 0$ ,  $H(s) = s$  and  $\bar{\phi}(y) = \bar{\omega} \vee \bar{r} + (1 + L) |z_1| + (\bar{b} + 1 + \underline{\omega}^{-1}) y$ . Observe that for all  $z \in Z$ ,

$$\int \ln^+ y G^\theta(\Pi_p(z); dy) \leq \int y G^\theta(\Pi_p(z); dy) = r \Pi_p(z) \leq \bar{r} \bar{V}(z).$$

Hence, using that for all  $\theta \in \Theta$ ,  $\pi_1^\theta V < \infty$ , we obtain **(BG-3)(viii)**, establishing the equivalence-class consistency of the MLE for the NBIN-GARCH( $p, p$ ) model.



To conclude the strong consistency of the MLE, it remains to check (a) and (b) in [Theorem 5.3.3](#). Condition (a) clearly holds by the identifiability of the negative binomial distribution. Note that as a consequence of the earlier proof, we have, for all  $\theta \in \Theta$ ,  $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.,

$$\Psi^\theta \langle Y_{-\infty:0} \rangle = \sum_{k=0}^{\infty} \mathbf{A}^k(\boldsymbol{\omega} + Y_{-k} \mathbf{b}).$$

By stationarity, for all  $\theta \in \Theta$  and  $t \in \mathbb{Z}$ , we have,  $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s.,

$$\Psi^\theta \langle Y_{-\infty:t} \rangle = \sum_{k=0}^{\infty} \mathbf{A}^k(\boldsymbol{\omega} + Y_{-k+t} \mathbf{b}).$$

Moreover, if

$$\Psi^\theta \langle Y_{-\infty:0} \rangle = \Psi^{\theta_\star} \langle Y_{-\infty:0} \rangle, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.},$$

then by stationarity, for all  $t \in \mathbb{Z}$ ,

$$\Psi^\theta \langle Y_{-\infty:t} \rangle = \Psi^{\theta_\star} \langle Y_{-\infty:t} \rangle, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.},$$

and thus

$$\sum_{k=0}^{\infty} \mathbf{A}^k(\boldsymbol{\omega} + Y_{-k+t} \mathbf{b}) = \sum_{k=0}^{\infty} \mathbf{A}_\star^k(\boldsymbol{\omega}_\star + Y_{-k+t} \mathbf{b}_\star), \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.} \quad (5.88)$$

This implies

$$(\mathbf{b} - \mathbf{b}_\star) Y_t = \sum_{k=1}^{\infty} \mathbf{A}_\star^k(\boldsymbol{\omega}_\star + Y_{-k+t} \mathbf{b}_\star) - \sum_{k=1}^{\infty} \mathbf{A}^k(\boldsymbol{\omega} + Y_{-k+t} \mathbf{b}), \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.} \quad (5.89)$$

Conditionally on  $\sigma(Y_{-\infty:t-1})$ ,  $Y_t$  is negative binomial variable with parameter  $(r_\star, p_\star) \in (0, \infty) \times (0, 1)$ . Thus, the left-hand side of (5.89) is  $\tilde{\mathbb{P}}^{\theta_\star}$ -a.s. constant only if  $\mathbf{b} = \mathbf{b}_\star$ , implying  $b_1^\star = b_1$ . Then, (5.88) reduces to

$$\sum_{k=1}^{\infty} \mathbf{A}^k(\boldsymbol{\omega} + Y_{-k+t} \mathbf{b}) = \sum_{k=1}^{\infty} \mathbf{A}_\star^k(\boldsymbol{\omega}_\star + Y_{-k+t} \mathbf{b}_\star), \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-a.s.} \quad (5.90)$$

Thus, by repeating the same argument as to obtain  $\mathbf{b} = \mathbf{b}_\star$ , (5.90) and so on yield that, for all integer  $k \geq 1$ ,

$$(\mathbf{A}^k - \mathbf{A}_\star^k) \mathbf{b}_\star = 0.$$

And by (5.83), we have  $\mathbf{A} = \mathbf{A}_\star$ , yielding that  $(a_{1:p}, b_{2:p}) = (a_{1:p}^\star, b_{2:p}^\star)$ . It is immediate from (5.88) that  $\boldsymbol{\omega} = \boldsymbol{\omega}_\star$ , and this is equivalent to having that  $\omega = \omega^\star$ . Thus, (b) is satisfied and the strong consistency therefore follows.  $\square$

**Proposition 5.4.15.** *Suppose that  $\theta \in \Theta$  satisfies*

$$|\lambda|_{\max} (\mathbf{A}^T \mathbf{A} + r (\mathbf{A}^T \mathbf{b} \boldsymbol{\delta}_p^T + \boldsymbol{\delta}_p \mathbf{b}^T \mathbf{A}) + r(r+1) \boldsymbol{\delta}_p \mathbf{b}^T \mathbf{b} \boldsymbol{\delta}_p^T) < 1. \quad (5.91)$$

*In addition, assume that the true parameter  $\theta_*$  lies in the interior of  $\Theta$  and satisfies (5.83), then [Theorem 5.3.7](#) holds.*

**Remark 5.4.16.** Condition (5.91) implies Condition (5.82). Thus the stationarity and ergodicity for the model follow. If moreover (5.83) is satisfied, then the strong consistency, that is,  $\lim_{n \rightarrow \infty} \hat{\theta}_{z,n} = \theta_*$ ,  $\tilde{\mathbb{P}}^{\theta_*}$ -a.s., holds for any  $z \in \mathbf{Z}$ .

Before proving [Proposition 5.4.15](#), let us show the following result.

**Lemma 5.4.17.** *Assume (5.91). Then*

$$\tilde{\mathbb{E}}^{\theta_*}[Y_1^2] < \infty. \quad (5.92)$$

*Proof.* Now, assume that (5.91) holds. To obtain (5.92) it is sufficient to show that there exists a triplet  $(q, \lambda, \beta) \in \mathbb{Z}_+ \times (0, 1) \times \mathbb{R}_+$  such that  $R^q \tilde{V} < \lambda \tilde{V} + \beta$  with  $\tilde{V}(z) = z^T z$ . Note first that for  $U \sim \mathcal{NB}(r, x/(1+x))$ , we have  $\mathbb{E}[U] = rx$  and  $\mathbb{E}[U^2] = r(r+1)x^2 + rx$ . Note also that the matrix  $\mathbf{M} := \mathbf{A}^T \mathbf{A} + r (\mathbf{A}^T \mathbf{b} \boldsymbol{\delta}_p^T + \boldsymbol{\delta}_p \mathbf{b}^T \mathbf{A}) + r(r+1) \boldsymbol{\delta}_p \mathbf{b}^T \mathbf{b} \boldsymbol{\delta}_p^T$  is symmetric positive semidefinite. Thus, there exist a diagonal matrix  $\mathbf{D} \geq 0$  and an orthogonal matrix  $\mathbf{Q}$  such that  $\mathbf{M} = \mathbf{QDQ}^T$ ; and, moreover,  $|\lambda|_{\max}(\mathbf{M}) = |\lambda|_{\max}(\mathbf{D})$ .

For all  $z \in \mathbf{Z}$ , we have

$$\begin{aligned} R\tilde{V}(z) &= \int \tilde{V}(\boldsymbol{\omega} + \mathbf{A}z + y\mathbf{b}) G(\Pi_p(z); dy) \\ &= \int (\boldsymbol{\omega} + \mathbf{A}z + y\mathbf{b})^T (\boldsymbol{\omega} + \mathbf{A}z + y\mathbf{b}) G(\boldsymbol{\delta}_p^T z; dy) \\ &= L_1(z) + z^T \mathbf{M} z = L_1(z) + z^T \mathbf{QDQ}^T z \\ &\leq L_1(z) + |\lambda|_{\max}(\mathbf{D}) z^T \mathbf{QQ}^T z = L_1(z) + |\lambda|_{\max}(\mathbf{M}) \tilde{V}(z), \end{aligned}$$

where  $L_1(z) = \boldsymbol{\omega}^T \boldsymbol{\omega} + (2\boldsymbol{\omega}^T \mathbf{A} + r2\boldsymbol{\omega}^T \mathbf{b} \boldsymbol{\delta}_p^T + r\mathbf{b}^T \mathbf{b} \boldsymbol{\delta}_p^T) z$ , which is a linear function in  $z$ . Since  $|\lambda|_{\max}(\mathbf{M}) < 1$ , and

$$\lim_{|z| \rightarrow \infty} \frac{R\tilde{V}(z)}{\tilde{V}(z)} \leq |\lambda|_{\max}(\mathbf{M}) < 1,$$

then there exists  $(\beta, \lambda) \in \mathbb{R}_+ \times (0, 1)$  such that for all  $z \in \mathbf{Z}$ ,

$$R\tilde{V}(z) \leq \lambda \tilde{V}(z) + \beta,$$

completing the proof.  $\square$

*Proof. Proof of Proposition 5.4.15.* Assumption (AG-1) holds by Remark 5.4.16. We now show that (CG-1)–(CG-7) are satisfied. From the definition of  $\Psi_y^\theta(z)$ , we have  $a(\theta, y) = \mathbf{A}$  and  $b(\theta, y) = \boldsymbol{\omega} + y\mathbf{b}$ , both admitting linear forms in  $\theta$ . From compactness of  $\Theta$  and by noting that (5.91) implies (CG-1)(ii) and that (5.92) also implies

$$\tilde{\mathbb{E}}^{\theta^*} [Y_1] < \infty, \quad (5.93)$$

it follows that (CG-1) holds. Assumptions (CG-2) and (CG-3) are clearly satisfied. Assumption (CG-4) immediately follows by Remark 5.4.16. We have

$$\begin{aligned} \ell_{z,k}^\theta &= \ln g^\theta (\Pi_p (\Psi^\theta \langle Y_{1:k-1} \rangle (z)); Y_k) \\ &= \ln \Gamma(Y_k + r) - (Y_k + r) \ln (1 + \boldsymbol{\delta}_p^T \Psi^\theta \langle Y_{1:k-1} \rangle (z)) \\ &\quad + Y_k \ln (\boldsymbol{\delta}_p^T \Psi^\theta \langle Y_{1:k-1} \rangle (z)) - \ln \Gamma(r) - \ln(Y_k!), \end{aligned}$$

$$\begin{aligned} \ell_k^\theta &= \ln g^\theta (\Pi_p (\Psi^\theta \langle Y_{-\infty:k-1} \rangle); Y_k) \\ &= \ln \Gamma(Y_k + r) - (Y_k + r) \ln (1 + \boldsymbol{\delta}_p^T \Psi^\theta \langle Y_{-\infty:k-1} \rangle) \\ &\quad + Y_k \ln (\boldsymbol{\delta}_p^T \Psi^\theta \langle Y_{-\infty:k-1} \rangle) - \ln \Gamma(r) - \ln(Y_k!). \end{aligned}$$

Then from Lemma 5.3.4, for all  $\theta = \theta_{0:2p+1}$  with  $\theta_{2p+1} = r$ , for all  $i \neq 2p+1$ ,  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\begin{aligned} \frac{\partial \ell_{z,k}^\theta}{\partial \theta_i} &= \left( \frac{Y_k}{\boldsymbol{\delta}_p^T \Psi^\theta \langle Y_{1:k-1} \rangle (z)} - \frac{Y_k + r}{1 + \boldsymbol{\delta}_p^T \Psi^\theta \langle Y_{1:k-1} \rangle (z)} \right) \boldsymbol{\delta}_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{1:k-1} \rangle (z) \\ \frac{\partial \ell_{z,k}^\theta}{\partial \theta_{2p+1}} &= \Gamma_2(Y_k + r) - \Gamma_2(r) - \ln (1 + \boldsymbol{\delta}_p^T \Psi^\theta \langle Y_{1:k-1} \rangle (z)) \end{aligned}$$

and

$$\begin{aligned} \frac{\partial \ell_k^\theta}{\partial \theta_i} &= \left( \frac{Y_k}{\boldsymbol{\delta}_p^T \Psi^\theta \langle Y_{-\infty:k-1} \rangle} - \frac{Y_k + r}{1 + \boldsymbol{\delta}_p^T \Psi^\theta \langle Y_{-\infty:k-1} \rangle} \right) \boldsymbol{\delta}_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{-\infty:k-1} \rangle \\ \frac{\partial \ell_k^\theta}{\partial \theta_{2p+1}} &= \Gamma_2(Y_k + r) - \Gamma_2(r) - \ln (1 + \boldsymbol{\delta}_p^T \Psi^\theta \langle Y_{-\infty:k-1} \rangle), \end{aligned}$$

where  $\Gamma_2$  is the digamma function defined by, for all  $u > 0$ ,

$$\Gamma_2(u) := \frac{d}{du} \ln \Gamma(u) = \frac{\int_0^\infty e^{-t} t^{u-1} \ln t \, dt}{\int_0^\infty e^{-t} t^{u-1} \, dt}.$$

For  $u \geq 3$ , the digamma function satisfies

$$\begin{aligned} |\Gamma_2(u)| &\leq \frac{\int_0^\infty e^{-t} t^{u-1} |\ln t| dt}{\Gamma(u)} \leq \frac{\int_0^\infty e^{-t} t^{u-1} (t + t^{-1}) dt}{\Gamma(u)} \\ &= \frac{\Gamma(u+1) + \Gamma(u-1)}{\Gamma(u)} = u + (u-1)^{-1} \leq u+1, \end{aligned} \quad (5.94)$$

where we have used that  $|\ln t| \leq (t + t^{-1})$  whenever  $t > 0$ .

It is straight forward to have that, for all  $i$ ,  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\begin{aligned} \left| \frac{\partial \ell_{z,k}^\theta}{\partial \theta_i} - \frac{\partial \ell_k^\theta}{\partial \theta_i} \right| &\leq \left| \delta_p^T \Psi^\theta \langle Y_{-\infty:k-1} \rangle - \delta_p^T \Psi^\theta \langle Y_{1:k-1} \rangle(z) \right| \\ &+ \left| \frac{Y_k + r}{1 + \delta_p^T \Psi^\theta \langle Y_{1:k-1} \rangle(z)} - \frac{Y_k}{\delta_p^T \Psi^\theta \langle Y_{1:k-1} \rangle(z)} \right| \\ &- \frac{Y_k + r}{1 + \delta_p^T \Psi^\theta \langle Y_{-\infty:k-1} \rangle} + \frac{Y_k}{\delta_p^T \Psi^\theta \langle Y_{-\infty:k-1} \rangle} \left| \delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{-\infty:k-1} \rangle \right| \\ &+ \left| \frac{Y_k + r}{1 + \delta_p^T \Psi^\theta \langle Y_{1:k-1} \rangle(z)} - \frac{Y_k}{\delta_p^T \Psi^\theta \langle Y_{1:k-1} \rangle(z)} \right| \\ &\times \left| \delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{-\infty:k-1} \rangle - \delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{1:k-1} \rangle(z) \right|. \end{aligned}$$

Since  $\delta_p^T \Psi^\theta \langle Y_{1:k-1} \rangle(z) \geq \underline{\omega}$  and so that  $\delta_p^T \Psi^\theta \langle Y_{-\infty:k-1} \rangle \geq \underline{\omega}$ , we obtain that for all  $i$ ,  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\begin{aligned} \left| \frac{\partial \ell_{z,k}^\theta}{\partial \theta_i} - \frac{\partial \ell_k^\theta}{\partial \theta_i} \right| &\leq \left| \delta_p^T \Psi^\theta \langle Y_{-\infty:k-1} \rangle - \delta_p^T \Psi^\theta \langle Y_{1:k-1} \rangle(z) \right| \\ &+ \left( \frac{2Y_k + r}{\underline{\omega}} \right) \left| \delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{-\infty:k-1} \rangle \right| \left| \delta_p^T \Psi^\theta \langle Y_{-\infty:k-1} \rangle - \delta_p^T \Psi^\theta \langle Y_{1:k-1} \rangle(z) \right| \\ &+ \left( \frac{2Y_k + r}{\underline{\omega}} \right) \left| \delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{-\infty:k-1} \rangle - \delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{1:k-1} \rangle(z) \right|. \end{aligned}$$

Note that for all  $z \in \mathbf{Z}$  and  $k \geq 1$ , we have,  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\Psi^\theta \langle Y_{-\infty:k-1} \rangle - \Psi^\theta \langle Y_{1:k-1} \rangle(z) = \mathbf{A}^{k-1} (\Psi^\theta \langle Y_{-\infty:0} \rangle - z),$$

$\frac{\partial}{\partial \theta_{2p+1}} \Psi^\theta \langle Y_{1:k-1} \rangle (z) = \frac{\partial}{\partial \theta_{2p+1}} \Psi^\theta \langle Y_{-\infty:k-1} \rangle = 0$  and, for all  $i \neq 2p+1$ ,

$$\begin{aligned} \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{-\infty:k-1} \rangle - \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{1:k-1} \rangle (z) \\ = \frac{\partial (\mathbf{A}^{k-1})}{\partial \theta_i} (\Psi^\theta \langle Y_{-\infty:0} \rangle - z) + \mathbf{A}^{k-1} \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{-\infty:0} \rangle. \end{aligned}$$

By proceeding similarly as in the proof of [Lemma 5.3.4](#) and noting that  $\sup_{\theta \in \Theta} |\boldsymbol{\delta}_p^T \Psi^\theta \langle Y_{-\infty:0} \rangle|$  and  $\sup_{\theta \in \Theta} |\boldsymbol{\delta}_p^T \frac{\partial}{\partial \theta} \Psi^\theta \langle Y_{-\infty:0} \rangle|$  are finite  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s., then there exist a constant  $\rho_1 \in (0, 1)$  and a  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s. finite random variable  $\tilde{M}_1 = \tilde{M}_1(z)$  such that for all  $k \geq 1$ ,  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$|\boldsymbol{\delta}_p^T \Psi^\theta \langle Y_{-\infty:k-1} \rangle - \boldsymbol{\delta}_p^T \Psi^\theta \langle Y_{1:k-1} \rangle (z)| \leq \tilde{M}_1 \rho_1^k,$$

and for all  $i$ ,  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\left| \boldsymbol{\delta}_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{-\infty:k-1} \rangle - \boldsymbol{\delta}_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{1:k-1} \rangle (z) \right| \leq \tilde{M}_1 \rho_1^k.$$

The last two inequalities imply that, for all  $\theta \in \Theta$ ,  $i \in \{1, \dots, 2p+2\}$  and  $k \in \mathbb{Z}_+^*$ ,  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\left| \frac{\partial \ell_{z,k}^\theta}{\partial \theta_i} - \frac{\partial \ell_k^\theta}{\partial \theta_i} \right| \leq \tilde{M}_1 \rho_1^k \left[ 1 + \left( \frac{2Y_k + \bar{r}}{\underline{\omega}} \right) \left( 1 + \sup_{\theta \in \Theta} \left| \boldsymbol{\delta}_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{-\infty:k-1} \rangle \right| \right) \right].$$

Note also that from [Lemma 5.6.6](#), we have, for all  $i$ ,

$$\tilde{\mathbb{E}}^{\theta^*} \left[ \sup_{\theta \in \Theta} \left| \boldsymbol{\delta}_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{-\infty:0} \rangle \right| \right] < \infty. \quad (5.95)$$

Thus, by stationarity and from [\(5.93\)](#) and [\(5.95\)](#), [Lemma 5.6.6](#) implies that there exist a constant  $\rho_2 \in (0, 1)$  and a  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s. finite random variable  $\tilde{M}_2 = \tilde{M}_2(z)$  such that for all  $\theta \in \Theta$ ,  $i \in \{0, \dots, 2p+1\}$  and  $k \geq 1$ ,  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\left| \frac{\partial \ell_{z,k}^\theta}{\partial \theta_i} - \frac{\partial \ell_k^\theta}{\partial \theta_i} \right| \leq \rho_2^k \tilde{M}_2.$$

Therefore, Assumption [\(CG-5\)](#) holds. We now check [\(CG-6\)](#). We have

$$\begin{aligned} \ell_1^\theta &= \ln g^\theta (\Pi_p (\Psi^\theta \langle Y_{-\infty:0} \rangle); Y_1) \\ &= \ln \Gamma(Y_1 + r) - (Y_1 + r) \ln (1 + \boldsymbol{\delta}_p^T \Psi^\theta \langle Y_{-\infty:0} \rangle) \\ &\quad - Y_1 \ln (\boldsymbol{\delta}_p^T \Psi^\theta \langle Y_{-\infty:0} \rangle) - \ln \Gamma(r) - \ln(Y_1!). \end{aligned}$$

Write  $\theta = \theta_{0:2p+1}$ . Then for all  $i$  and  $j$  such that  $i \neq 2p+1$  and  $j \neq 2p+1$ , we have

$$\frac{\partial \ell_1^\theta}{\partial \theta_i} = \left( \frac{Y_1}{\delta_p^T \Psi^\theta \langle Y_{-\infty:0} \rangle} - \frac{Y_1 + r}{1 + \delta_p^T \Psi^\theta \langle Y_{-\infty:0} \rangle} \right) \delta_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{-\infty:0} \rangle, \quad (5.96)$$

$$\frac{\partial \ell_1^\theta}{\partial \theta_{2p+2}} = \Gamma_2(Y_1 + r) - \Gamma_2(r) - \ln(1 + \delta_p^T \Psi^\theta \langle Y_{-\infty:0} \rangle), \quad (5.97)$$

and

$$\begin{aligned} \frac{\partial^2 \ell_1^\theta}{\partial \theta_i \partial \theta_j} &= \left( \frac{Y_1}{\delta_p^T \Psi^\theta \langle Y_{-\infty:0} \rangle} - \frac{Y_1 + r}{1 + \delta_p^T \Psi^\theta \langle Y_{-\infty:0} \rangle} \right) \delta_p^T \frac{\partial^2}{\partial \theta_i \partial \theta_j} \Psi^\theta \langle Y_{-\infty:0} \rangle \\ &+ \left( \frac{Y_1 + r}{(1 + \delta_p^T \Psi^\theta \langle Y_{-\infty:0} \rangle)^2} - \frac{Y_1}{(\delta_p^T \Psi^\theta \langle Y_{-\infty:0} \rangle)^2} \right) \delta_p^T \frac{\partial \Psi^\theta \langle Y_{-\infty:0} \rangle}{\partial \theta_i} \delta_p^T \frac{\partial \Psi^\theta \langle Y_{-\infty:0} \rangle}{\partial \theta_j}, \\ \frac{\partial^2 \ell_1^\theta}{\partial \theta_i \partial \theta_{2p+1}} &= - (1 + \delta_p^T \Psi^\theta \langle Y_{-\infty:0} \rangle)^{-1} \delta_p^T \frac{\partial \Psi^\theta \langle Y_{-\infty:0} \rangle}{\partial \theta_i}, \\ \frac{\partial^2 \ell_1^\theta}{\partial \theta_{2p+1} \partial \theta_{2p+1}} &= \Gamma_3(Y_1 + r) - \Gamma_3(r), \end{aligned}$$

where  $\Gamma_3$  is the trigamma function defined by, for all  $u > 0$ ,

$$\Gamma_3(u) := \frac{d}{du} \Gamma_2(u) = \frac{\int_0^\infty e^{-t} t^{u-1} |\ln t|^2 dt}{\Gamma(u)} - \Gamma_2(u)^2.$$

Note that when  $u \geq 3$ , the trigamma function satisfies

$$\begin{aligned} |\Gamma_3(u)| &\leq \frac{\int_0^\infty e^{-t} t^{u-1} (t + t^{-1})^2 dt}{\Gamma(u)} + \Gamma_2(u)^2 \\ &= \frac{\Gamma(u+2) + 2\Gamma(u) + \Gamma(u-2)}{\Gamma(u)} + \Gamma_2(u)^2 \leq 2(u+1)^2 + 3. \end{aligned} \quad (5.98)$$

Using similar lines as the proof of [Lemma 5.3.4](#) together with [\(5.92\)](#), it can be shown that, for all  $i$  and  $j$ ,

$$\tilde{\mathbb{E}}^{\theta_\star} \left[ \sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{-\infty:0} \rangle \right|^2 \right] < \infty \quad (5.99)$$

and

$$\tilde{\mathbb{E}}^{\theta_\star} \left[ \sup_{\theta \in \Theta} \left| \frac{\partial^2}{\partial \theta_i \partial \theta_j} \Psi^\theta \langle Y_{-\infty:0} \rangle \right| \right] < \infty. \quad (5.100)$$

From (5.96), (5.97) and (5.94), we have, for all  $i \neq 2p + 1$ ,  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\begin{aligned} \left| \frac{\partial \ell_1^\theta}{\partial \theta_i} \right| &\leq \left( \frac{2Y_k + \bar{r}}{\underline{\omega}} \right) \sup_{\theta \in \Theta} \left| \boldsymbol{\delta}_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{-\infty:0} \rangle \right|, \\ \left| \frac{\partial \ell_1^\theta}{\partial \theta_{2p+2}} \right| &\leq Y_1 + \tilde{C} + \ln \left( 1 + \sup_{\theta \in \Theta} \left| \boldsymbol{\delta}_p^T \frac{\partial}{\partial \theta_i} \Psi^\theta \langle Y_{-\infty:0} \rangle \right| \right), \end{aligned}$$

where  $\tilde{C}$  is a positive constant. Using tower property, (5.92) and (5.99), we obtain that for all  $i$ ,

$$\tilde{\mathbb{E}}^{\theta^*} \left[ \sup_{\theta \in \Theta} \left| \frac{\partial \ell_1^\theta}{\partial \theta_i} \right|^2 \right] < \infty.$$

By similar argument, we can obtain from (5.92), (5.98), (5.99) and (5.100) that, for all  $i$  and  $j$ ,

$$\tilde{\mathbb{E}}^{\theta^*} \left[ \sup_{\theta \in \Theta} \left| \frac{\partial^2 \ell_1^\theta}{\partial \theta_i \partial \theta_j} \right| \right] < \infty.$$

Thus, (CG-6)-(i) and (CG-6)-(ii) hold. Note that for all  $i, j$ , we have,  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\frac{\partial}{\partial \theta_i} g^\theta (\Pi_p (\Psi^\theta \langle Y_{-\infty:0} \rangle); Y_1) = g^\theta (\Pi_p (\Psi^\theta \langle Y_{-\infty:0} \rangle); Y_1) \frac{\partial \ell_1^\theta}{\partial \theta_i}$$

and

$$\begin{aligned} &\frac{\partial^2}{\partial \theta_i \partial \theta_j} g^\theta (\Pi_p (\Psi^\theta \langle Y_{-\infty:0} \rangle); Y_1) \\ &= g^\theta (\Pi_p (\Psi^\theta \langle Y_{-\infty:0} \rangle); Y_1) \frac{\partial^2 \ell_1^\theta}{\partial \theta_i \partial \theta_j} + \frac{\partial}{\partial \theta_i} g^\theta (\Pi_p (\Psi^\theta \langle Y_{-\infty:0} \rangle); Y_1) \frac{\partial \ell_1^\theta}{\partial \theta_j} \end{aligned}$$

Then, (CG-6)-(iii) follows by observing that for all  $i$ ,  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\left| \frac{\partial}{\partial \theta_i} g^\theta (\Pi_p (\Psi^\theta \langle Y_{-\infty:0} \rangle); Y_1) \right| = g^\theta (\Pi_p (\Psi^\theta \langle Y_{-\infty:0} \rangle); Y_1) \left| \frac{\partial \ell_1^\theta}{\partial \theta_i} \right| \leq \sup_{\theta \in \Theta} \left| \frac{\partial \ell_1^\theta}{\partial \theta_i} \right|,$$

and

$$\tilde{\mathbb{E}}^{\theta^*} \left[ \sup_{\theta \in \Theta} \left| \frac{\partial \ell_1^\theta}{\partial \theta_i} \right| \right] \leq \left( \tilde{\mathbb{E}}^{\theta^*} \left[ \sup_{\theta \in \Theta} \left| \frac{\partial \ell_1^\theta}{\partial \theta_i} \right|^2 \right] \right)^{1/2} < \infty.$$

Similarly, (CG-6)-(iv) holds. For concluding, it remains to check (CG-7). By Remark 5.3.8, it is sufficient to show that (5.34) holds. Now let  $\boldsymbol{\alpha} =$

$(\alpha_0, \dots, \alpha_{2p+1}) \in \mathbb{R}^{2p+2}$  and for convenience, write  $\theta = (\theta_0, \dots, \theta_{2p+1}) = (\omega, a_{1:p}, b_{2:p}, b_1, r)$ . We have  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\begin{aligned} \boldsymbol{\alpha}^T \frac{\partial \ell_1^{\theta^*}}{\partial \theta} &= \left( \frac{Y_1}{\boldsymbol{\delta}_p^T \Psi^{\theta^*} \langle Y_{-\infty:0} \rangle} - \frac{Y_1 + r^*}{1 + \boldsymbol{\delta}_p^T \Psi^{\theta^*} \langle Y_{-\infty:0} \rangle} \right) \sum_{i=0}^{2p} \alpha_i \boldsymbol{\delta}_p^T \frac{\partial \Psi^{\theta^*} \langle Y_{-\infty:0} \rangle}{\partial \theta_i} \\ &\quad + \alpha_{2p+1} [\Gamma_2(Y_1 + r^*) - \Gamma_2(r^*) - \ln(1 + \boldsymbol{\delta}_p^T \Psi^{\theta^*} \langle Y_{-\infty:0} \rangle)]. \end{aligned}$$

Since  $Y_1$  conditionally on  $\sigma(Y_{-\infty:0})$  is a negative binomial variable, then  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,  $\boldsymbol{\alpha}^T \frac{\partial \ell_1^{\theta^*}}{\partial \theta} = 0$  implies

$$\alpha_{2p+1} [\Gamma_2(r^*) + \ln(1 + \boldsymbol{\delta}_p^T \Psi^{\theta^*} \langle Y_{-\infty:0} \rangle)] = 0, \quad \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.}$$

Since  $X_1^{\theta^*} = \boldsymbol{\delta}_p^T Z_1^{\theta^*} = \boldsymbol{\delta}_p^T \Psi^{\theta^*} \langle Y_{-\infty:0} \rangle$  is a random variable, then  $\alpha_{2p+1} = 0$ . Thus,  $\boldsymbol{\alpha}^T \frac{\partial \ell_1^{\theta^*}}{\partial \theta} = 0$ ,  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s., also implies

$$\sum_{i=0}^{2p} \alpha_i \boldsymbol{\delta}_p^T \frac{\partial \Psi^{\theta^*} \langle Y_{-\infty:0} \rangle}{\partial \theta_i} = 0, \quad \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.}$$

The rest of the proof follows by the same argument as in the proof of [Proposition 5.4.7-\(iii\)](#) and is thus omitted here. The proof therefore follows.  $\square$

## 5.5 Empirical Study

### 5.5.1 Sharpness of Ergodicity Condition (5.56) for the Log-linear Poisson GARCH Model

We assess the sharpness of our general ergodicity condition (5.56) for the log-linear Poisson GARCH model by simulating several time series with different values of the condition level

$$\nu_L := \bigvee_{(k_1, \dots, k_p) \in \{1, 2\}^p} \left\| \prod_{j=1}^p \mathbf{A}_{k_j} \right\|$$

and in the cases where  $p = 1$  and  $2$ . Each generated time series holds 2000 observations. [Figure 5.1](#) presents three log-linear Poisson GARCH time series of order (1, 1) with different parameters: Series ODM(1, 1)-(a) corresponds to the model with parameter  $(\omega, a_1, b_1) = (-.03, -.7, 1.2)$  and  $\nu_L = .7$ ; Series ODM(1, 1)-(b) corresponds to the model with parameter



$(\omega, a_1, b_1) = (-.03, .57, .38)$  and  $\nu_L = .95$ ; and Series ODM(1, 1)-(c) corresponds to the model with parameter  $(\omega, a_1, b_1) = (-.03, .603, .402)$  and  $\nu_L = 1.005$ . Remark that in Series ODM(1, 1)-(a), we allow the parameter  $b_1$  to be greater than one. [Figure 5.2](#) exhibits three log-linear Poisson GARCH time series of order (2, 2) with different parameters. Series ODM(2, 2)-(a) corresponds to the model with parameter  $(\omega, a_1, a_2, b_1, b_2) = (-.02, .4, .1, .3, .1)$  and  $\nu_L = .9$ ; Series ODM(2, 2)-(b) corresponds to the model with parameter  $(\omega, a_1, a_2, b_1, b_2) = (-.02, .44, .11, .33, .11)$  and  $\nu_L = .99$ ; and Series ODM(2, 2)-(c) corresponds to the model with parameter  $(\omega, a_1, a_2, b_1, b_2) = (-.02, .446, .1115, .3345, .1115)$  and  $\nu_L = 1.0062$ . For both [Figure 5.1](#) and [Figure 5.2](#), it turns out that when the condition level  $\nu_L$  is far away from its boundary, the processes appear quite stable. However, the level of instability seems to increase as the condition level  $\nu_L$  gets closer to one. When this boundary is passed, the models appear to be unstable. We should remark that when  $\nu_L < 1$ , the stable behavior of these processes is persistent regardless of their starting points.

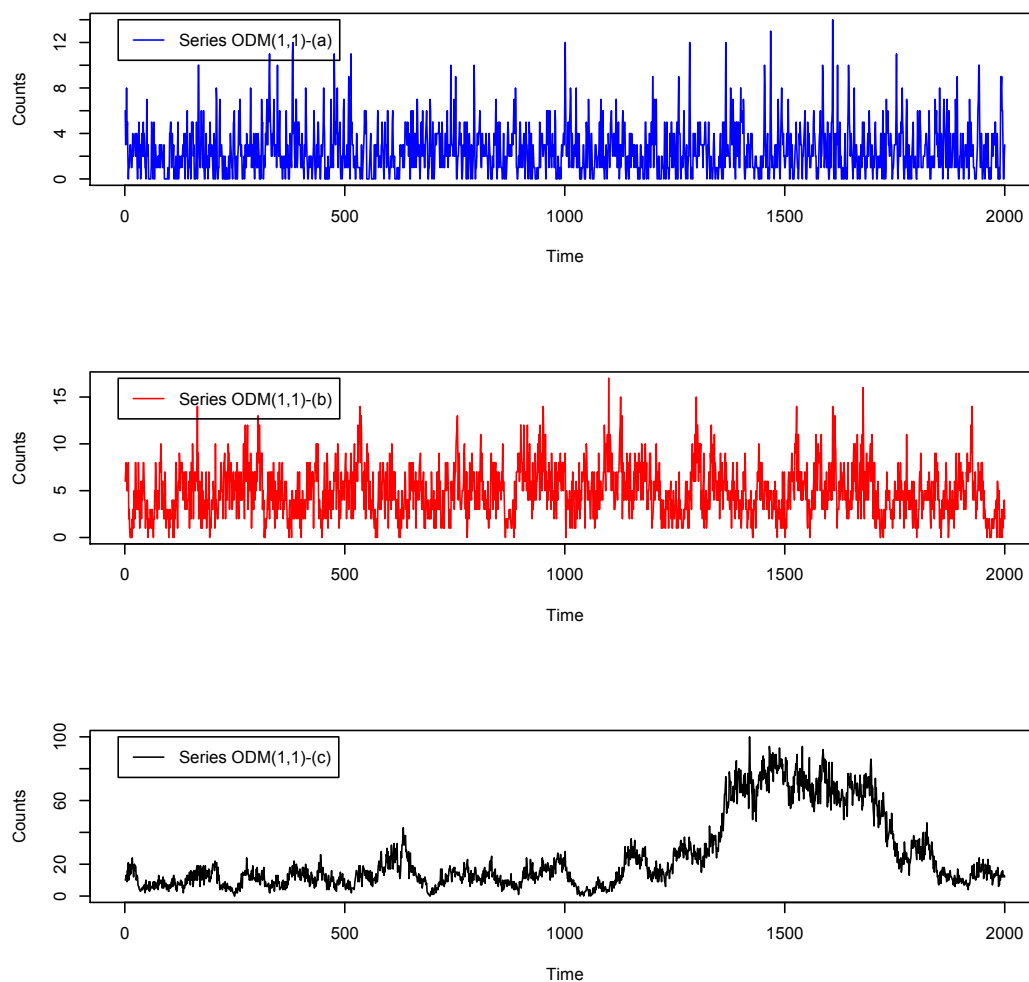


Figure 5.1 – Simulated time series for the log-linear Poisson GARCH of order (1, 1). Series ODM(1,1)-(a), -(b) and -(c) corresponds to the parameters  $(\omega, a_1, b_1)$  equal to  $(-.03, -.7, 1.2)$ ,  $(-.03, .57, .38)$  and  $(-.03, .603, .402)$ , respectively.

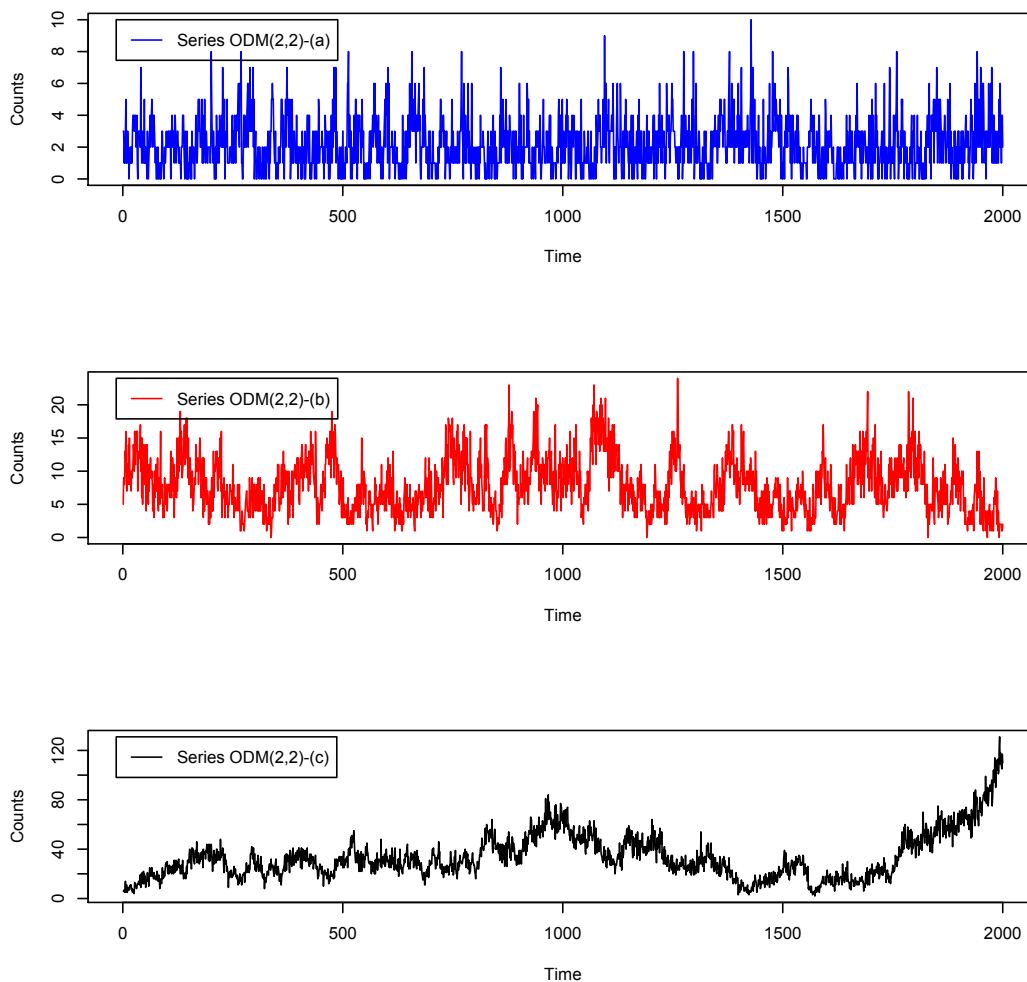


Figure 5.2 – Simulated time series for the log-linear Poisson GARCH of order (2,2). Series ODM(2,2)-(a), -(b) and -(c) corresponds to the parameters  $(\omega, a_1, a_2, b_1, b_2)$  equal to  $(-.02, .4, .1, .3, .1)$ ,  $(-.02, .44, .11, .33, .11)$  and  $(-.02, .446, .1115, .3345, .1115)$ , respectively.

### 5.5.2 Data Example

Our purpose here is to give some empirical evidence on some particular time series data for which higher-order ODMs may be more appropriate than the first-order ones. Our data set, provided by U.S. Geology Survey (USGS) (downloadable on: <http://earthquake.usgs.gov/earthquakes/search/>), contains a time series of numbers of earthquakes occurring hourly around the

globe with magnitudes larger than 4.0 Richters (a level that is detectable by humans), consisting of 8759 observations during the year 2014, from January 1 to December 31. The time series and its serial correlation plots are shown in [Figure 5.3](#) and [Figure 5.4](#), respectively, exhibiting that data possesses a long memory. In our analysis, we used the first 8559 observations for conducting the parameter estimation and the last 200 observations for evaluating the prediction performance of the fitted models. We fitted the data by several different models, by the NBIN-GARCH and the log-linear Poisson GARCH models with different orders: (1, 1), (1, 2), (2, 1), (2, 2), (3, 1), (3, 2) and (3, 3). For estimating the parameters, we use the constrained nonlinear optimization function `auglag` (Augmented Lagrangian Minimization Algorithm) from the package `alabama` (Augmented Lagrangian Adaptive Barrier Minimization Algorithm) in R. The values of AIC, BIC and average squared prediction error (ASPE),  $ASPE = m^{-1} \sum_{k=1}^m (Y_k - \hat{Y}_k)^2$ , where  $\hat{Y}_k$  is the predictor of the observation  $Y_k$ —hence the mean process evaluated at the MLE, for these models are reported in [Table 5.1](#). It shows that, using either the AIC or BIC criterion, the higher-order models are preferred for both the NBIN-GARCH and the log-linear Poisson GARCH, opposed to their first-order counterparts. Among the fitted models, for the NBIN-GARCH, the AIC suggests the model of order (3, 3) corresponding to the estimator  $(\hat{\omega}, \hat{a}_1, \hat{a}_2, \hat{a}_3, \hat{b}_1, \hat{b}_2, \hat{b}_3, \hat{r}) = (.035, .424, .068, .222, .021, 1.036 \times 10^{-8}, 5.681 \times 10^{-8}, 7.431)$ , while the BIC favors the model of order (3, 1) corresponding to the estimated parameter  $(\hat{\omega}, \hat{a}_1, \hat{a}_2, \hat{a}_3, \hat{b}_1, \hat{r}) = (.036, .414, .067, .232, .022, 7.131)$ . We observe that the values of  $\hat{b}_2$  and  $\hat{b}_3$  are so small that they might be insignificant to be included. However, to discard these coefficients from the model properly, we may need to conduct a hypothesis testing and this is beyond the scope of this thesis. Meanwhile, let us keep these terms in the model for convenience. For the log-linear Poisson GARCH models, the AIC suggests the model of (3, 3) corresponding to the estimator  $(\hat{\omega}, \hat{a}_1, \hat{a}_2, \hat{a}_3, \hat{b}_1, \hat{b}_2, \hat{b}_3) = (-.043, .495, .142, .091, .256, 1.084 \times 10^{-5}, -.017)$ , whereas the BIC picks the model with order (2, 1) corresponding to the estimator  $(\hat{\omega}, \hat{a}_1, \hat{a}_2, \hat{b}_1) = (-.041, .471, .234, .251)$ . The estimators for the NBIN-GARCH(1, 1) and for the log-linear Poisson GARCH(1, 1) are  $(\hat{\omega}, \hat{a}_1, \hat{b}_1, \hat{r}) = (.041, .719, .019, 6.997)$  and  $(\hat{\omega}, \hat{a}_1, \hat{b}_1) = (-.028, .722, .226)$ , respectively. [Table 5.1](#) also shows that the ASPE values of the models selected by either the AIC or the BIC criterion are lower than those of the first-order ones. This suggests that the higher-order models outperform in both cases of the NBIN-GARCH and the log-linear Poisson GARCH. However, there are barely any differences between these values for the models selected by the AIC and by the BIC, respectively. By comparing across the models, the selected log-linear Poisson GARCH model seems to yield prediction performance as well as the

selected NBIN-GARCH model in our data. [Figure 5.5](#) and [Figure 5.6](#) respectively depict the predicted series of first-order models and the ones chosen by AIC and BIC, compared with the real observations, for the NBIN-GARCH and the log-linear Poisson GARCH.

Table 5.1 – AIC, BIC and ASPE values for the NBIN-GARCH and the log-linear Poisson GARCH models with different orders: (1, 1), (1, 2), (2, 1), (2, 2), (3, 1), (3, 2) and (3, 3).

	NBIN-GARCH						
	(1, 1)	(1, 2)	(2, 1)	(2, 2)	(3, 1)	(3, 2)	(3, 3)
AIC	30102	30097	30083	30078	30074	30070	<b>30065</b>
BIC	30130	30132	30118	30120	<b>30117</b>	30119	30122
ASPE	1.633	1.633	1.618	1.633	1.614	1.614	1.613
	Log-linear Poisson GARCH						
	(1, 1)	(1, 2)	(2, 1)	(2, 2)	(3, 1)	(3, 2)	(3, 3)
AIC	30632	30621	30616	30614	30610	30608	<b>30607</b>
BIC	30653	30649	<b>30644.5</b>	30649	30645.4	30651	30656
ASPE	1.654	1.658	1.616	1.616	1.613	1.615	1.617

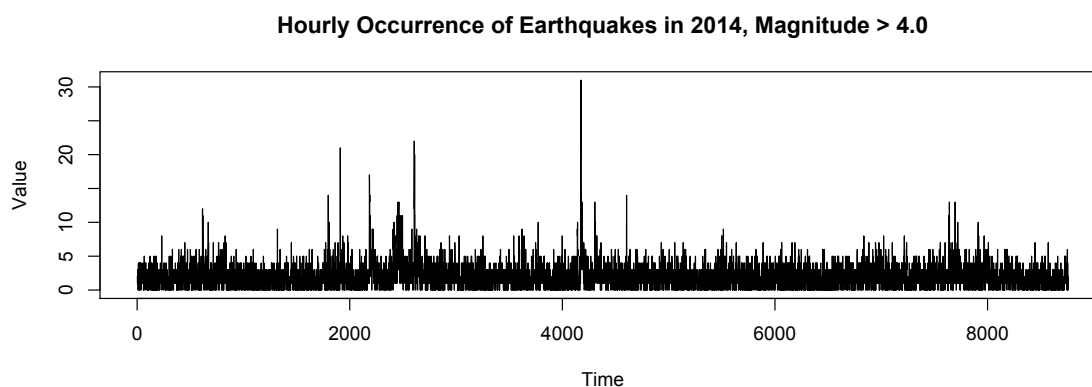


Figure 5.3 – Time series of numbers of earthquakes in 2014, from January 1 to December 31 (hourly).

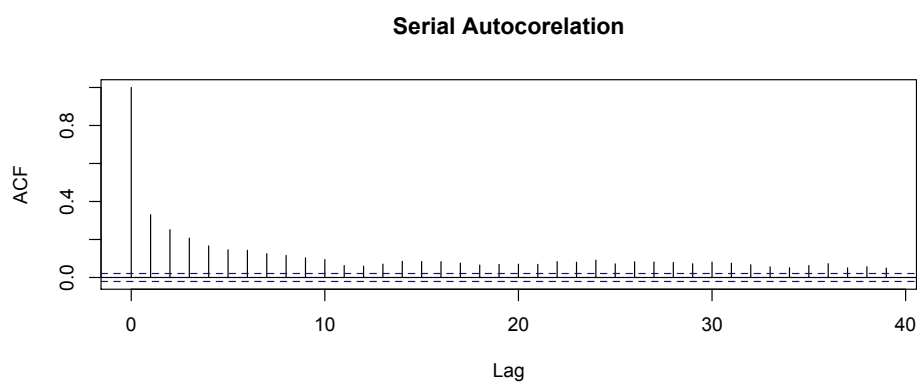


Figure 5.4 – Sample autocorrelation function.

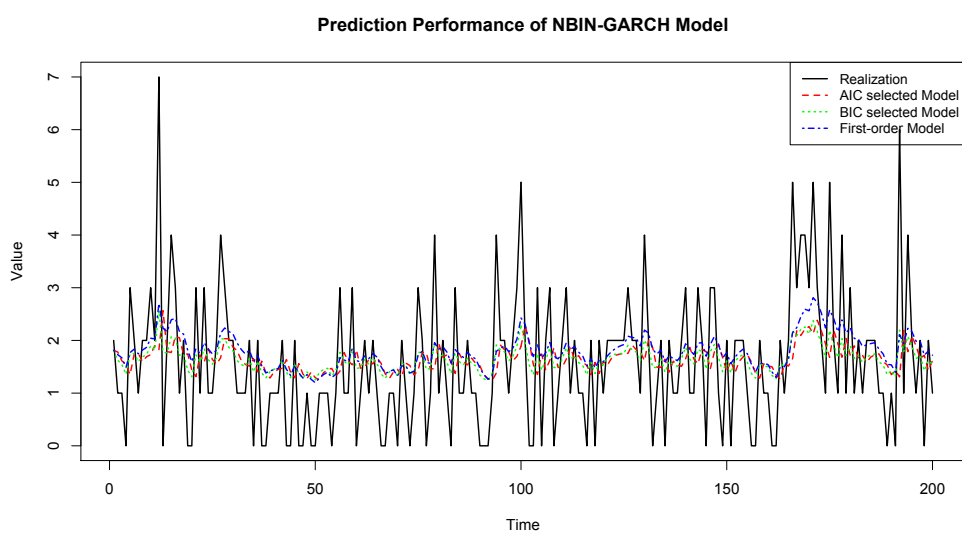


Figure 5.5 – Predicted series by the NBIN-GARCH of order (1,1) (dashed blue), by the one selected by the AIC (long dashed red) and by the one selected by the BIC (small dotted green), compared to the true series (black).

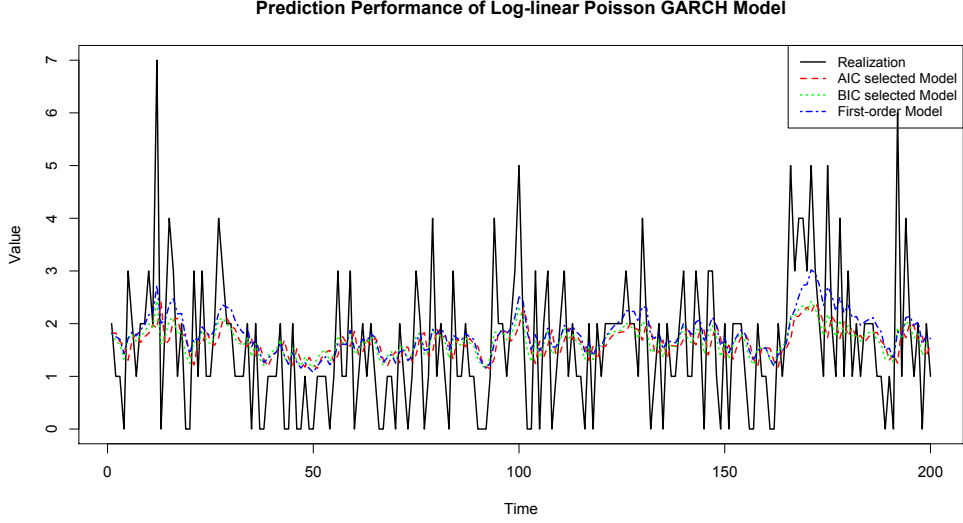


Figure 5.6 – Predicted series by the log-linear Poisson GARCH of order (1, 1) (dashed blue), by the one selected by the AIC (long dashed red) and by the one selected by the BIC (small dotted green), compared to the true series (black).

## 5.6 Postponed Proofs

### 5.6.1 Proof of Lemma 5.3.4

*Proof.* From (5.24), for all  $n \in \mathbb{Z}_+^*$ ,  $\theta \in \Theta$ ,  $z \in \mathbb{Z}$ , and  $y_{1:n} \in \mathbb{Y}^n$ , we have

$$\Psi^\theta \langle y_{1:n} \rangle (z) = \left( \prod_{\ell=1}^n a(\theta, y_{n-\ell+1}) \right) z + \sum_{k=1}^n \left( \prod_{\ell=1}^{k-1} a(\theta, y_{n-\ell}) \right) b(\theta, y_{n-k+1}). \quad (5.101)$$

Then, for all  $m \in \mathbb{Z}_+$ ,  $\theta \in \Theta$  and  $z \in \mathbb{Z}$ , we get that

$$\Psi^\theta \langle Y_{-m:0} \rangle (z) = \left( \prod_{\ell=0}^m a(\theta, Y_{-\ell}) \right) z + \sum_{k=0}^m \left( \prod_{\ell=0}^{k-1} a(\theta, Y_{-\ell}) \right) b(\theta, Y_{-k}).$$

By (CG-1)(ii), we have

$$\limsup_{m \rightarrow \infty} \sup_{\theta \in \Theta} \left\| \left( \prod_{\ell=0}^m a(\theta, Y_{-\ell}) \right) z \right\| = 0, \quad \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.}$$

Moreover, we have

$$\sum_{k=0}^{\infty} \left\| \left( \prod_{\ell=0}^{k-1} a(\theta, Y_{-\ell}) \right) b(\theta, Y_{-k}) \right\| \leq \bar{C} \sum_{k=0}^{\infty} \rho^k \sup_{\theta \in \Theta} \|b(\theta, Y_{-k})\|,$$

for some positive constant  $\bar{C}$ . By (CG-1)(iii) and Lemma 5.6.6, the right-hand side of the above inequality converges  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s. This implies that for all  $z \in \mathbb{Z}$ , the sequence  $\{\Psi^\theta \langle Y_{-m:0} \rangle(z) : m \in \mathbb{Z}_+\}$  converges uniformly on  $\Theta$  to a limit  $\Psi^\theta \langle Y_{-\infty:0} \rangle := \sum_{k=0}^{\infty} \left( \prod_{\ell=0}^{k-1} a(\theta, Y_{-\ell}) \right) b(\theta, Y_{-k})$ ,  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s., showing that (5.15) is satisfied. Next we show that (5.16) is satisfied. Note that for all  $m \in \mathbb{Z}_+$

$$\Psi^{\theta^*} \langle Y_{-m:0} \rangle(Z_{-m}) = \left( \prod_{\ell=0}^m a(\theta_*, Y_{-\ell}) \right) Z_{-m} + \sum_{k=0}^m \left( \prod_{\ell=0}^{k-1} a(\theta_*, Y_{-\ell}) \right) b(\theta_*, Y_{-k}).$$

Since  $\{Z_{-m} : m \in \mathbb{Z}_+\}$  is stationary under  $\mathbb{P}^{\theta^*}$  so that it is bounded in probability, and since  $\lim_{m \rightarrow \infty} \prod_{\ell=0}^m a(\theta_*, Y_{-\ell}) = 0$ ,  $\mathbb{P}^{\theta^*}$ -a.s., it then implies that  $\lim_{m \rightarrow \infty} \left( \prod_{\ell=0}^m a(\theta_*, Y_{-\ell}) \right) Z_{-m} = 0$  in  $\mathbb{P}^{\theta^*}$ -probability. From the proof above, we have

$$\lim_{m \rightarrow \infty} \sum_{k=0}^m \left( \prod_{\ell=0}^{k-1} a(\theta_*, Y_{-\ell}) \right) b(\theta_*, Y_{-k}) = \Psi^{\theta^*} \langle Y_{-\infty:0} \rangle, \quad \mathbb{P}^{\theta^*}\text{-a.s.}$$

Note also that for all  $m \in \mathbb{Z}_+$ ,  $Z_1 = \Psi^{\theta^*} \langle Y_{-m:0} \rangle(Z_{-m})$ ,  $\mathbb{P}^{\theta^*}$ -a.s. We therefore get that

$$\Psi^{\theta^*} \langle Y_{-\infty:0} \rangle = Z_1, \quad \mathbb{P}^{\theta^*}\text{-a.s.},$$

showing (5.16). In addition, (CG-1)(iii) implies

$$\begin{aligned} \tilde{\mathbb{E}}^{\theta^*} \left[ \sup_{\theta \in \Theta} \|\Psi^\theta \langle Y_{-\infty:0} \rangle\| \right] &\leq \bar{C} \sum_{k=0}^{\infty} \rho^k \tilde{\mathbb{E}}^{\theta^*} \left[ \sup_{\theta \in \Theta} \|b(\theta, Y_{-k})\| \right] \\ &\leq \frac{\bar{C}}{1-\rho} \tilde{\mathbb{E}}^{\theta^*} \left[ \sup_{\theta \in \Theta} \|b(\theta, Y_1)\| \right] < \infty, \end{aligned}$$

showing (5.27).

For any  $i$ , we have



$$\begin{aligned}
 & \frac{\partial \Psi^\theta \langle Y_{-m:0} \rangle(z)}{\partial \theta_i} \\
 &= \frac{\partial [(\prod_{\ell=0}^m a(\theta, Y_{-\ell})) z]}{\partial \theta_i} + \sum_{k=0}^m \frac{\partial \left[ \left( \prod_{\ell=0}^{k-1} a(\theta, Y_{-\ell}) \right) b(\theta, Y_{-k}) \right]}{\partial \theta_i} \\
 &= \sum_{\ell'=0}^m \left( \prod_{\ell=0}^m a_{\ell',i}(\theta, Y_{-\ell}) \right) z + \sum_{k=0}^m \left( \sum_{\ell'=0}^{k-1} \prod_{\ell=0}^{k-1} a_{\ell',i}(\theta, Y_{-\ell}) \right) b(\theta, Y_{-k}) \\
 &+ \sum_{k=0}^m \left( \prod_{\ell=0}^{k-1} a(\theta, Y_{-\ell}) \right) \frac{\partial b(\theta, Y_{-k})}{\partial \theta_i},
 \end{aligned}$$

where for all  $i, \ell, \ell'$ ,

$$a_{\ell',i}(\theta, Y_{-\ell}) = \begin{cases} a(\theta, Y_{-\ell}) & \text{if } \ell \neq \ell' \\ \frac{\partial}{\partial \theta_i} a(\theta, Y_{-\ell}) & \text{if } \ell = \ell'. \end{cases} \quad (5.102)$$

From (CG-1)(ii), we have

$$\begin{aligned}
 & \left\| \sum_{\ell'=0}^m \left( \prod_{\ell=0}^m a_{\ell',i}(\theta, Y_{-\ell}) \right) z \right\| \leq \sum_{\ell'=0}^m \left\| \prod_{\ell=0}^m a_{\ell',i}(\theta, Y_{-\ell}) \right\| \|z\| \\
 & \leq \sum_{\ell'=0}^m C^2 \rho^m \left\| \frac{\partial a(\theta, Y_{-\ell'})}{\partial \theta_i} \right\| \|z\| \\
 & \leq C^2 \|z\| \left( \frac{1+\rho}{2} \right)^m \sum_{\ell'=0}^m \left( \frac{2\rho}{1+\rho} \right)^{\ell'} \sup_{\theta \in \Theta} \left\| \frac{\partial a(\theta, Y_{-\ell'})}{\partial \theta_i} \right\|.
 \end{aligned}$$

We have from (CG-1)(iv) and by Lemma 5.6.6,

$$\sum_{\ell'=0}^{\infty} \left( \frac{2\rho}{1+\rho} \right)^{\ell'} \sup_{\theta \in \Theta} \left\| \frac{\partial a(\theta, Y_{-\ell'})}{\partial \theta_i} \right\| < \infty, \quad \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.}$$

Thus,  $\tilde{\mathbb{P}}^{\theta^*}\text{-a.s.}$ ,  $\left\| \sum_{\ell'=0}^m \left( \prod_{\ell=0}^m a_{\ell',i}(\theta, Y_{-\ell}) \right) z \right\|$  converges uniformly (in  $\theta$ ) to zero. Moreover, there exist a constant  $\rho_1 \in (0, 1)$  and a  $\tilde{\mathbb{P}}^{\theta^*}\text{-a.s.}$  finite random variable  $\tilde{M}_1 > 0$  with  $\tilde{\mathbb{E}}^{\theta^*}[\tilde{M}_1] < \infty$  such that for all  $m \in \mathbb{Z}_+$ ,

$$\sum_{\ell'=0}^m \sup_{\theta \in \Theta} \left\| \prod_{\ell=0}^m a_{\ell',i}(\theta, Y_{-\ell}) \right\| \leq \tilde{M}_1 \rho_1^m, \quad \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.} \quad (5.103)$$

Next, we show that  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\sum_{k=0}^m \left( \sum_{\ell'=0}^{k-1} \prod_{\ell=0}^{k-1} a_{\ell',i}(\theta, Y_{-\ell}) \right) b(\theta, Y_{-k})$$

converges uniformly on  $\Theta$ . We obtain from (5.103),  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\begin{aligned} \sum_{k=0}^m \left\| \left( \sum_{\ell'=0}^{k-1} \prod_{\ell=0}^{k-1} a_{\ell',i}(\theta, Y_{-\ell}) \right) b(\theta, Y_{-k}) \right\| &\leq \tilde{M}_1 \sum_{k=0}^m \rho_1^{k-1} \sup_{\theta \in \Theta} \|b(\theta, Y_{-k})\| \\ &\leq \tilde{M}_1 \sum_{k=0}^{\infty} \rho_1^{k-1} \sup_{\theta \in \Theta} \|b(\theta, Y_{-k})\|, \end{aligned}$$

which is finite from (CG-1)(iii) and Lemma 5.6.6. To conclude (5.25), it remains to show that  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\sum_{k=0}^m \left( \prod_{\ell=0}^{k-1} a(\theta, Y_{-\ell}) \right) \frac{\partial b(\theta, Y_{-k})}{\partial \theta_i}$$

converges uniformly on  $\Theta$ . We have  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s.,

$$\begin{aligned} \sum_{k=0}^m \left\| \left( \prod_{\ell=0}^{k-1} a(\theta, Y_{-\ell}) \right) \frac{\partial b(\theta, Y_{-k})}{\partial \theta_i} \right\| &\leq \sum_{k=0}^m C \rho^k \sup_{\theta \in \Theta} \left\| \frac{\partial b(\theta, Y_{-k})}{\partial \theta_i} \right\| \\ &\leq C \sum_{k=0}^{\infty} \rho^k \sup_{\theta \in \Theta} \left\| \frac{\partial b(\theta, Y_{-k})}{\partial \theta_i} \right\|, \end{aligned}$$

which is finite by (CG-1)(iv) and Lemma 5.6.6. Thus the sequence  $\left\{ \frac{\partial \Psi^\theta \langle Y_{-m:0} \rangle(z)}{\partial \theta} : m \in \mathbb{Z}_+ \right\}$   $\tilde{\mathbb{P}}^{\theta^*}$ -a.s. converges uniformly on  $\Theta$  and this follows that for all  $z \in \mathbb{Z}$  and  $\theta \in \Theta$ ,

$$\lim_{m \rightarrow \infty} \frac{\partial \Psi^\theta \langle Y_{-m:0} \rangle(z)}{\partial \theta} = \frac{\partial}{\partial \theta} \lim_{m \rightarrow \infty} \Psi^\theta \langle Y_{-m:0} \rangle(z) = \frac{\partial \Psi^\theta \langle Y_{-\infty:0} \rangle}{\partial \theta}, \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.},$$

showing that the quantity  $\frac{\partial \Psi^\theta \langle Y_{-\infty:0} \rangle}{\partial \theta}$  is  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s. well defined. The fact that (5.28) holds follows from (CG-1)(iv), (5.27) and (5.103).

For any  $i, j$ , we have

$$\begin{aligned}
 & \frac{\partial^2 \Psi^\theta \langle Y_{-m:0} \rangle (z)}{\partial \theta_j \partial \theta_i} \\
 &= \sum_{\ell'=0}^m \frac{\partial}{\partial \theta_j} \left( \prod_{\ell=0}^m a_{\ell',i}(\theta, Y_{-\ell}) \right) z \\
 &+ \sum_{k=0}^m \frac{\partial}{\partial \theta_j} \left[ \left( \sum_{\ell'=0}^{k-1} \prod_{\ell=0}^{k-1} a_{\ell',i}(\theta, Y_{-\ell}) \right) b(\theta, Y_{-k}) \right] \\
 &+ \sum_{k=0}^m \frac{\partial}{\partial \theta_j} \left[ \left( \prod_{\ell=0}^{k-1} a(\theta, Y_{-\ell}) \right) \frac{\partial b(\theta, Y_{-k})}{\partial \theta_i} \right] \\
 &= U_{1,m}(\theta) + U_{2,m}(\theta) + U_{3,m}(\theta) + U_{4,m}(\theta) + U_{5,m}(\theta),
 \end{aligned}$$

where

$$\begin{aligned}
 U_{1,m}(\theta) &= \sum_{\ell'=0}^m \sum_{\ell''=0}^m \left( \prod_{\ell=0}^m a_{\ell',\ell'',i,j}(\theta, Y_{-\ell}) \right) z, \\
 U_{2,m}(\theta) &= \sum_{k=0}^m \left( \sum_{\ell'=0}^{k-1} \sum_{\ell''=0}^{k-1} \prod_{\ell=0}^{k-1} a_{\ell',\ell'',i,j}(\theta, Y_{-\ell}) \right) b(\theta, Y_{-k}), \\
 U_{3,m}(\theta) &= \sum_{k=0}^m \left( \sum_{\ell'=0}^{k-1} \prod_{\ell=0}^{k-1} a_{\ell',i}(\theta, Y_{-\ell}) \right) \frac{\partial b(\theta, Y_{-k})}{\partial \theta_j}, \\
 U_{4,m}(\theta) &= \sum_{k=0}^m \left( \sum_{\ell'=0}^{k-1} \prod_{\ell=0}^{k-1} a_{\ell',j}(\theta, Y_{-\ell}) \right) \frac{\partial b(\theta, Y_{-k})}{\partial \theta_i}, \\
 U_{5,m}(\theta) &= \sum_{k=0}^m \left( \prod_{\ell=0}^{k-1} a(\theta, Y_{-\ell}) \right) \frac{\partial^2 b(\theta, Y_{-k})}{\partial \theta_j \partial \theta_i},
 \end{aligned}$$

and with for all  $i, j, \ell, \ell', \ell''$ ,

$$a_{\ell',\ell'',i,j}(\theta, Y_{-\ell}) = \begin{cases} a_{\ell',i}(\theta, Y_{-\ell}) & \text{if } \ell \neq \ell'' \\ \frac{\partial a_{\ell',i}(\theta, Y_{-\ell})}{\partial \theta_j} & \text{if } \ell = \ell'', \end{cases} \quad (5.104)$$

and from (5.102),

$$\frac{\partial a_{\ell',i}(\theta, Y_{-\ell})}{\partial \theta_j} = \begin{cases} \frac{\partial a(\theta, Y_{-\ell})}{\partial \theta_j} & \text{if } \ell \neq \ell' \\ \frac{\partial^2 a(\theta, Y_{-\ell})}{\partial \theta_j \partial \theta_i} & \text{if } \ell = \ell'. \end{cases} \quad (5.105)$$

To establish that  $\left\{ \frac{\partial^2 \Psi^\theta \langle Y_{-m;0} \rangle (z)}{\partial \theta_j \partial \theta_i} : m \in \mathbb{Z}_+ \right\}$  is almost surely uniformly convergent on  $\Theta$ , it requires proving that for all  $i \in \{1, \dots, 5\}$ , the sequences  $\{U_{i,m}(\theta) : m \in \mathbb{Z}_+\}$  converges uniformly on  $\Theta$ . We have

$$\begin{aligned} \|U_{1,m}(\theta)\| &= \left\| \sum_{\ell'=0}^m \sum_{\ell''=0}^m \left( \prod_{\ell=0}^m a_{\ell',\ell'',i,j}(\theta, Y_{-\ell}) \right) z \right\| \\ &\leq \sum_{\ell'=0}^m \sum_{\ell''=0}^m \left\| \prod_{\ell=0}^m a_{\ell',\ell'',i,j}(\theta, Y_{-\ell}) \right\| \|z\| \\ &\stackrel{(1)}{\leq} \tilde{M}_1^2 \|z\| \sum_{\ell'=0}^m \sum_{\ell''=0}^m \rho_1^m \sup_{\theta \in \Theta} \left\| \frac{\partial a_{\ell',i}(\theta, Y_{-\ell''})}{\partial \theta_j} \right\| \\ &\stackrel{(2)}{\leq} \tilde{M}_1^2 \|z\| \sum_{\ell''=0}^m \rho_1^m \left( m \sup_{\theta \in \Theta} \left\| \frac{\partial a(\theta, Y_{-\ell''})}{\partial \theta_j} \right\| + \sup_{\theta \in \Theta} \left\| \frac{\partial^2 a(\theta, Y_{-\ell''})}{\partial \theta_j \partial \theta_i} \right\| \right), \end{aligned}$$

where  $\stackrel{(1)}{\leq}$  follows from (5.103) and the submultiplicativity of the matrix norm  $\|\cdot\|$ , and  $\stackrel{(2)}{\leq}$  follows from (5.105). By assuming

$$\tilde{\mathbb{E}}^{\theta_*} \left[ \ln^+ \sup_{\theta \in \Theta} \left\| \frac{\partial a(\theta, Y_1)}{\partial \theta_i} \right\| \right] < \infty$$

and

$$\tilde{\mathbb{E}}^{\theta_*} \left[ \ln^+ \sup_{\theta \in \Theta} \left\| \frac{\partial^2 a(\theta, Y_1)}{\partial \theta_j \partial \theta_i} \right\| \right] < \infty,$$

then there exist a constant  $\rho_2 \in (0, 1)$  and a  $\tilde{\mathbb{P}}^{\theta_*}$ -a.s. finite random variable  $\tilde{M}_2 > 0$  with  $\tilde{\mathbb{E}}^{\theta_*}[\tilde{M}_2] < \infty$  such that for all  $m \in \mathbb{Z}_+$ ,

$$\leq \sum_{\ell'=0}^m \sum_{\ell''=0}^m \left\| \prod_{\ell=0}^m a_{\ell',\ell'',i,j}(\theta, Y_{-\ell}) \right\| \leq \tilde{M}_2 \rho_2^m, \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.}$$

Consequently, for all  $z \in \mathbb{Z}$ ,

$$\lim_{m \rightarrow \infty} \sup_{\theta \in \Theta} \|U_{1,m}(\theta)\| = 0, \quad \tilde{\mathbb{P}}^{\theta_*}\text{-a.s.}$$

Thus,  $\tilde{\mathbb{P}}^{\theta_*}$ -a.s., the sequence  $\{U_{1,m}(\theta) : m \in \mathbb{Z}_+\}$  converges to zero uniformly on  $\Theta$ . Next, we have

$$\begin{aligned}
 \|U_{2,m}(\theta)\| &= \sum_{k=0}^m \left\| \left( \sum_{\ell'=0}^{k-1} \sum_{\ell''=0}^{k-1} \prod_{\ell=0}^{k-1} a_{\ell',\ell'',i,j}(\theta, Y_{-\ell}) \right) b(\theta, Y_{-k}) \right\| \\
 &\leq \sum_{k=0}^m \left( \sum_{\ell'=0}^{k-1} \sum_{\ell''=0}^{k-1} \left\| \prod_{\ell=0}^{k-1} a_{\ell',\ell'',i,j}(\theta, Y_{-\ell}) \right\| \right) \|b(\theta, Y_{-k})\| \\
 &\leq \tilde{M}_2 \sum_{k=0}^m \rho^{k-1} \sup_{\theta \in \Theta} \|b(\theta, Y_{-k})\|.
 \end{aligned}$$

Since  $\lim_{m \rightarrow \infty} \sum_{k=0}^m \rho^{k-1} \sup_{\theta \in \Theta} \|b(\theta, Y_{-k})\| < \infty$ , then  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s., the sequence  $\{U_{2,m}(\theta) : m \in \mathbb{Z}_+\}$  converges uniformly on  $\Theta$ . To show that  $\{U_{3,m}(\theta) : m \in \mathbb{Z}_+\}$  and  $\{U_{4,m}(\theta) : m \in \mathbb{Z}_+\}$  converge uniformly on  $\Theta$ ,  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s., we proceed similarly as above and apply (5.103) and (CG-1)(iv). Finally, by the same argument, the almost surely uniform convergence of  $\{U_{5,m}(\theta) : m \in \mathbb{Z}_+\}$  follows from (CG-1)(ii) and (CG-1)(v). The  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s. uniform convergence of  $\left\{ \frac{\partial^2 \Psi^\theta \langle Y_{-m;0} \rangle(z)}{\partial \theta_j \partial \theta_i} : m \in \mathbb{Z}_+ \right\}$  on  $\Theta$  then implies that  $\theta \mapsto \frac{\partial^2 \Psi^\theta \langle Y_{-\infty;0} \rangle}{\partial \theta_j \partial \theta_i}$  is  $\tilde{\mathbb{P}}^{\theta^*}$ -a.s. well defined and

$$\frac{\partial^2 \Psi^\theta \langle Y_{-\infty;0} \rangle}{\partial \theta_j \partial \theta_i} = \lim_{m \rightarrow \infty} \frac{\partial^2 \Psi^\theta \langle Y_{-m;0} \rangle(z)}{\partial \theta_j \partial \theta_i}, \quad \tilde{\mathbb{P}}^{\theta^*}\text{-a.s.},$$

showing (5.26). The proof for (5.29) follows by similar arguments as for (5.27) and (5.28).  $\square$

### 5.6.2 Proof of Theorem 5.3.12

The proof of this theorem follows similar lines of Douc et al. [2015]. To proceed, let us first recall a more general set of conditions derived in Douc et al. [2013] in which the following definition is required.

**Definition 5.6.1.** Let  $\bar{H}$  be a probability kernel from  $Z^2$  to  $\mathcal{Y}^{\otimes 2} \otimes \mathcal{P}(\{0, 1\})$  satisfying the following marginal conditions, for all  $(z, z') \in Z^2$  and  $B \in \mathcal{Y}$ ,

$$\begin{cases} \bar{H}((z, z'); B \times \mathcal{Y} \times \{0, 1\}) = H(z; B) = G(\Pi_p(z); B), \\ \bar{H}((z, z'); \mathcal{Y} \times B \times \{0, 1\}) = H(z'; B) = G(\Pi_p(z'); B), \end{cases} \quad (5.106)$$

and such that the following coupling condition holds:

$$\bar{H}((z, z'); \{(y, y) : y \in \mathcal{Y}\} \times \{1\}) = \bar{H}((z, z'); \mathcal{Y}^2 \times \{1\}). \quad (5.107)$$

Define the following quantities successively.

- The trace measure of  $\bar{H}((z, z'); \cdot)$  on the set  $\{(y, y) : y \in \mathcal{Y}\} \times \{1\}$  is denoted by

$$\check{H}((z, z'); B) = \bar{H}((z, z'); \{(y, y) : y \in B\} \times \{1\}), \quad B \in \mathcal{Y}.$$

- The probability kernel  $\bar{R}$  from  $(\mathbb{Z}^2, \mathcal{Z}^{\otimes 2})$  to  $(\mathbb{Z}^2 \times \{0, 1\}, \mathcal{Z}^{\otimes 2} \otimes \mathcal{P}(\{0, 1\}))$  is defined for all  $z, z' \in \mathbb{Z}^2$  and  $A \in \mathcal{X}^{\otimes 2}$  by

$$\bar{R}((z, z'); A \times \{1\}) = \int_{\mathcal{Y}} \mathbb{1}_A(\psi_y(z), \psi_y(z')) \check{H}((z, z'); dy).$$

- The measurable function  $\alpha$  from  $\mathbb{Z}^2$  to  $[0, 1]$  is defined by

$$\alpha(z, z') = \bar{R}((z, z'); \mathbb{Z}^2 \times \{1\}) = \bar{H}((z, z'); \mathcal{Y}^2 \times \{1\}). \quad (5.108)$$

- The kernel  $\hat{R}$  is defined for all  $(z, z') \in \mathbb{Z}^2$  and  $A \in \mathcal{Z}^{\otimes 2}$  by

$$\hat{R}((z, z'); A) = \begin{cases} \frac{\bar{R}((z, z'); A \times \{1\})}{\alpha(z, z')} & \text{if } \alpha(z, z') > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (5.109)$$

We can now introduce the so-called *contracting condition* which yields ergodicity.

- (AG-8) There exists a kernel  $\bar{H}$  yielding  $\alpha$  and  $\hat{R}$  as in [Definition 5.6.1](#), a measurable function  $W : \mathbb{Z}^2 \rightarrow [1, \infty)$  satisfying [Conditions \(AG-7\)\(ii\)](#) and [\(AG-7\)\(iii\)](#) and real numbers  $(D, \zeta_1, \zeta_2, \rho) \in (\mathbb{R}_+)^3 \times (0, 1)$  such that for all  $(z, z') \in \mathbb{X}^2$  and, for all  $n \geq 1$ ,

$$\hat{R}^n((z, z'); dz) \leq D\rho^n dz(z, z'), \quad (5.110)$$

$$\hat{R}^n((z, z'); dz \times W) \leq D\rho^n d_{\mathbb{Z}^2}^{\zeta_1}(z, z') W^{\zeta_2}(z, z'). \quad (5.111)$$

Under [Conditions \(AG-3\)](#), [\(AG-4\)](#), [\(AG-5\)](#) and [\(AG-8\)](#) and by combining [Theorem 6](#), [Proposition 8](#) and [Lemma 7](#) in [Douc et al. \[2013\]](#), we immediately obtain the following result.

**Theorem 5.6.2.** *Assume [\(AG-3\)](#), [\(AG-4\)](#), [\(AG-5\)](#) and [\(AG-8\)](#). Then the Markov kernel  $K$  admits a unique invariant distribution  $\pi$  and  $\pi_1(\bar{V}) < \infty$  for any  $\bar{V} : \mathbb{X} \rightarrow \mathbb{R}_+$  such that  $\bar{V} \lesssim V$ .*

Assumptions [\(AG-3\)](#), [\(AG-4\)](#) and [\(AG-5\)](#) are quite usual and easy to check. The key point to obtain ergodicity is thus to construct  $\bar{H}$  satisfying [\(AG-8\)](#). For this, we can also rely on the following result which is an adaption of [[Douc et al., 2013](#), Lemma 9].

**Lemma 5.6.3.** *Assume that there exists  $(\varrho, \beta, D_1, D_2, \ell, \ell') \in (0, 1) \times \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{Z}_+^* \times \mathbb{Z}_+^*$  such that for all  $(z, z') \in \mathbf{Z}^2$ ,*

$$\hat{R}((z, z'); \mathbb{1}_{(D_1 d_{\mathbf{Z}}(z, z'), \infty)} \circ d_{\mathbf{Z}}) = 0, \quad (5.112)$$

$$\hat{R}^\ell((z, z'); \mathbb{1}_{(\varrho d_{\mathbf{Z}}(z, z'), \infty)} \circ d_{\mathbf{Z}}) = 0, \quad (5.113)$$

and

$$\hat{R}W \leq D_2W, \quad (5.114)$$

$$\hat{R}^{\ell'}W \leq W + \beta. \quad (5.115)$$

Then, (5.110) and (5.111) hold.

**Remark 5.6.4.** Lemma 5.6.3 still holds if we replace (5.114) and (5.115) by

$$\hat{R}^{\ell'+r}W \leq W + \beta, \quad r \in \{0, \dots, \ell' - 1\}.$$

*Proof.* Without loss of generality, we can assume that  $\ell = \ell' = m$ , otherwise we can take  $m = \ell \vee \ell'$ . For positive integer  $n$ , write  $n = km + r$ , where  $r \in \{0, \dots, m - 1\}$  and  $k \in \mathbb{Z}_+$ . We have

$$\begin{aligned} \hat{R}^n((z, z'); d_{\mathbf{Z}}) &= \hat{R}^{km+r}((z, z'); d_{\mathbf{Z}}) = \int \hat{R}^{km}((z, z'); dz_1 dz'_1) \hat{R}^r((z_1, z'_1); d_{\mathbf{Z}}) \\ &= \int \hat{R}^{km}((z, z'); dz_1 dz'_1) \hat{R}^r((z_1, z'_1); d_{\mathbf{Z}} \mathbb{1}_{[0, D_1 d_{\mathbf{Z}}(z_1, z'_1)]} \circ d_{\mathbf{Z}}) \\ &\leq \int \hat{R}^{km}((z, z'); dz_1 dz'_1) D_1^r d_{\mathbf{Z}}(z_1, z'_1) \\ &\leq D_1^r \int \hat{R}^{(k-1)m}((z, z'); dz_2 dz'_2) \hat{R}^m((z_2, z'_2); dz_1 dz'_1) d_{\mathbf{Z}}(z_1, z'_1) \\ &\leq \varrho D_1^r \int \hat{R}^{(k-1)m}((z, z'); dz_2 dz'_2) d_{\mathbf{Z}}(z_2, z'_2) \\ &\dots \\ &\leq \varrho^k D_1^r d_{\mathbf{Z}}(z, z') \leq \varrho^{-1} \left( \bigvee_{r=0}^{m-1} D_1^r \right) (\varrho^{1/m})^n d_{\mathbf{Z}}(z, z'). \end{aligned}$$

Similarly, we have

$$\begin{aligned}
 \hat{R}^n((z, z'); d_{\mathbf{Z}} \times W) &= \int \hat{R}^{km}((z, z'); dz_1 dz'_1) \hat{R}^r((z_1, z'_1); d_{\mathbf{Z}} \times W) \\
 &\leq \int \hat{R}^{km}((z, z'); dz_1 dz'_1) D_1^r D_2^r d_{\mathbf{Z}}(z_1, z'_1) W(z_1, z'_1) \\
 &\leq D_1^r D_2^r \int \hat{R}^{(k-1)m}((z, z'); dz_2 dz'_2) \hat{R}^m((z_2, z'_2); dz_1 dz'_1) d_{\mathbf{Z}}(z_1, z'_1) W(z_1, z'_1) \\
 &\leq \varrho D_1^r D_2^r \int \hat{R}^{(k-1)m}((z, z'); dz_2 dz'_2) d_{\mathbf{Z}}(z_2, z'_2) (W(z_2, z'_2) + \beta) \\
 &\dots \\
 &\leq D_1^r D_2^r \varrho^k (W(z, z') + \beta k) d_{\mathbf{Z}}(z, z').
 \end{aligned}$$

Since for all  $\varrho \in (0, 1)$ ,  $\lim_{k \rightarrow \infty} k \varrho^k = 0$ , then there exists  $\bar{D} > 0$  and  $\bar{\varrho} \in [\varrho, 1)$  such that

$$\begin{aligned}
 \hat{R}^n((z, z'); d_{\mathbf{Z}} \times W) &\leq \bar{D} \bar{\varrho}^{k-1} W(z, z') d_{\mathbf{Z}}(z, z') \\
 &\leq \bar{D} (\bar{\varrho}^{1/m})^n W(z, z') d_{\mathbf{Z}}(z, z').
 \end{aligned}$$

Setting  $D = \bar{D} \vee (\varrho^{-1} (\bigvee_{r=0}^{m-1} D_1^r))$  and  $\rho = \varrho^{1/m} \vee \bar{\varrho}^{1/m}$ , we get (5.110) and (5.111).  $\square$

Now we can prove that our set of conditions is sufficient.

*Proof of Theorem 5.3.12.* We only need to show that (AG-6) and (AG-7) imply (AG-8). We preface our proof by the following lemma, which is adapted from [Douc et al., 2015, Lemma 5].

**Lemma 5.6.5.** *Assume (AG-7)(i). Then one can define a kernel  $\bar{H}$  as in Definition 5.6.1 with the same  $\alpha$  given in (5.108). Moreover, the kernel  $\hat{R}$  defined by (5.109) satisfies, for all  $(z, z') \in \mathbf{Z}^2$  such that  $\alpha(z, z') > 0$  and all measurable functions  $f : \mathbf{Z}^2 \rightarrow \mathbb{R}_+$ ,*

$$\hat{R}((z, z'); f) = G(\phi(z, z'); \tilde{f}) \quad \text{with} \quad \tilde{f}(y) = f(\Psi_y(z), \Psi_y(z')). \quad (5.116)$$

Let us complete the proof of Theorem 5.3.12 before proving this lemma. By Lemma 5.6.5 and Lemma 5.6.3, it remains to check that (5.112) and (5.114) hold for all  $(z, z') \in \mathbf{Z}^2$ . Observe that by definition of  $\hat{R}$ , we have for all  $n \in \mathbb{Z}_+^*$ ,  $(z, z') \in \mathbf{Z}^2$  and measurable function  $f : \mathbf{Z}^2 \rightarrow \mathbb{R}$ ,

$$\hat{R}^n((z, z'); f) = J_n((z, z'); f).$$



So, condition (AG-7)(iv) can be equivalently stated as: there exist  $m \in \mathbb{Z}_+^*$  and  $D > 0$  such that

$$\sup_{(z, z') \in \mathbb{Z}^2} \left( \hat{R}^m((z, z'); W) - W(z, z') \right) < \infty$$

and

$$\sup_{(z, z') \in \mathbb{Z}^2} \left( \hat{R}((z, z'); W) - DW(z, z') \right) < \infty.$$

Thus we can find  $(\beta, D_2) \in \mathbb{R} \times (0, \infty)$  such that (5.114) and (5.115) hold for all  $(z, z') \in \mathbb{Z}^2$ .

Now, let  $(z, z') \in \mathbb{Z}^2$ . When  $z = z'$ , then for all  $n \in \mathbb{Z}_+^*$  and  $y_{1:n} \in \mathbb{Y}^n$ ,  $d_{\mathbb{Z}}(\Psi\langle y_{1:n} \rangle(z), \Psi\langle y_{1:n} \rangle(z')) = 0$ , implying that (5.112) and (5.113) hold with any nonnegative  $D_1$  and  $\varrho$ . For  $z \neq z'$ , let  $D_1$  and  $\varrho$  be defined respectively by

$$D_1 = \bigvee_{k=1}^{\ell-1} \sup_{\substack{(z, z', y_{1:k}) \in \mathbb{Z}^2 \times \mathbb{Y} \\ z \neq z'}} \frac{d_{\mathbb{Z}}(\Psi\langle y_{1:k} \rangle(z), \Psi\langle y_{1:k} \rangle(z'))}{d_{\mathbb{Z}}(z, z')},$$

and

$$\varrho = \sup_{\substack{(z, z', y_{1:\ell}) \in \mathbb{Z}^2 \times \mathbb{Y} \\ z \neq z'}} \frac{d_{\mathbb{Z}}(\Psi\langle y_{1:\ell} \rangle(z), \Psi\langle y_{1:\ell} \rangle(z'))}{d_{\mathbb{Z}}(z, z')}.$$

By (AG-6), we have  $D_1 \in (0, \infty)$  and  $\varrho \in (0, 1)$ . Therefore, Conditions (5.112) and (5.113) hold for all  $(z, z') \in \mathbb{Z}^2$  with the above  $D_1$  and  $\varrho$ .  $\square$

We conclude this section with the postponed

*Proof of Lemma 5.6.5.* Let  $(z, z') \in \mathbb{Z}^2$ . We define  $\bar{H}((z, z'); \cdot)$  as the distribution of  $(Y, Y', U)$  drawn as follows. We first draw a random variable  $\bar{Y}$  taking values in  $\mathbb{Y}$  with density  $g(\phi(z, z'); \cdot)$  with respect to  $\nu$ . Then we define  $(Y, Y', U)$  by separating the two cases,  $\alpha(z, z') = 1$  and  $\alpha(z, z') < 1$ .

- Suppose that  $\alpha(z, z') = 1$ . Then from (AG-7)(i), we have

$$H(z; \cdot) = H(z'; \cdot) = G(\phi(z, z'); \cdot).$$

In this case, we set  $(Y, Y', U) = (\bar{Y}, \bar{Y}, 1)$ .

- Suppose now that  $\alpha(z, z') < 1$ . Then, using (5.37), the functions

$$(1 - \alpha(z, z'))^{-1} [g(\Pi_p(z); \cdot) - \alpha(z, z')g(\phi(z, z'); \cdot)]$$

and

$$(1 - \alpha(z, z'))^{-1} [g(\Pi_p(z'); \cdot) - \alpha(z, z')g(\phi(z, z'); \cdot)],$$

are probability density functions with respect to  $\nu$  and we let  $\Lambda$  and  $\Lambda'$  be two independent random variables taking values in  $\mathsf{Y}$  drawn with these two density functions, respectively. In this case we draw  $U$  independently according to a Bernoulli variable with mean  $\alpha(z, z')$  and set

$$(Y, Y') = \begin{cases} (\bar{Y}, \bar{Y}') & \text{if } U = 1, \\ (\Lambda, \Lambda') & \text{if } U = 0. \end{cases}$$

One can easily check that the so defined kernel  $\bar{H}$  satisfies (5.106) and (5.107). Moreover, for all  $(z, z') \in \mathsf{Z}^2$ ,

$$\bar{H}((z, z'); \mathsf{Y}^2 \times \{1\}) = \mathbb{P}(U = 1) = \alpha(z, z'),$$

which is compatible with (5.108). The kernel  $\hat{R}$  is defined by setting  $\hat{R}((z, z'); \cdot)$  as the conditional distribution of  $(Z, Z') = (\Psi_Y(z), \Psi_Y(z'))$  given that  $U = 1$ . To complete the proof of Lemma 5.6.5, observe that for any measurable  $f : \mathsf{Z}^2 \rightarrow \mathbb{R}_+$ , we have, for all  $(z, z') \in \mathsf{Z}^2$  such that  $\alpha(z, z') > 0$ ,

$$\begin{aligned} \hat{R}((z, z'); f) &= \mathbb{E}[f(\Psi_Y(z), \Psi_Y(z')) \mid U = 1] \\ &= \mathbb{E}[f(\Psi_{\bar{Y}}(z), \Psi_{\bar{Y}}(z'))] \\ &= G(\phi(z, z'); \tilde{f}), \end{aligned}$$

where  $\tilde{f}(y) = f(\Psi_y(z), \Psi_y(z'))$  for all  $y \in \mathsf{Y}$ . □

### 5.6.3 Proof of Lemma 5.4.2

If  $a_i \geq 0$ ,  $b_i \geq 0$ , for all  $i \in \{1, \dots, p\}$ , and  $\sum_{i=1}^p (a_i + b_i) < 1$ , it is trivial that  $|\lambda|_{\max}(\mathbf{A} + \mathbf{b}\boldsymbol{\delta}_p^T) < 1$ . Otherwise, there would exist some  $\lambda$  such that  $|\lambda|^{-1} \leq 1$  and

$$\left| \sum_{i=1}^p (a_i + b_i) \lambda^{-i} \right| \leq \sum_{i=1}^p (a_i + b_i) |\lambda|^{-i} < 1.$$

This would imply  $P(\lambda) = \lambda^{2p-1} (1 - \sum_{k=1}^p (a_k + b_k) \lambda^{-k}) \neq 0$ . Suppose on the other hand that  $\lambda^* = |\lambda|_{\max}(\mathbf{A} + \mathbf{b}\boldsymbol{\delta}_p^T) < 1$ . From (5.46), we can write  $P(\lambda)$  as a product of  $\lambda^{p-1}$  and  $Q(\lambda) = \lambda^p - \sum_{k=1}^p (a_k + b_k) \lambda^{p-k}$ . Note that  $\lambda^* = 0$  if and only if  $\sum_{i=1}^p (a_i + b_i) = 0$ . Then we can suppose that  $0 < \lambda^* < 1$ , then there exists  $j \in \{1, \dots, p\}$  such that  $a_j + b_j > 0$ . Furthermore, note that for all  $i \in \{1, \dots, p\}$ ,  $(a_i + b_i) \lambda^{*i} \geq (a_i + b_i) \lambda^{*p}$  and this inequality is strict

when  $i = j$ . Thus from Perron-Frobenius theorem for nonnegative matrix,  $Q(\lambda^*) = 0$ , and we have

$$0 < \lambda^{*p} \left( 1 - \sum_{i=1}^p (a_i + b_i) \right).$$

This completes the proof since  $\lambda^* > 0$ .

#### 5.6.4 Proof of Lemma 5.4.11

If  $\vartheta \leq 0$ , then (5.59) holds since  $(1 + Y)^\vartheta \leq 1$ . Note that for  $s \in (0, 1]$ , we have

$$\mathbb{E}[(1 + Y)^s] \leq (\mathbb{E}[1 + Y])^s \leq (1 + \zeta)^s \leq 1 + \zeta^s.$$

Thus, (5.59) holds when  $\vartheta \in (0, 1]$ . For  $\vartheta \in (1, 3/2]$ , then there exists  $s_1 \in (0, 1/2]$  such that  $\vartheta = 1 + s_1$ . By Cauchy-Schwartz inequality, we have

$$\begin{aligned} \mathbb{E}[(1 + Y)^\vartheta] &= \mathbb{E}[(1 + Y)(1 + Y)^{s_1}] \leq (\mathbb{E}[(1 + Y)^2])^{1/2} (\mathbb{E}[(1 + Y)^{2s_1}])^{1/2} \\ &\leq (1 + 3\zeta + \zeta^2)^{1/2} (1 + \zeta)^{s_1} \leq c_1 + c_2 \zeta^\vartheta, \end{aligned}$$

for some constants  $c_1, c_2 > 0$ . If  $\vartheta \in (3/2, 2]$ , then there exists  $s_2 \in (0, 1/2]$  such that  $\vartheta = 3/2 + s_2$  and

$$\begin{aligned} \mathbb{E}[(1 + Y)^\vartheta] &= \mathbb{E}[(1 + Y)^{3/2}(1 + Y)^{s_2}] \leq (\mathbb{E}[(1 + Y)^3])^{1/2} (\mathbb{E}[(1 + Y)^{2s_2}])^{1/2} \\ &\leq (1 + 7\zeta + 6\zeta^2 + \zeta^3)^{1/2} (1 + \zeta)^{s_2} \leq c_3 + c_4 \zeta^\vartheta, \end{aligned}$$

for some constants  $c_3, c_4 > 0$ . Finally, if  $\vartheta > 2$ , then there exists  $s_3 \in (0, 1]$  such that  $\vartheta = \lfloor \vartheta \rfloor + s_3$  and we have

$$\begin{aligned} \mathbb{E}[(1 + Y)^\vartheta] &= \mathbb{E}[(1 + Y)^{\lfloor \vartheta \rfloor} (1 + Y)^{s_3}] \\ &\leq (\mathbb{E}[(1 + Y)^{2\lfloor \vartheta \rfloor}])^{1/2} (\mathbb{E}[(1 + Y)^{2s_3}])^{1/2} \\ &\leq \left( 1 + 2^{\lfloor \vartheta \rfloor} (\mathbb{E}[Y^{2\lfloor \vartheta \rfloor}])^{1/2} \right) (c_5 + c_6 \zeta^{s_3}), \end{aligned}$$

for some constants  $c_5, c_6 > 0$ . The proof then follows by observing that

$$\lim_{\zeta \rightarrow \infty} \frac{(\mathbb{E}[Y^{2\lfloor \vartheta \rfloor}])^{1/2}}{\zeta^{\lfloor \vartheta \rfloor}} = 1.$$

**Lemma 5.6.6.** *Let  $\{U_k : k \in \mathbb{Z}_+\}$  be a stationary sequence of real-valued random variables on  $(\Omega, \mathcal{F}, \mathbb{P})$ . Assume that  $\mathbb{E}(\ln^+ |U_0|) < \infty$ . Then, for all  $\eta \in (0, 1)$ ,*

$$\lim_{k \rightarrow \infty} \eta^k U_k = 0, \quad \mathbb{P}\text{-a.s.}$$

*Proof.* See [Douc et al., 2013, Lemma 34]. □

## Chapter 6

# Conclusions and Future Perspectives

This thesis focuses on developing theory of parameter estimation for a general class of time series models, namely, the class of partially observed bivariate Markov chains. Time series models belonging to or embedded in this class that yield potential applications to time series of counts are particularly considered.

We establish under very mild assumptions that the limit of the normalized log-likelihood is maximized when the parameters belong to the equivalence class of the true parameter, serving as a key step for obtaining the consistency of the MLE in well-specified models. This result is obtained in the general framework of partially observed Markov models and is addressed by relying on the uniqueness of the invariant distribution of the Markov chain associated to the complete data, regardless of its rate of convergence to the equilibrium. Due to its generality this result can be applied to many classes of interest including the class of HMMs and the class of observation-driven models, an important class for dealing with time series of counts.

The consistency of the MLE in the sense of [Leroux \[1992\]](#) is developed independently for the class of fully dominated partially observed Markov models and the class of observation-driven models. The result obtained in the former class is derived under mild sufficient conditions, covering many interesting models such as HMMs, Markov switching models and some versions of Markov chains in random environment. For observation-driven models, the result is derived, for both first-order and general-order settings, under sets of assumptions that are easy to check in specific examples. In addition, strong consistency and the asymptotic normality of the MLE are also treated for some special but important cases of general-order observation-driven models.

This thesis can be further enriched and can lead to future research in

several directions.

1. In our proof of the identification of the maximizing set of the limit of the normalized log-likelihood, we require the Markov structure of the models. Could this result still hold for other general settings other than Markov models?
2. Now a unified approach for identifying the maximizing set of the normalized log-likelihood for the class of partially observed Markov models is obtained. However, to our best knowledge, showing that the MLE converges for this class is still carried out by separating into cases. Could a unified result for this problem be obtained also for the class of partially observed Markov models?
3. The proof of the asymptotic normality of the MLE generally requires strong consistency of the MLE. However, whether the equivalence-class consistency is sufficient for the proof has not been known yet. Thus, it would be worthwhile if we could show that this is indeed the case.
4. The proof of the asymptotic normality of the MLE for the general-order observation-driven models is obtained only in a specific form of the hidden state equation. This result can of course be extended to greater generality under similar approach.

# Résumé en français

## 1 Introduction

Cette thèse s'est déroulée au sein du Laboratoire Traitement et Communication de l'Information (LTCI) à Télécom ParisTech, sous la co-direction de François Roueff, Professeur à Télécom ParisTech, et de Randal Douc, Professeur à Télécom SudParis, de mars 2013 à mars 2016. Elle se concentre sur le développement de la théorie de l'estimation du maximum de vraisemblance pour une classe générale de modèles de séries temporelles, c'est-à-dire la classe de chaînes de Markov bivariées partiellement observées. Appartenant ou étant compris dans cette classe, les modèles de séries temporelles qui fournissent des applications potentielles aux séries temporelles de comptage sont particulièrement considérés.

### 1.1 Contexte

Un modèle de Markov partiellement observé ou *partially observed Markov model* (PMM) est un processus de Markov bivarié où seule une partie du processus complet est observé – d'où les termes « partiellement observé », et pour l'autre composante, la variable d'état, les termes « caché » ou « inaperçu » –, et où toute inférence statistique doit être faite en utilisant des échantillons de réalisations du processus d'observation. La classe des PMM contient notamment deux classes remarquables de modèles statistiques communément utilisées pour examiner les séquences temporellement dépendantes et les données de séries temporelles, c'est-à-dire la classe des modèles de Markov cachés ou *hidden Markov models* (HMM) et la classe des modèles « observation-driven » ou *observation-driven models* (ODM). Une propriété commune et cependant importante à ces deux sous-classes réside dans le fait que les processus cachés des variables d'état sont des chaînes de Markov par nature. Cette propriété permet d'établir des propriétés statistiques importantes telles que la stationnarité et l'ergodicité pour les chaînes entières en s'appuyant sur les propriétés statistiques héritées des chaînes cachées.

Les HMM sont largement appliqués dans de nombreux domaines, de la biologie (Churchill [1992], Krogh et al. [1994]) à l'analyse des séries temporelles (MacDonald and Zucchini [1997]), en passant par l'économétrie (Hull and White [1987], Mamon and Elliott [2007]), la finance (Zhang [2004]), la neurophysiologie (Fredkin and Rice [1987]), le traitement du signal (Crouse et al. [1998], Krishnamurthy and Chung [2007]) et la reconnaissance vocale (Rabiner [1989], Huang et al. [1990], Juang and Rabiner [1991]). Théorie et inférence pour cette classe de modèles ont été développées par Baum et ses collaborateurs dans une série de contributions : Baum and Petrie [1966], Baum and Eagon [1967], Baum and Sell [1968], Petrie [1969], Baum et al. [1970], Baum [1972] (voir aussi Ephraim and Merhav [2002] pour une brève note historique sur les HMM). Ces études ont été par la suite développées dans une série de travaux par Finesso [1991], Leroux [1992], Mevel [1997], Bickel et al. [1998], Jensen and Petersen [1999], Le Gland and Mevel [2000], Douc and Matias [2001], Bickel et al. [2002], Mevel and Finesso [2004], Genon-Catalot and Laredo [2006], Douc et al. [2009, 2011], Douc and Moulines [2012] (voir aussi Cappé et al. [2005] et les références qui y figurent pour une discussion approfondie). Bien qu'il y ait eu pour cette classe importante des études substantielles, aucun des résultats théoriques ne sont encore jusqu'ici devenus standards et ils ont donc toujours besoin d'être améliorés.

Les ODM ont été introduits par Cox [1981] pour décrire des modèles statistiques ayant la caractéristique clé que la variable d'état cachée peut être représentée explicitement comme une fonction déterministe des variables de l'observation passée. Cette propriété est intéressante d'un point de vue pratique, puisque la fonction de vraisemblance (conditionnelle) associée et ses dérivés sont faciles à calculer, l'estimation du paramètre est donc relativement simple, et la prédiction, qui est de première importance dans beaucoup d'applications de séries temporelles, est immédiate. Le célèbre modèle GARCH(1,1) introduit par Bollerslev [1986], de même que la plupart des modèles dérivés de celui-ci, sont des exemples typiques d'ODM (voir Bollerslev [2008] pour une liste de certains d'entre eux). De plus, ces modèles sont particulièrement utiles pour traiter les données de séries temporelles de nature discrète, quand les modèles classiques autorégressifs ne sont plus appropriés (voir [Kedem and Fokianos, 2002, Chapitre 4] pour un compte-rendu complet et Tjøstheim [2012] pour un examen récent). Les applications des ODM peuvent être trouvées dans une large palette de disciplines, telles que l'économie (Pindyck and Rubinfeld [1998]), les études environnementales (Bhaskaran et al. [2013]), l'épidémiologie et les études de santé publique (Zeger [1988], Davis et al. [1999], Ferland et al. [2006]), la finance (Liesenfeld and Richard [2003], Rydberg and Shephard [2003], Fokianos and Tjøstheim [2011], Francq and Zakoian [2011]) et la dynamique des populations (Ives

et al. [2003]).

Théorie et inférence pour cette classe de modèles, quand les observations ont des valeurs d'entiers, ont été développées dans une série de contributions dont [Streett \[2000\]](#), [Davis et al. \[2003\]](#), [Heinen \[2003\]](#), [Ferland et al. \[2006\]](#), [Fokianos et al. \[2009\]](#), [Franke \[2010\]](#), [Woodard et al. \[2010\]](#), [Fokianos and Tjøstheim \[2011\]](#), [Henderson et al. \[2011\]](#), [Neumann \[2011\]](#), [Davis and Liu \[2012\]](#), [Doukhan et al. \[2012\]](#), [Douc et al. \[2013\]](#), [Fokianos et al. \[2013\]](#), [Christou and Fokianos \[2014, 2015\]](#) et [Douc et al. \[2015\]](#). Comme dans le cas des HMM, aucun des résultats théoriques ne sont encore jusqu'ici devenus standards et ils ont donc toujours besoin d'être améliorés.

Au-delà des HMM et des ODM, la classe des PMM comprend beaucoup d'autres modèles intéressants ayant des applications potentielles dans les séries temporelles et d'autres domaines, comme les modèles de Markov à changement de régime (voir [Hamilton \[1988, 1989\]](#)), beaucoup des chaînes de Markov en environnement aléatoire ou *Markov chains in random environment* introduites par [Cogburn \[1980\]](#) et [Bourgin and Cogburn \[1981\]](#), les champs aléatoires de Markov par paires ou *pairwise Markov random fields* (voir [Pieczynski and Tebbache \[2000\]](#)) et d'autres cas étudiés par exemple par [Pieczynski \[2003\]](#) et [Ephraïm and Mark \[2012\]](#). Cependant, théorie et inférence pour ce cadre général restent encore insuffisantes dans la littérature.

## 1.2 Motivations et problématiques

Cette thèse est principalement sous-tendue par les questions suivantes, rencontrées pour les HMM et les ODM en particulier (et par conséquent dans les PMM en général).

### 1.2.1 Consistance

L'estimation du maximum de vraisemblance (EMV) est une méthode naturelle et répandue pour identifier un modèle paramétrique de séries temporelles à partir d'un échantillon de réalisations du processus observé. Dans le cadre d'un modèle bien spécifié, c'est-à-dire quand la loi des observations est entièrement décrite par un membre paramétré de la classe des PMM, il est d'un grand intérêt d'obtenir la consistance de l'estimateur, à savoir sa convergence vers le vrai paramètre, soit  $\theta_*$ , lorsque la taille de l'échantillon tend vers l'infini. La démonstration comprend en général deux étapes importantes :

1. l'EMV converge vers l'ensemble  $\Theta_*$  qui maximise la log-vraisemblance asymptotique normalisée ; et
2. cet ensemble  $\Theta_*$  se réduit à l'ensemble du seul vrai paramètre  $\theta_*$ .



La deuxième étape est souvent mentionnée comme résolvant le problème d'*identifiabilité*, mais elle peut en réalité être divisée en deux sous-problèmes :

- 2.1 montrer que tout paramètre de l'ensemble des maxima  $\Theta_*$  produit la même distribution du processus d'observation que celle du vrai paramètre  $\theta_*$ ; et
- 2.2 montrer que pour une taille d'échantillons suffisamment large, l'ensemble de tels paramètres se réduit à  $\{\theta_*\}$ .

Cependant, au moins dans le cas des HMM, le problème 2.2 ci-dessus apparaît comme difficile à résoudre; voir [Allman et al. \[2009\]](#), [Gassiat et al. \[2016\]](#) et les références qui y figurent pour s'informer des dernières avancées concernant les HMM. Effectivement, le problème 2.1 peut être résolu de manière indépendante, et avec l'étape 1 vue plus haut, ceci implique directement que l'EMV est consistant dans un sens faible, à savoir que le paramètre estimé converge vers l'ensemble de tous les paramètres associés à la même loi que celle de l'échantillon observé. La consistance dans ce sens est appelée *consistance de classe d'équivalence*, comme introduite par [Leroux \[1992\]](#).

### ***Problème de consistance dans les HMM***

Pour les HMM, la consistance de l'EMV est de grande importance; elle se présente autant comme sujet d'étude (voir [Baum and Petrie \[1966\]](#), [Petrie \[1969\]](#), [Leroux \[1992\]](#), [Le Gland and Mevel \[2000\]](#), [Le Gland and Mevel \[2000\]](#), [Douc and Matias \[2001\]](#), [Douc et al. \[2004\]](#), [Genon-Catalot and Laredo \[2006\]](#), [Douc et al. \[2011\]](#)) que comme hypothèse de base (voir [Bickel et al. \[1998\]](#), [Jensen and Petersen \[1999\]](#)). Parmi ces contributions, le résultat obtenu dans l'article de [Douc et al. \[2011\]](#) semble assez général et peut être appliqué à plusieurs versions de HMM ayant des espaces d'état et des espaces d'observation généraux. Dans cette contribution, l'identification de l'ensemble des maxima  $\Theta_*$  de la log-vraisemblance asymptotique normalisée est réalisée en utilisant la condition de *séparation exponentielle des mesures*, qui est apparemment simple à appliquer quand le modèle sous-jacent est géométriquement ergodique. Cependant, pour les modèles non-géométriquement ergodiques tels que la chaîne de Markov polynomialement ergodique (voir [Tuominen and Tweedie \[1994\]](#) et [Jarnier and Roberts \[2002\]](#) pour l'existence et quelques études d'un tel exemple), vérifier cette condition peut être laborieux et compliqué. La caractérisation de l'ensemble des maxima  $\Theta_*$  (et donc de la consistance de classe d'équivalence de l'EMV) reste par conséquent une question délicate pour les HMM.

***Problème de la consistance dans les ODM***

Dans les ODM, la fonction de vraisemblance et ses dérivés sont faciles à calculer, ce qui rend l'estimation au moins numériquement possible. La consistance de l'EMV peut cependant être fastidieuse et est souvent obtenue à l'aide de calculs spécifiques au modèle étudié. Pour la variable observée discrète en particulier, des résultats généraux de consistance ont été obtenus récemment dans [Davis and Liu \[2012\]](#) ou [Douc et al. \[2013\]](#) (voir aussi [Henderson et al. \[2011\]](#) pour l'existence de solutions stationnaires et ergodiques pour quelques ODM). Néanmoins, dans ces contributions, la manière de prouver que l'ensemble des maxima  $\Theta_*$  se réduit à  $\{\theta_*\}$  nécessite la vérification de conditions spécifiques dans chaque exemple donné, et semble avoir des difficultés à affirmer le résultat dans un contexte plus général, par exemple lorsque la distribution des observations compte tenu de la variable cachée dépend également d'un paramètre inconnu. Des exemples de ces modèles sont le modèle NBIN-GARCH introduit par [Zhu \[2011\]](#) et le modèle NM-GARCH proposé par [Haas et al. \[2004\]](#) et [Alexander and Lazar \[2006\]](#). À notre connaissance, la consistance (complète) de l'EMV n'a pas été traitée pour le modèle NBIN-GARCH (voir [Davis and Liu \[2012\]](#) pour un traitement partiel); et la preuve usuelle de la consistance de l'EMV pour le modèle GARCH ne peut pas être directement adaptée au modèle NM-GARCH.

Il est donc d'un grand intérêt d'aborder ce problème de consistance dans l'esprit général des PMM, et de savoir si l'approche actuelle est en mesure de rendre compte des versions insolubles des HMM et des ODM mentionnées ci-dessus.

**1.2.2 ODM d'ordre général**

Dans de nombreuses séries temporelles, la résolution du problème 1 et du problème 2.1 de la section 1.2.1, soit l'établissement la consistance de l'EMV, nécessite généralement de supposer ou de prouver que le processus d'observation est stationnaire et ergodique. Dans le cadre des HMM, les solutions stationnaires et ergodiques pour le modèle sont héritées de la stationnarité et de l'ergodicité du processus caché sous-jacent et sont généralement étudiées en utilisant l'hypothèse de  $\varphi$ -irréductibilité. Cependant, il n'en va pas de même lorsque le modèle étudié est placé dans le cadre des ODM. Dans ce dernier cas, il s'avère qu'une approche plus subtile est nécessaire, au moins dans les cas pour lesquels le processus observé prend des valeurs entières. La difficulté alors, bien que par nature il soit une chaîne de Markov, peut provenir de la dégénérescence du processus caché de l'état de son équation d'état. À ce stade, plusieurs méthodes de résolution de ce problème ont été

proposées. Celles-ci comprennent la technique de perturbation (voir [Fokianos and Tjøstheim \[2011\]](#)), l'approche de la contractilité (voir [Neumann \[2011\]](#)), l'approche de la dépendance faible (voir [Doukhan et al. \[2012\]](#)) et l'approche basée sur la théorie des chaînes de Markov sans irréductibilité introduite par [Douc et al. \[2013\]](#). Parmi ces méthodes, le résultat obtenu récemment par [Douc et al. \[2013\]](#) semble capable de fonctionner avec de nombreux modèles d'intérêt se trouvant dans la classe d'ODM de premier ordre, indépendamment du fait que le processus d'observation soit discret ou continu. Cependant, on ne sait pas jusqu'à présent si ce même résultat s'applique à un contexte plus général et plus souple d'ODM d'ordre général ou de modèles de type GARCH d'ordre général. La même question, telle que soumise par [Douc et al. \[2013\]](#) et mise en lumière par [Tjøstheim \[2015\]](#), se pose pour les propriétés asymptotiques de l'EMV pour la classe des ODM d'ordre général.

### 1.3 Organisation de la thèse

L'objectif principal de cette thèse est d'apporter des réponses aux problèmes spécifiques mentionnés ci-dessus. Elle est structurée comme suit. Dans la section [2](#), nous fournissons le résultat du problème [2.1](#), c'est-à-dire l'identification de l'ensemble des maxima de la log-vraisemblance asymptotique normalisée, sous des hypothèses assez faibles et dans le cadre général des PMM. La section [3](#) est dédiée à l'étude de la convergence et de la consistance (de classe d'équivalence) de l'EMV pour la classe des PMM entièrement dominés. La section [4](#) traite de l'ergodicité et de la consistance de l'EMV pour la classe des ODM de premier ordre. La section [5](#) généralise les résultats de la section [4](#) à la classe des ODM d'ordre général. La normalité asymptotique de l'EMV est également traitée dans ce cadre général. Conclusions et perspectives d'avenir de la thèse sont présentées dans la section [6](#).

### 1.4 Conférences et publications

#### Conférences avec actes :

- C1** R. Douc, F. Roueff et T. Sim. Identifiability conditions for partially-observed Markov chains. *11th International Vilnius Conference on Probability Theory and Mathematical Statistics*, Juin 2014.
- C2** R. Douc, F. Roueff et T. Sim. The maximizing set of the asymptotic normalized log-likelihood for partially observed Markov chains. *Paris-Southeast Asia Conference in Mathematical Finance*, Février 2015.

**Articles de journal :**

- J1** R. Douc, F. Roueff et T. Sim. Handy sufficient conditions for the convergence of the maximum likelihood estimator in observation-driven models. *Lithuanian Mathematical Journal*, 55(3) : 367-392, 2015. URL <https://hal.archives-ouvertes.fr/hal-01080955>.
- J2** R. Douc, F. Roueff et T. Sim. The maximizing set of the asymptotic normalized log-likelihood for partially observed Markov chains. À paraître bientôt dans *Annals of Applied Probability*. URL <https://hal.archives-ouvertes.fr/hal-01080955>.

## 2 Identification de l'ensemble des maxima de la log-vraisemblance asymptotique normalisée dans les PMM

### 2.1 Cadre général : modèles de Markov partiellement observés

Soient  $(\mathsf{X}, \mathcal{X})$  et  $(\mathsf{Y}, \mathcal{Y})$  deux espaces boréliens et  $\Theta$  un ensemble de paramètres. Considérons un modèle statistique déterminé par une classe de noyaux markoviens  $\{K^\theta : \theta \in \Theta\}$  sur  $(\mathsf{X} \times \mathsf{Y}) \times (\mathcal{X} \otimes \mathcal{Y})$ . Notons  $\mathbb{P}_\xi^\theta$  la loi de probabilité (et  $\mathbb{E}_\xi^\theta$  l'espérance correspondante) induite sur  $(\mathsf{X} \times \mathsf{Y})^{\mathbb{Z}_+}$  par une chaîne de Markov  $\{(X_k, Y_k) : k \in \mathbb{Z}_+\}$  ayant le noyau de transition  $K^\theta$  et la distribution initiale  $\xi$  sur  $\mathsf{X} \times \mathsf{Y}$ . Dans le cas où  $\xi$  est une masse de Dirac en  $(x, y)$ , nous allons simplement écrire  $\mathbb{P}_{(x,y)}^\theta$ . Pour les chaînes de Markov partiellement observées, c'est-à-dire quand un échantillon  $Y_{1:n} := (Y_1, \dots, Y_n) \in \mathsf{Y}^n$  du deuxième composant est observé seulement, il est pratique d'écrire  $K^\theta$  tel que

$$K^\theta((x, y); dx'dy') = Q^\theta((x, y); dx')G^\theta((x, y, x'); dy'), \quad (1)$$

où  $Q^\theta$  et  $G^\theta$  sont des noyaux de probabilité sur  $(\mathsf{X} \times \mathsf{Y}) \times \mathcal{X}$  et sur  $(\mathsf{X} \times \mathsf{Y} \times \mathsf{X}) \times \mathcal{Y}$ , respectivement. Nous considérons maintenant le cadre général suivant.

**Définition 2.1.** On dit que le modèle de Markov  $\{K^\theta : \theta \in \Theta\}$  de la forme (1) est partiellement dominé s'il existe une mesure  $\sigma$ -finie  $\nu$  sur  $\mathsf{Y}$  telle que pour tout  $(x, y), (x', y') \in \mathsf{X} \times \mathsf{X}$ ,

$$G^\theta((x, y, x'); dy') = g^\theta((x, y, x'); y')\nu(dy'), \quad (2)$$

où la fonction de densité conditionnelle  $g^\theta$  satisfait par ailleurs

$$g^\theta((x, y, x'); y') > 0, \quad \text{pour tout } (x, y), (x', y') \in \mathbf{X} \times \mathbf{Y}. \quad (3)$$

Il résulte de (2) que, pour tout  $(x, y) \in \mathbf{X} \times \mathbf{Y}$ ,  $A \in \mathcal{X}$  et  $B \in \mathcal{Y}$ ,

$$K^\theta((x, y); A \times B) = \int_B \kappa^\theta\langle y, y' \rangle(x; A) \nu(dy'),$$

où pour tout  $y, y' \in \mathbf{Y}$ ,  $\kappa^\theta\langle y, y' \rangle$  est un noyau défini sur  $(\mathbf{X}, \mathcal{X})$  par

$$\kappa^\theta\langle y, y' \rangle(x; dx') := Q^\theta((x, y); dx') g^\theta((x, y, x'); y'). \quad (4)$$

La dynamique du processus de Markov bivarié partiellement observé peut être représentée comme dans la figure 2.1.

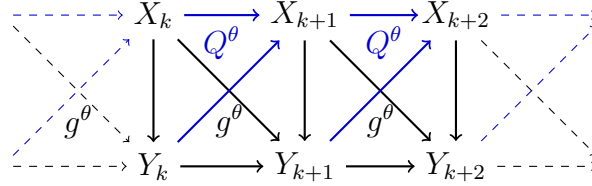


FIGURE 2.1 – Représentation graphique du modèle de Markov bivarié partiellement observé.

Notons que, en général, le noyau  $\kappa^\theta\langle y, y' \rangle$  est non normalisé car  $\kappa^\theta\langle y, y' \rangle(x; \mathbf{X})$  peut être différent de un. En outre, du fait que  $Q^\theta((x, y); \cdot)$  est une probabilité sur  $(\mathbf{X}, \mathcal{X})$  et de la condition (3), on a pour tout  $(x, y, y') \in \mathbf{X} \times \mathbf{Y} \times \mathbf{Y}$ ,  $\kappa^\theta\langle y, y' \rangle(x; \mathbf{X}) > 0$ .

Dans le cas des modèles bien spécifiés, il est supposé que les observations  $Y_{1:n}$  sont générées par un processus  $\{(X_k, Y_k) : k \in \mathbb{Z}_+\}$ , qui suit la distribution  $\mathbb{P}_{\xi_\star}^{\theta_\star}$  associée à un paramètre inconnu  $\theta_\star \in \Theta$  et une distribution initiale inconnue  $\xi_\star$  (généralement,  $\xi_\star$  est telle que, sous la loi  $\mathbb{P}_{\xi_\star}^{\theta_\star}$ ,  $\{Y_k : k \in \mathbb{Z}_+\}$ , c'est une séquence stationnaire). Pour former une estimation consistante de  $\theta_\star$  sur la base des observations  $Y_{1:n}$  seulement, c'est-à-dire sans accès au processus caché  $\{X_k : k \in \mathbb{Z}_+\}$ , on définit l'estimateur du maximum de vraisemblance (EMV)  $\hat{\theta}_{\xi, n}$  par

$$\hat{\theta}_{\xi, n} \in \operatorname{argmax}_{\theta \in \Theta} L_{\xi, n}(\theta), \quad (5)$$

où  $L_{\xi,n}(\theta)$  est la fonction de la log-vraisemblance (conditionnelle) des observations sous le paramètre  $\theta$  avec une distribution initiale arbitraire  $\xi$  sur  $\mathbf{X} \times \mathbf{Y}$ , c'est-à-dire

$$\begin{aligned} L_{\xi,n}(\theta) &:= \ln \int \prod_{k=0}^{n-1} Q^\theta((x_k, y_k); dx_{k+1}) g^\theta((x_k, y_k, x_{k+1}); y_{k+1}) \xi(dx_0 dy_0) \\ &= \ln \int \kappa^\theta\langle y_0, y_1 \rangle \kappa^\theta\langle y_1, y_2 \rangle \cdots \kappa^\theta\langle y_{n-1}, y_n \rangle(x_0; \mathbf{X}) \xi(dx_0 dy_0). \end{aligned} \quad (6)$$

Cela correspond au logarithme de la densité conditionnelle de  $Y_{1:n}$  sachant  $(X_0, Y_0)$ , densité qui est intégrée selon  $\xi$ . Dans la pratique,  $\xi$  est souvent pris comme une masse de Dirac en  $(x, y)$  avec  $x$  arbitrairement choisi et  $y$  égal à l'observation  $Y_0$  quand elle est disponible. Dans ce contexte, une façon classique (voir par exemple [Leroux \[1992\]](#)) de prouver la consistance de l'EMV  $\hat{\theta}_{\xi,n}$  peut être décomposée dans les étapes suivantes. La première étape est de montrer que  $\hat{\theta}_{\xi,n}$  est, avec une probabilité qui tend vers un, dans un voisinage de l'ensemble

$$\Theta_\star := \operatorname{argmax}_{\theta \in \Theta} \tilde{\mathbb{E}}^{\theta_\star} [\ln p^{\theta, \theta_\star}(Y_1 | Y_{-\infty:0})]. \quad (7)$$

Cette formule implique deux quantités qui n'ont pas encore été définies, car elles peuvent nécessiter des hypothèses supplémentaires : la première est l'espérance  $\tilde{\mathbb{E}}^\theta$ , qui correspond à la distribution  $\tilde{\mathbb{P}}^\theta$  d'une séquence  $\{Y_k : k \in \mathbb{Z}\}$  conformément au noyau  $K^\theta$  ; la seconde est la densité  $p^{\theta, \theta_\star}(\cdot | \cdot)$ , qui apparaît lors de la prise de la limite, sous la probabilité  $\tilde{\mathbb{P}}^{\theta_\star}$ , de la  $\tilde{\mathbb{P}}^\theta$ -densité conditionnelle de  $Y_1$  sachant  $Y_{-m:0}$ , quand  $m$  tend vers l'infini. Dans de nombreux cas, ces quantités apparaissent naturellement parce que le modèle est ergodique et que la log-vraisemblance normalisée  $n^{-1}L_{\xi,n}(\theta)$  peut être approchée par

$$\frac{1}{n} \sum_{k=1}^n \ln p^{\theta, \theta_\star}(Y_k | Y_{-\infty:k-1}).$$

Nous allons fournir ci-dessous certaines hypothèses générales, l'hypothèse **(K-1)** et l'hypothèse **(K-2)**, qui donnent des définitions précises de  $\tilde{\mathbb{P}}^\theta$  et  $p^{\theta, \theta'}(\cdot | \cdot)$ . La deuxième étape consiste à prouver que l'ensemble  $\Theta_\star$  dans (7) est lié au vrai paramètre  $\theta_\star$  d'une manière exploitable. Idéalement, on pourrait avoir  $\Theta_\star = \{\theta_\star\}$ , ce qui fournirait la consistance de  $\hat{\theta}_{\xi,n}$  pour estimer  $\theta_\star$ . Dans ce travail, notre premier objectif est de fournir un ensemble d'hypothèses générales qui assurent que  $\Theta_\star$  est exactement l'ensemble des paramètres  $\theta$  tels que  $\tilde{\mathbb{P}}^\theta = \tilde{\mathbb{P}}^{\theta_\star}$ . Puis ce résultat garantit que l'estimateur converge vers l'ensemble des paramètres compatibles avec la vraie distribution stationnaire

des observations. De plus, si le modèle  $\{\tilde{\mathbb{P}}^\theta : \theta \in \Theta\}$  est identifiable, alors cet ensemble se réduit à  $\{\theta_\star\}$  et la consistance de  $\hat{\theta}_{\xi,n}$  est directement obtenue.

Pour compléter notre cadre général, nous énonçons l'hypothèse principale concernant le modèle, la notation qui en découle et les définitions utilisées tout au long de la thèse.

(K-1) Pour tout  $\theta \in \Theta$ , le noyau  $K^\theta$  admet uniquement la probabilité stationnaire  $\pi^\theta$ .

**Définition 2.2.** Sous l'hypothèse (K-1), on note par  $\pi_1^\theta$  et par  $\pi_2^\theta$  les distributions marginales de  $\pi^\theta$  sur  $\mathbf{X}$  et  $\mathbf{Y}$ , respectivement. On note aussi par  $\mathbb{P}^\theta$  et par  $\tilde{\mathbb{P}}^\theta$  les distributions de probabilités définies respectivement comme suit :

- a)  $\mathbb{P}^\theta$  désigne l'extension de  $\mathbb{P}_{\pi^\theta}^\theta$  sur  $(\mathbf{X} \times \mathbf{Y})^{\mathbb{Z}}$ .
- b)  $\tilde{\mathbb{P}}^\theta$  est la projection correspondante sur la composante  $\mathbf{Y}^{\mathbb{Z}}$ .

On utilise également les symboles  $\mathbb{E}^\theta$  et  $\tilde{\mathbb{E}}^\theta$  pour désigner les espérances correspondantes aux  $\mathbb{P}^\theta$  et  $\tilde{\mathbb{P}}^\theta$ , respectivement. De plus, pour tout  $\theta, \theta' \in \Theta$ , on écrit  $\theta \sim \theta'$  si et seulement si  $\tilde{\mathbb{P}}^\theta = \tilde{\mathbb{P}}^{\theta'}$ . Cela définit une relation d'équivalence sur l'ensemble des paramètres  $\Theta$ , et la classe d'équivalence qui correspond à  $\theta$  est notée par  $[\theta] := \{\theta' \in \Theta : \theta \sim \theta'\}$ .

La relation d'équivalence  $\sim$  a été introduite par [Leroux \[1992\]](#) comme une alternative à la condition d'identifiabilité classique.

## 2.2 Identification de l'ensemble des maxima $\Theta_\star$

Outre l'hypothèse (K-1), notre résultat principal est indiqué dans l'hypothèse générale suivante.

(K-2) Pour tout  $\theta \neq \theta'$  dans  $\Theta$ , il existe un noyau de probabilité  $\Phi^{\theta,\theta'}$  sur  $\mathbf{Y}^{\mathbb{Z}-} \times \mathcal{X}$  tel que pour tout  $A \in \mathcal{X}$ ,

$$\frac{\int_{\mathbf{X}} \Phi^{\theta,\theta'}(Y_{-\infty;0}; dx_0) \kappa^\theta \langle Y_0, Y_1 \rangle(x_0; A)}{\int_{\mathbf{X}} \Phi^{\theta,\theta'}(Y_{-\infty;0}; dx_0) \kappa^\theta \langle Y_0, Y_1 \rangle(x_0; \mathbf{X})} = \Phi^{\theta,\theta'}(Y_{-\infty;1}; A), \quad \tilde{\mathbb{P}}^{\theta'}\text{-p.s.}$$

Donnons un aperçu de la formule qui apparaît en (K-2) et expliquons pourquoi il est si important de considérer les cas  $\theta = \theta'$  et  $\theta \neq \theta'$  séparément. Puisque  $\mathbf{X}$  est un espace borélien, le théorème 6.3 de [Kallenberg \[2002\]](#) s'applique et la distribution conditionnelle de  $X_0$  sachant  $Y_{-\infty;0}$  sous la loi  $\mathbb{P}^\theta$

définit un noyau de probabilité noté  $\Phi^\theta$ . En effet, le noyau  $\Phi^\theta$  satisfait, pour tout  $A \in \mathcal{X}$ ,

$$\frac{\int_{\mathbf{X}} \Phi^\theta(Y_{-\infty:0}; dx_0) \kappa^\theta \langle Y_0, Y_1 \rangle (x_0; A)}{\int_{\mathbf{X}} \Phi^\theta(Y_{-\infty:0}; dx_0) \kappa^\theta \langle Y_0, Y_1 \rangle (x_0; \mathbf{X})} = \Phi^\theta(Y_{-\infty:1}; A), \quad \tilde{\mathbb{P}}^\theta\text{-p.s.} \quad (8)$$

L'hypothèse (K-2) affirme que le noyau  $\Phi^{\theta, \theta'}$  satisfait une identité similaire  $\tilde{\mathbb{P}}^{\theta'}$ -p.s. pour  $\theta' \neq \theta$ .

Il convient de noter que le dénominateur du ratio affiché dans (K-2) peut être écrit  $p^{\theta, \theta'}(Y_1 | Y_{-\infty:0})$ , où, pour tout  $y \in \mathbf{Y}$  et  $y_{-\infty:0} \in \mathbf{Y}^{\mathbb{Z}^-}$ ,

$$p^{\theta, \theta'}(y | y_{-\infty:0}) := \int_{\mathbf{X}} \Phi^{\theta, \theta'}(y_{-\infty:0}; dx_0) \kappa^\theta \langle y_0, y \rangle (x_0; \mathbf{X}) \quad (9)$$

est une densité conditionnelle par rapport à la mesure  $\nu$ , car pour tout  $(x, y) \in \mathbf{X} \times \mathbf{Y}$ ,  $\int \kappa^\theta \langle y, y' \rangle (x; \mathbf{X}) \nu(dy') = 1$ .

Puisque  $\mathbf{Y}$  est un espace borélien, le théorème 6.3 de [Kallenberg \[2002\]](#) s'applique et la distribution conditionnelle de  $Y_{1:n}$  sachant  $Y_{-\infty:0}$  définit un noyau de probabilité. Et puisque  $\tilde{\mathbb{P}}^\theta(Y_{1:n} \in \cdot)$  est dominée par  $\nu^{\otimes n}$ , une densité conditionnelle est à son tour définie par rapport à  $\nu^{\otimes n}$ , qu'on note  $p_n^\theta(\cdot | \cdot)$ , de telle sorte que pour tout  $B \in \mathcal{Y}^{\otimes n}$ ,

$$\tilde{\mathbb{P}}^\theta(Y_{1:n} \in B | Y_{-\infty:0}) = \int_B p_n^\theta(y_{1:n} | Y_{-\infty:0}) \nu(dy_1) \cdots \nu(dy_n), \quad \tilde{\mathbb{P}}^\theta\text{-p.s.} \quad (10)$$

Le résultat principal de l'identifiabilité peut être énoncé comme suit.

**Théorème 2.3.** *Supposons que l'hypothèse (K-1) soit satisfaite et définissons  $\mathbb{P}^\theta$ ,  $\tilde{\mathbb{P}}^\theta$  et  $[\theta]$  suivant la définition 2.2. Supposons que l'hypothèse (K-2) soit également satisfaite. Pour tout  $\theta, \theta' \in \Theta$ , on définit  $p^{\theta, \theta'}(Y_1 | Y_{-\infty:0})$  par (9) si  $\theta \neq \theta'$ , et par  $p^{\theta, \theta}(Y_1 | Y_{-\infty:0}) = p_1^\theta(Y_1 | Y_{-\infty:0})$ , comme dans (10), sinon. Alors, pour tout  $\theta_\star \in \Theta$ , on a*

$$\operatorname{argmax}_{\theta \in \Theta} \tilde{\mathbb{E}}^{\theta_\star} [\ln p^{\theta, \theta_\star}(Y_1 | Y_{-\infty:0})] = [\theta_\star].$$

Ce résultat se fonde uniquement sur les hypothèses (K-1) et (K-2), et ne nécessite pas de taux vers l'équilibre. Cela est en réalité en profond contraste avec les résultats existants dans ces modèles, où l'identification de  $\Theta_\star$  est établie sous l'hypothèse de la séparation exponentielle des mesures (voir [Douc et al. \[2011\]](#)) ou sous l'hypothèse de l'ergodicité géométrique (voir [Douc et al. \[2004\]](#)).



### 2.3 Construction du noyau $\Phi^{\theta, \theta'}$ comme limite rétrogressive

Outre l'hypothèse (K-1), le théorème 2.3 dépend fondamentalement de l'hypothèse (K-2). Ces hypothèses assurent l'existence du noyau de probabilité  $\Phi^{\theta, \theta'}$  qui implique la définition de  $p_1^{\theta, \theta'}(\cdot|\cdot)$ . On explique maintenant comment le noyau  $\Phi^{\theta, \theta'}$  est obtenu comme limite des noyaux explicites dérivés de  $K^\theta$  sous  $\mathbb{P}^{\theta'}$ . Cela s'appliquera généralement aux ODMs, traités dans la section 2.5, mais aussi dans le cas plus classique des HMMs, comme expliqué dans la section 2.4. Une approche habituelle est de définir le noyau  $\Phi^{\theta, \theta'}$  comme la limite faible des noyaux suivants.

**Définition 2.4.** Soit  $n \in \mathbb{Z}_+^*$ . Pour tout  $\theta \in \Theta$  et  $x \in \mathsf{X}$ , on définit le noyau de probabilité  $\Phi_{x,n}^\theta$  sur  $\mathsf{Y}^{n+1} \times \mathcal{X}$  par, pour tout  $y_{0:n} \in \mathsf{Y}^{n+1}$  et  $A \in \mathcal{X}$ ,

$$\Phi_{x,n}^\theta(y_{0:n}; A) := \frac{\int_{\mathsf{X}^{n-1} \times A} \prod_{k=0}^{n-1} \kappa^\theta \langle y_k, y_{k+1} \rangle (x_k; dx_{k+1})}{\int_{\mathsf{X}^n} \prod_{k=0}^{n-1} \kappa^\theta \langle y_k, y_{k+1} \rangle (x_k; dx_{k+1})} \quad \text{avec } x_0 = x.$$

Il est important de noter que  $\Phi_{x,n}^\theta(Y_{0:n}; \cdot)$  est la distribution conditionnelle de  $X_n$  sachant  $Y_{1:n}$  sous  $\mathbb{P}_{(x, Y_0)}^\theta$ . Pour obtenir le noyau  $\Phi^{\theta, \theta'}$  voulu, on prend, pour une valeur de  $x$  (choisie de manière adéquate), la limite de  $\Phi_{x,n}^\theta(y_{0:n}; \cdot)$  quand  $n \rightarrow \infty$  et pour une séquence  $y_{0:n}$  correspondant à une trajectoire sous  $\tilde{\mathbb{P}}^{\theta'}$ . L'affirmation précise est fournie dans l'hypothèse (K-3) suivante, qui nécessite la définition ci-dessous. Pour tout  $\theta \in \Theta$  et pour toute fonction mesurable  $f \geq 0$  définie sur  $\mathsf{X}$ , on fixe

$$\mathcal{F}_f^\theta := \{x \mapsto \kappa^\theta \langle y, y' \rangle (x; f) : (y, y') \in \mathsf{Y}^2\}.$$

On peut maintenant énoncer l'hypothèse comme suit :

(K-3) Pour tout  $\theta \neq \theta' \in \Theta$ , il existe  $x \in \mathsf{X}$ , un noyau de probabilité  $\Phi^{\theta, \theta'}$  sur  $\mathsf{Y}^{\mathbb{Z}^-} \times \mathcal{X}$  et une classe dénombrable  $\mathcal{F}$  des fonctions  $\mathsf{X} \rightarrow \mathbb{R}_+$  mesurables, tels que pour tout  $f \in \mathcal{F}$ ,

$$\tilde{\mathbb{P}}^{\theta'} \left( \forall f' \in \mathcal{F}_f^\theta \cup \{f\}, \lim_{m \rightarrow \infty} \Phi_{x,m}^\theta(Y_{-m:0}; f') = \Phi^{\theta, \theta'}(Y_{-\infty:0}; f') < \infty \right) = 1.$$

Le lemme suivant montre que, à condition que  $\mathcal{F}$  soit suffisamment riche, l'hypothèse (K-3) peut être utilisée directement pour obtenir l'hypothèse (K-2). On dit qu'une classe de fonctions  $\mathsf{X} \rightarrow \mathbb{R}$  se sépare si, pour des mesures de probabilité  $\mu_1$  et  $\mu_2$  sur  $(\mathsf{X}, \mathcal{X})$ , l'égalité entre  $\mu_1(f)$  et  $\mu_2(f)$  sur  $f$  dans la classe implique l'égalité des deux mesures.

**Lemme 2.5.** *Supposons que l'hypothèse (K-3) se tient et que  $\mathcal{F}$  est une classe de fonctions qui sépare des mesures et contient  $\mathbb{1}_X$ . Alors, le noyau  $\Phi^{\theta, \theta'}$  satisfait l'hypothèse (K-2).*

## 2.4 Application aux HMM

### 2.4.1 Définitions et hypothèses

Les modèles de Markov cachés appartiennent à une sous-classe des modèles de Markov partiellement observés et sont définis comme suit :

**Définition 2.6.** On considère un modèle de Markov partiellement observé et partiellement dominé, comme donné dans la définition 2.1, avec les noyaux de Markov  $\{K^\theta : \theta \in \Theta\}$ . On dira que ce modèle est un modèle de Markov caché si le noyau  $K^\theta$  satisfait

$$K^\theta((x, y); dx' dy') = Q^\theta(x; dx')G^\theta(x'; dy'). \quad (11)$$

De plus, dans ce contexte, on suppose toujours que  $(X, d_X)$  est un espace métrique séparable complet et on note par  $\mathcal{X}$  la  $\sigma$ -algèbre borélienne associée.

Dans (11),  $Q^\theta$  et  $G^\theta$  sont des noyaux de transition sur  $X \times \mathcal{X}$  et  $X \times \mathcal{Y}$ , respectivement. Puisque le modèle est partiellement dominé, on note par  $g^\theta$  la dérivée de Radon–Nikodym de  $G^\theta(x; \cdot)$  correspondante par rapport à la mesure dominante  $\nu$  : pour tout  $(x, y) \in X \times Y$ ,  $\frac{dG^\theta(x; \cdot)}{d\nu}(y) = g^\theta(x; y)$ . On peut observer directement que le noyau non-normalisé  $\kappa^\theta \langle y, y' \rangle$  défini dans (4) ne dépend plus de  $y$ , et dans ce cas, on peut écrire :

$$\kappa^\theta \langle y, y' \rangle(x; dx') = \kappa^\theta \langle y' \rangle(x; dx') = Q^\theta(x; dx')g^\theta(x'; y').$$

La dynamique du processus de Markov caché peut être illustrée par la figure 2.2.

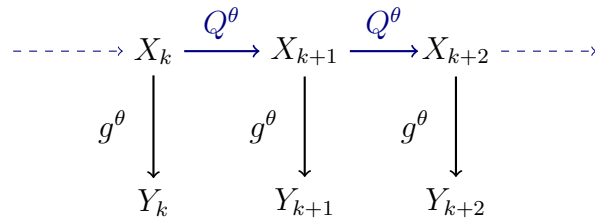


FIGURE 2.2 – Représentation graphique du modèle de Markov caché.

Pour chaque entier  $n \geq 1$ ,  $\theta \in \Theta$  et chaque séquence  $y_{0:n-1} \in \mathcal{Y}^n$ , considérons le noyau non-normalisé  $\mathbf{L}^\theta \langle y_{0:n-1} \rangle$  sur  $\mathbf{X} \times \mathcal{X}$  défini par, pour tout  $x_0 \in \mathbf{X}$  et  $A \in \mathcal{X}$ ,

$$\mathbf{L}^\theta \langle y_{0:n-1} \rangle(x_0; A) = \int \cdots \int \left[ \prod_{k=0}^{n-1} g^\theta(x_k; y_k) Q^\theta(x_k; dx_{k+1}) \right] \mathbb{1}_A(x_n).$$

Ainsi, l'EMV  $\hat{\theta}_{\xi, n}$ , associé aux observations  $Y_{0:n-1}$  avec une distribution initiale arbitraire  $\xi$  sur  $\mathbf{X}$ , est défini par

$$\hat{\theta}_{\xi, n} \in \operatorname{argmax}_{\theta \in \Theta} \xi \mathbf{L}^\theta \langle Y_{0:n-1} \rangle \mathbb{1}_{\mathbf{X}}.$$

Ici, on démontre l'application du résultat principal (le théorème 2.3) dans le contexte des HMM en considérant les hypothèses de [Douc and Moulines \[2012\]](#) dans le cas particulier de blocs de taille 1 ( $r = 1$ ). Mais avant de lister les hypothèses principales, on rappelle d'abord la définition d'un ensemble de Doeblin local (*local Doeblin set*) comme introduit dans [\[Douc and Moulines, 2012, Définition 1\]](#).

**Définition 2.7.** Un ensemble  $C$  est dit de Doeblin local par rapport à la famille des noyaux  $\{Q^\theta : \theta \in \Theta\}$  s'il existe des constantes positives  $\epsilon_C^-$ ,  $\epsilon_C^+$ , et une famille de mesures de probabilité  $\{\lambda_C^\theta : \theta \in \Theta\}$ , telles que, pour tout  $\theta \in \Theta$ ,  $\lambda_C^\theta(C) = 1$ , et pour tout  $A \in \mathcal{X}$  et  $x \in C$ ,

$$\epsilon_C^- \lambda_C^\theta(A) \leq Q^\theta(x; A \cap C) \leq \epsilon_C^+ \lambda_C^\theta(A).$$

On considère maintenant l'ensemble des hypothèses suivantes.

(D-1) Il existe une mesure  $\sigma$ -finie  $\mu$  sur  $(\mathbf{X}, \mathcal{X})$  qui domine  $Q^\theta(x; \cdot)$  pour tout  $(x, \theta) \in \mathbf{X} \times \Theta$ . De plus, en notant  $q^\theta(x; x') := \frac{dQ^\theta(x; \cdot)}{d\mu}(x')$ , on a

$$q^\theta(x; x') > 0, \quad \text{pour tout } (x, x', \theta) \in \mathbf{X} \times \mathbf{X} \times \Theta.$$

(D-2) Pour tout  $y \in \mathcal{Y}$ , on a  $\sup_{\theta \in \Theta} \sup_{x \in \mathbf{X}} g^\theta(x; y) < \infty$ .

(D-3) (a) Pour tout  $\theta_\star \in \Theta$ , il existe un ensemble  $K \in \mathcal{Y}$  avec  $\tilde{\mathbb{P}}^{\theta_\star}(Y_0 \in K) > 2/3$  tel que pour tout  $\eta > 0$ , il existe un ensemble de Doeblin local  $C \in \mathcal{X}$  par rapport à  $\{Q^\theta : \theta \in \Theta\}$  satisfaisant, pour tout  $\theta \in \Theta$  et tout  $y \in K$ ,

$$\sup_{x \in C^c} g^\theta(x; y) \leq \eta \sup_{x \in \mathbf{X}} g^\theta(x; y) < \infty.$$

(b) Pour tout  $\theta_\star \in \Theta$ , il existe un ensemble  $D \in \mathcal{X}$  satisfaisant

$$\inf_{\theta \in \Theta} \inf_{x \in D} Q^\theta(x; D) > 0 \quad \text{et} \quad \tilde{\mathbb{E}}^{\theta_\star} \left[ \ln^- \inf_{\theta \in \Theta} \inf_{x \in D} g^\theta(x; Y_0) \right] < \infty.$$

(D-4) Pour tout  $\theta_\star \in \Theta$ ,  $\tilde{\mathbb{E}}^{\theta_\star} \left[ \ln^+ \sup_{\theta \in \Theta} \sup_{x \in \mathbb{X}} g^\theta(x; Y_0) \right] < \infty$ .

(D-5) Il existe  $p \in \mathbb{Z}_+$  tel que pour chaque  $x \in \mathbb{X}$  et  $n \geq p$ , la fonction  $\theta \mapsto \mathbf{L}^\theta \langle Y_{0:n} \rangle(x; \mathbb{X})$  est  $\tilde{\mathbb{P}}^{\theta_\star}$ -p.s. continue sur  $\Theta$ .

En comparaison, les hypothèses (D-3), (D-4) et (D-5) et (3) dans la définition 2.1 correspondent aux hypothèses (A1), (A2) et (A3) dans Douc and Moulines [2012], où les blocs sont de taille  $r = 1$ .

Dans ces modèles, le noyau  $\Phi_{x,n}^\theta$  introduit dans la définition 2.4 s'écrit :

$$\Phi_{x,n}^\theta(y_{0:n}; A) = \frac{\int_{\mathbb{X}^{n-1} \times A} \prod_{k=0}^{n-1} Q^\theta(x_k; dx_{k+1}) g^\theta(x_{k+1}; y_{k+1})}{\int_{\mathbb{X}^n} \prod_{k=0}^{n-1} Q^\theta(x_k; dx_{k+1}) g^\theta(x_{k+1}; y_{k+1})} \quad \text{avec } x_0 = x.$$

### 2.4.2 Consistance de classe d'équivalence

On peut maintenant énoncer le résultat principal concernant la consistance de classe d'équivalence de l'EMV pour les HMM.

**Théorème 2.8.** *Supposons que (K-1) se tient et définissons  $\mathbb{P}^\theta$ ,  $\tilde{\mathbb{P}}^\theta$  et la classe d'équivalence  $[\theta]$  comme dans la définition 2.2. Par ailleurs, supposons que  $(\Theta, \Delta)$  est un espace métrique compact et que les hypothèses (D-1)–(D-5) se tiennent. Alors, pour toute mesure de probabilité  $\xi$ ,*

$$\lim_{n \rightarrow \infty} \Delta(\hat{\theta}_{\xi,n}, [\theta_\star]) = 0, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-p.s.}$$

### 2.4.3 Un exemple polynomialement ergodique

Comme application du théorème 2.8, on considère le modèle HMM suivant. On pose  $\mathbb{X} = \mathbb{R}_+$ ,  $\mathcal{X} = \mathcal{B}(\mathbb{R}_+)$ ,  $\mathbb{Y} = \mathbb{R}$  et  $\mathcal{Y} = \mathcal{B}(\mathbb{R})$  et on définit un HMM sur  $\mathbb{X} \times \mathbb{Y}$  par les récursions suivantes :

$$\begin{aligned} X_k &= (X_{k-1} + U_k - m)^+, \\ Y_k &= aX_k + V_k, \end{aligned} \tag{12}$$

où  $(m, a) \in \mathbb{R}_+^* \times \mathbb{R}$ , et la séquence  $\{(U_k, V_k) : k \in \mathbb{Z}_+\}$  est indépendante et identiquement distribuée (i.i.d.) et est indépendante de  $X_0$ . Ce modèle de Markov  $\{X_k : k \in \mathbb{Z}_+\}$  a été proposé par [Tuominen and Tweedie \[1994\]](#) et considéré de plus par [Järner and Roberts \[2002\]](#) comme un exemple de chaîne de Markov polynomialement ergodique, dans des hypothèses spécifiques faites sur  $U_k$ 's. C'est-à-dire que, si  $U_k$ 's sont centrées et  $\mathbb{E}[e^{\lambda U_k^+}] = \infty$  pour chaque  $\lambda > 0$ , on peut montrer que la chaîne  $\{X_k : k \in \mathbb{Z}_+\}$  n'est pas géométriquement ergodique. Dans une telle situation, la condition de séparation exponentielle de mesures introduite dans [Douc et al. \[2011\]](#) semble difficile à vérifier. Pour établir la consistance de classe d'équivalence de l'EMV de ce modèle, on suppose de plus que  $U_0$  et  $V_0$  sont indépendantes et centrées, et que les deux admettent des densités par rapport à la mesure de Lebesgue  $\lambda$  sur  $\mathbb{R}$ , notées  $r$  et  $h$ , respectivement. On suppose également que

(E-1) on a :

- (a) la densité  $r$  est continue et positive sur  $\mathbb{R}$  ;
- (b) il existe une constante  $\alpha > 2$  telle que  $r(u)|u|^{\alpha+1}$  est strictement bornée de  $\infty$  quand  $|u| \rightarrow \infty$  et de 0 quand  $u \rightarrow \infty$  ;

(E-2) on a :

- (a) la densité  $h$  est continue et positive sur  $\mathbb{R}$ , et  $\lim_{|v| \rightarrow \infty} h(v) = 0$  ;
- (b) il existe une constante  $\beta \in [1, \alpha - 1)$  (avec  $\alpha$  donnée par (E-1)) et  $b, c > 0$  satisfaisant  $\mathbb{E}(|V_0|^\beta) < \infty$  et  $h(v) \geq be^{-c|v|^\beta}$  pour tout  $v \in \mathbb{R}$ .

Par exemple, une distribution de Pareto symétrique avec un paramètre strictement plus large que 2 satisfait (E-1), et à condition que  $\alpha > 3$ , (E-2) se tient avec une distribution gaussienne centrée. Le modèle est paramétré par  $\theta = (m, a) \in \Theta := [\underline{m}, \overline{m}] \times [\underline{a}, \overline{a}]$  où  $0 < \underline{m} < \overline{m}$  et  $\underline{a} < \overline{a}$ . Dans ce modèle, la transition markovienne  $Q^\theta$  de  $\{X_k : k \in \mathbb{Z}_+\}$  admet une densité de transition  $q^\theta$  par rapport à la mesure dominante  $\mu(dx) = \lambda(dx) + \delta_0(dx)$ , que l'on peut écrire ainsi : pour tout  $(x, x') \in \mathbb{R}_+^2$ ,

$$q^\theta(x; x') = r(x' - x + m) \mathbb{1}\{x' > 0\} + \left( \int_{-\infty}^{m-x} r(u) du \right) \mathbb{1}\{x' = 0\}.$$

Par ailleurs, (12) implique que

$$g^\theta(x; y) = h(y - ax).$$

On a maintenant le résultat.

**Proposition 2.9.** *Considérons le HMM défini comme dans (12) dans les hypothèses (E-1) et (E-2). Alors (K-1) se tient, et on définit  $\mathbb{P}^\theta$ ,  $\tilde{\mathbb{P}}^\theta$  et la classe d'équivalence  $[\theta]$  comme dans la définition 2.2. De plus, pour chaque mesure de probabilité  $\xi$ , l'EMV  $\hat{\theta}_{\xi,n}$  est consistant de classe d'équivalence, c'est-à-dire que pour tout  $\theta_* \in \Theta$ ,*

$$\lim_{n \rightarrow \infty} \Delta(\hat{\theta}_{\xi,n}, [\theta_*]) = 0, \quad \tilde{\mathbb{P}}^{\theta_*}\text{-p.s.}$$

## 2.5 Application aux ODM

### 2.5.1 Définitions et notation

Les ODM forment une sous-classe des modèles de Markov partiellement dominés et partiellement observés. Ils sont définis formellement comme suit :

**Définition 2.10.** Considérons un modèle de Markov partiellement observé et partiellement dominé comme donné dans la définition 2.1 avec pour noyaux de Markov  $\{K^\theta : \theta \in \Theta\}$ . On dit que ce modèle est un modèle observation-driven si le noyau  $K^\theta$  satisfait

$$K^\theta((x, y); dx' dy') = \delta_{\psi_y^\theta(x)}(dx') G^\theta(x'; dy'), \quad (13)$$

où  $\delta_a$  note la masse de Dirac au point  $a$ ,  $G^\theta$  est un noyau de probabilité sur  $\mathbf{X} \times \mathcal{Y}$  et  $\{(x, y) \mapsto \psi_y^\theta(x) : \theta \in \Theta\}$  est une famille de fonctions mesurables de  $(\mathbf{X} \times \mathbf{Y}, \mathcal{X} \otimes \mathcal{Y})$  à  $(\mathbf{X}, \mathcal{X})$ . On remarque que le processus  $\{X_k : k \geq 1\}$  en lui-même est une chaîne de Markov avec un noyau de transition défini par

$$R^\theta(x; A) = \int \mathbb{1}_A(\psi_y^\theta(x)) G^\theta(x; dy), \quad x \in \mathbf{X}, A \in \mathcal{X}. \quad (14)$$

Dans ce contexte, on suppose toujours que  $(\mathbf{X}, d_X)$  est un espace métrique séparable complet et  $\mathcal{X}$  représente la  $\sigma$ -algèbre borélienne associée.

Notons qu'une chaîne de Markov  $\{(X_k, Y_k) : k \in \mathbb{Z}_+\}$  avec le noyau de probabilité donné par (13) peut être définie de manière équivalente par les récursions suivantes :

$$\begin{aligned} X_{k+1} &= \psi_{Y_k}^\theta(X_k), \\ Y_{k+1} | X_{0:k+1}, Y_{0:k} &\sim G^\theta(X_{k+1}; \cdot), \end{aligned} \quad (15)$$

et peut être graphiquement représentée comme dans la figure 2.3.

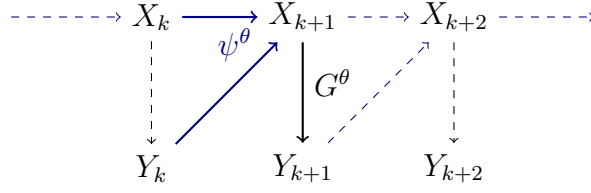


FIGURE 2.3 – Représentation graphique des ODM.

L'exemple le plus connu est le processus GARCH(1, 1), où  $G^\theta(x; \cdot)$  est une distribution centrée (dite gaussienne) de variance  $x$  et où  $\psi_y^\theta(x)$  est une fonction affine de  $x$  et de  $y^2$ .

Dans le cas spécifique de la définition 2.1, pour tout  $x \in \mathsf{X}$ , le noyau de probabilité  $G^\theta(x; \cdot)$  est dominé par une mesure  $\sigma$ -finie  $\nu$  sur  $(\mathsf{Y}, \mathcal{Y})$ , et on note  $g^\theta(x; \cdot)$  sa dérivée de Radon–Nikodym,  $g^\theta(x; y) = \frac{dG^\theta(x; \cdot)}{d\nu}(y)$ . Un ODM paramétré dominé est ainsi défini par la collection  $\{(g^\theta, \psi^\theta) : \theta \in \Theta\}$ . Par ailleurs, (3) peut être ré-écrite dans ce cas : pour tout  $(x, y) \in \mathsf{X} \times \mathsf{Y}$  et pour tout  $\theta \in \Theta$ ,

$$g^\theta(x; y) > 0.$$

Dans (K-1), on suppose que le modèle est bien spécifié, c'est-à-dire que l'échantillon d'observations  $(Y_1, \dots, Y_n)$  est distribué selon  $\tilde{\mathbb{P}}^{\theta_\star}$  pour un paramètre  $\theta_\star$  inconnu. L'inférence de  $\theta_\star$  est basée sur la vraisemblance conditionnelle de  $(Y_1, \dots, Y_n)$  sachant  $X_1 = x$  pour une valeur arbitraire  $x \in \mathsf{X}$ . La fonction de densité correspondante par rapport à la mesure  $\nu^{\otimes n}$  est, sous le paramètre  $\theta$ ,

$$y_{1:n} \mapsto \prod_{k=1}^n g^\theta(\psi^\theta\langle y_{1:k-1} \rangle(x); y_k),$$

où, pour tout vecteur  $y_{1:p} = (y_1, \dots, y_p) \in \mathsf{Y}^p$ ,  $\psi^\theta\langle y_{1:p} \rangle$  est la fonction  $\mathsf{X} \rightarrow \mathsf{X}$  définie comme la composition successive de  $\psi_{y_1}^\theta, \psi_{y_2}^\theta, \dots$ , et  $\psi_{y_p}^\theta$ ,

$$\psi^\theta\langle y_{1:p} \rangle = \psi_{y_p}^\theta \circ \psi_{y_{p-1}}^\theta \circ \dots \circ \psi_{y_1}^\theta, \quad (16)$$

avec la convention que  $\psi^\theta\langle y_{s:t} \rangle(x) = x$  pour  $s > t$ . Alors, l'EMV (conditionnel)  $\hat{\theta}_{x,n}$  du paramètre  $\theta$  correspondant est défini par

$$\hat{\theta}_{x,n} \in \operatorname{argmax}_{\theta \in \Theta} \mathbb{L}_{x,n}^\theta\langle Y_{1:n} \rangle, \quad (17)$$

où

$$\mathbb{L}_{x,n}^\theta\langle y_{1:n} \rangle := n^{-1} \ln \left( \prod_{k=1}^n g^\theta(\psi^\theta\langle y_{1:k-1} \rangle(x); y_k) \right). \quad (18)$$

On fournira des conditions simples pour obtenir la consistance de  $\hat{\theta}_{x,n}$  dans le sens que, avec une probabilité tendant vers un, pour une valeur de  $x$  bien choisie,  $\hat{\theta}_{x,n}$  appartient à un voisinage de la classe d'équivalence  $[\theta_\star]$  de  $\theta_\star$ , comme donnée dans la définition 2.2.

### 2.5.2 Identifiabilité

Considérons les hypothèses suivantes.

(C-1) Pour tout  $\theta \neq \theta_\star \in \Theta$ , il existe une valeur  $x \in X$  et une fonction mesurable  $\psi^{\theta, \theta_\star} \langle \cdot \rangle$  définie sur  $Y^{\mathbb{Z}^-}$  telles que

$$\lim_{m \rightarrow \infty} \psi^\theta \langle Y_{-m:0} \rangle (x) = \psi^{\theta, \theta_\star} \langle Y_{-\infty:0} \rangle, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-p.s.} \quad (19)$$

(C-2) Pour tout  $\theta \in \Theta$  et  $y \in Y$ , la fonction  $x \mapsto g^\theta(x; y)$  est continue sur  $X$ .

(C-3) Pour tout  $\theta \in \Theta$  et  $y \in Y$ , la fonction  $x \mapsto \psi_y^\theta(x)$  est continue sur  $X$ .

Dans les ODM, le noyau  $\kappa^\theta$  défini dans (4) s'écrit

$$\kappa^\theta \langle y, y' \rangle (x; dx') = g^\theta(x'; y') \delta_{\psi_y^\theta(x)}(dx') = g^\theta(\psi_y^\theta(x); y') \delta_{\psi_y^\theta(x)}(dx'),$$

et le noyau de probabilité  $\Phi_{x,n}^\theta$  dans la définition 2.4 se lit, pour tout  $x \in X$  et  $y_{0:n} \in Y^{n+1}$ ,

$$\Phi_{x,n}^\theta(y_{0:n}; \cdot) = \delta_{\psi^\theta \langle y_{0:n-1} \rangle (x)(\cdot)}$$

(la masse de Dirac au point  $\psi^\theta \langle y_{0:n-1} \rangle (x)$ ). En utilisant ces expressions, on obtient le résultat suivant qui est un cas spécial du théorème 2.3.

**Théorème 2.11.** *On suppose que (K-1) se tient dans le cadre de l'ODM et on définit  $\mathbb{P}^\theta$ ,  $\tilde{\mathbb{P}}^\theta$  et  $[\theta]$  comme dans la définition 2.2. On suppose que les hypothèses (C-1), (C-2) et (C-3) se tiennent, et on définit  $p^{\theta, \theta_\star}(\cdot | \cdot)$  en mettant, pour  $y_{-\infty:0} \in Y^{\mathbb{Z}^-}$   $\tilde{\mathbb{P}}^{\theta_\star}$  presque partout,*

$$p^{\theta, \theta_\star}(y_1 | y_{-\infty:0}) = \begin{cases} g^\theta(\psi^{\theta, \theta_\star} \langle y_{-\infty:0} \rangle; y_1) & \text{si } \theta \neq \theta_\star, \\ p_1^\theta(y_1 | y_{-\infty:0}) \text{ comme définie dans (10),} & \text{sinon.} \end{cases} \quad (20)$$

Alors, pour tout  $\theta_\star \in \Theta$ , on a

$$\operatorname{argmax}_{\theta \in \Theta} \tilde{\mathbb{E}}^{\theta_\star} [\ln p^{\theta, \theta_\star}(Y_1 | Y_{-\infty:0})] = [\theta_\star]. \quad (21)$$



### 3 Consistance de l'EMV pour la classe des PMM entièrement dominés

Dans cette section, nous examinons la consistance de l'EMV pour la classe des PMM entièrement dominés. En suivant l'approche de [Douc and Moulines \[2012\]](#) pour dériver la convergence de l'EMV à l'ensemble des minima du taux d'entropie relative dans des HMM mal spécifiés, dans des conditions suffisantes assez faibles, la convergence de l'EMV pour cette classe est établie. De plus, quand le modèle est bien spécifié, en suivant [Douc et al. \[2014\]](#), dans ces mêmes conditions, l'identifiabilité est également obtenue, établissant la consistance de classe d'équivalence de l'EMV. Ce résultat est obtenu comme produit de la convergence de l'EMV par blocs, que nous décrivons ci-dessous.

#### 3.1 Consistance de l'EMV par blocs

##### 3.1.1 Définitions et notation

Soient  $(\mathsf{X}, \mathcal{X})$  et  $(\mathsf{Z}, \mathcal{Z})$  deux espaces boréliens et  $\Theta$  un espace des paramètres, qui est toujours supposé être un espace métrique compact muni d'une métrique  $\Delta$ . Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace des probabilités, et  $\{Z_k : k \in \mathbb{Z}\}$  un processus stochastique prenant des valeurs sur  $(\mathsf{Z}, \mathcal{Z})$ , et dont on suppose qu'il est observé. Dans cette section, nous suppose l'hypothèse suivante.

(L-1) Le processus stochastique  $\{Z_k : k \in \mathbb{Z}\}$  est stationnaire et ergodique sur  $(\mathsf{Z}, \mathcal{Z})$ .

On note  $\mathbb{P}_{\mathsf{Z}}$  l'image de la probabilité  $\mathbb{P}$  induite par le processus stationnaire  $\{Z_k : k \in \mathbb{Z}\}$  sur l'espace produit  $(\mathsf{Z}^{\mathbb{Z}}, \mathcal{Z}^{\otimes \mathbb{Z}})$  et on note  $\mathbb{E}_{\mathsf{Z}}$  l'espérance associée. Soit  $\{\mathbf{L}^{\theta}\langle z \rangle : (\theta, z) \in \Theta \times \mathsf{Z}\}$  une famille de noyaux de transition sur  $(\mathsf{X}, \mathcal{X})$  telle que pour tout  $\theta \in \Theta$ , toute mesure de probabilité  $\chi$  sur  $(\mathsf{X}, \mathcal{X})$  et toute fonction mesurable positive  $f$  sur  $\mathsf{X}$ , la fonction  $z \mapsto \chi \mathbf{L}^{\theta}\langle z \rangle f$  est mesurable sur  $(\mathsf{Z}, \mathcal{Z})$ . Ici, on remarque que cette famille de noyaux de transition peut être ou non liée au processus stochastique observé  $\{Z_k : k \in \mathbb{Z}\}$ . On définit maintenant ce qui suit.

**Définition 3.1** (Noyau de vraisemblance). Soient  $\theta \in \Theta$ ,  $n \in \mathbb{Z}_+^*$  et  $z_{0:n-1} \in \mathsf{Z}^n$ . Le *noyau de vraisemblance (paramétrique)* par rapport à la famille  $\{\mathbf{L}^{\theta}\langle z \rangle : (\theta, z) \in \Theta \times \mathsf{Z}\}$  de l'échantillon  $z_{0:n-1}$ , noté  $\mathbf{L}_n^{\theta}\langle z_{0:n-1} \rangle$ , est défini comme l'application successive de noyaux de transition  $\mathbf{L}^{\theta}\langle z_0 \rangle, \dots, \mathbf{L}^{\theta}\langle z_{n-1} \rangle$  de la famille  $\{\mathbf{L}^{\theta}\langle z \rangle : (\theta, z) \in \Theta \times \mathsf{Z}\}$ . Formellement, pour tout  $(x_0, A) \in \mathsf{X} \times \mathcal{X}$ ,

$$\mathbf{L}_n^{\theta}\langle z_{0:n-1} \rangle(x_0; A) := \int \mathbf{L}^{\theta}\langle z_0 \rangle(x_0; dx_1) \cdots \mathbf{L}^{\theta}\langle z_{n-1} \rangle(x_{n-1}; dx_n) \mathbb{1}_A(x_n). \quad (22)$$

Par définition,  $\mathbf{L}_n^\theta \langle z_{0:n-1} \rangle$  définit également un noyau de transition sur  $(\mathbf{X}, \mathcal{X})$ . Par ailleurs, pour tout  $\theta \in \Theta$ ,  $n, p \in \mathbb{Z}_+^*$  et  $z_{0:n+p-1} \in \mathbf{Z}^{n+p}$ , l'identité :

$$\mathbf{L}_{n+p}^\theta \langle z_{0:n+p-1} \rangle = \mathbf{L}_n^\theta \langle z_{0:n-1} \rangle \mathbf{L}_p^\theta \langle z_{n:n+p-1} \rangle$$

se tient. Dans ce qui suit, on notera toujours  $\mathcal{P}(\mathbf{X}, \mathcal{X})$  l'ensemble des mesures de probabilité sur  $(\mathbf{X}, \mathcal{X})$ . Pour des commodités de notation, l'indice  $n$  sera omis de  $\mathbf{L}_n^\theta$  qui est été défini dans (22). Ensuite, On suppose toujours que pour tout  $\theta \in \Theta$ ,  $n \in \mathbb{Z}_+^*$ ,  $z_{0:n-1} \in \mathbf{Z}^n$  et toute mesure de probabilité  $\chi \in \mathcal{P}(\mathbf{X}, \mathcal{X})$ ,

$$0 < \chi \mathbf{L}^\theta \langle z_{0:n-1} \rangle \mathbb{1}_{\mathbf{X}} < \infty. \quad (23)$$

La définition suivante définit la version conditionnelle du noyau de vraisemblance.

**Définition 3.2.** Soit  $\theta \in \Theta$ . Pour tout  $m, n \in \mathbb{Z}_+$  avec  $m < n$ ,  $z_{m:n} \in \mathbf{Z}^{n-m+1}$ ,  $\chi \in \mathcal{P}(\mathbf{X}, \mathcal{X})$  et toute fonction mesurable  $f$  de  $(\mathbf{X}, \mathcal{X})$  à  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , on définit

$$\chi \mathbf{L}^\theta \langle z_{p:n} | z_{m:p-1} \rangle f := \frac{\chi \mathbf{L}^\theta \langle z_{m:n} \rangle f}{\chi \mathbf{L}^\theta \langle z_{m:p-1} \rangle \mathbb{1}_{\mathbf{X}}}, \quad m < p \leq n. \quad (24)$$

**Définition 3.3.** Soient  $n \in \mathbb{Z}_+^*$ , la séquence  $Z_{0:n-1}$  des observations réalisées du processus stochastique  $\{Z_k : k \in \mathbb{Z}\}$  et  $x_0 \in \mathbf{X}$ . La fonction  $\theta \mapsto \mathbf{L}_n^\theta \langle Z_{0:n-1} \rangle (x_0; \mathbf{X})$  est appelée la *fonction de L-vraisemblance (conditionnelle)* par rapport à la famille  $\{\mathbf{L}^\theta \langle z \rangle : (\theta, z) \in \Theta \times \mathbf{Z}\}$  des observations  $Z_{0:n-1}$  sachant  $X_0 = x_0$ . On définit l'*estimateur du maximum de L-vraisemblance* correspondant par rapport à la famille  $\{\mathbf{L}^\theta \langle z \rangle : (\theta, z) \in \Theta \times \mathbf{Z}\}$ , associé à une distribution de probabilité initiale arbitraire  $\chi \in \mathcal{P}(\mathbf{X}, \mathcal{X})$  par

$$\hat{\theta}_{\chi, n}^L \in \operatorname{argmax}_{\theta \in \Theta} \chi \mathbf{L}^\theta \langle Z_{0:n-1} \rangle \mathbb{1}_{\mathbf{X}}. \quad (25)$$

Quand  $\chi$  est  $\delta_x$ , la masse de Dirac au point  $x \in \mathbf{X}$ , l'estimateur du maximum de  $L$ -vraisemblance défini dans (25) sera simplement noté  $\hat{\theta}_{x, n}^L$ .

Dans cette présente section, nous voulons établir que sous certaines conditions, il existe un sous-ensemble non-vidé  $\Theta_\star^L \subseteq \Theta$  tel que  $\Delta(\hat{\theta}_{\chi, n}^L, \Theta_\star^L)$ , la distance entre  $\hat{\theta}_{\chi, n}^L$  et  $\Theta_\star^L$ , converge  $\mathbb{P}_Z$ -p.s. vers zéro quand  $n \rightarrow \infty$ . La preuve de ce résultat suit des arguments similaires à ceux de [Douc and Moulines \[2012\]](#). La première étape consiste à établir la limite de la  $L$ -vraisemblance, c'est-à-dire à montrer qu'il existe une fonction  $\mathbf{L}(\theta)$  en  $\theta \in \Theta$  telle que pour tout  $\theta \in \Theta$ ,

$$\lim_{n \rightarrow \infty} n^{-1} \ln \chi \mathbf{L}^\theta \langle Z_{0:n-1} \rangle \mathbb{1}_{\mathbf{X}} = \mathbf{L}(\theta), \quad \mathbb{P}_Z\text{-p.s.} \quad (26)$$

La seconde étape consiste à prouver que la valeur qui maximise  $\theta \mapsto \chi \mathbf{L}^\theta \langle Z_{0:n-1} \rangle \mathbb{1}_X$  converge  $\mathbb{P}_Z$ -p.s. vers l'ensemble des maxima  $\Theta_\star^L = \operatorname{argmax}_{\theta \in \Theta} \mathbf{L}(\theta)$ . On décrit maintenant brièvement la manière dont ces deux étapes essentielles est établie. Lors de la formulation de la première étape, on note tout d'abord que de la relation (24) de la définition 3.2, on peut écrire

$$n^{-1} \ln \chi \mathbf{L}^\theta \langle Z_{0:n-1} \rangle \mathbb{1}_X = n^{-1} \sum_{k=0}^{n-1} \ln \chi \mathbf{L}^\theta \langle Z_k | Z_{0:k-1} \rangle \mathbb{1}_X.$$

Ensuite, il est nécessaire de montrer qu'il existe une fonction  $\mathbb{P}_Z$ -intégrable  $\pi_Z^\theta : \mathbb{Z}^- \rightarrow \mathbb{R}$  telle que  $|\chi \mathbf{L}^\theta \langle Z_k | Z_{0:k-1} \rangle \mathbb{1}_X - \pi_Z^\theta(Z_{-\infty:k-1})|$  converge  $\mathbb{P}_Z$ -p.s. vers zéro quand  $k \rightarrow \infty$ . Ce résultat peut être obtenu en établissant l'oubli exponentiel de la distribution initiale (*the exponential forgetting of the initial distribution*) de la  $L$ -vraisemblance conditionnelle, et dans ce but, on emploie la technique de couplage introduite par Kleptsyna and Veretennikov [2008] et affinée par Douc et al. [2009]. Alors, puisque le processus  $\{Z_k : k \in \mathbb{Z}\}$  est stationnaire et ergodique, le théorème ergodique de Birkhoff s'applique, et (26) s'ensuit, avec  $\mathbf{L}(\theta) = \mathbb{E}_Z [\ln \pi_Z^\theta(Z_{-\infty:0})]$ , à condition que cette dernière quantité soit finie (l'approche utilisée ici a été proposée par Baum and Petrie [1966] dans le cadre des HMM d'espace d'état fini, puis généralisée par Douc et al. [2004] dans le cadre des HMM d'espace d'état général, mais sous des conditions techniques strictes). Quand la première étape est achevée, on applique Pfanzagl [1969] pour obtenir la deuxième étape, à condition qu'on puisse montrer que  $\theta \mapsto \mathbf{L}(\theta)$  est semi-continue.

### 3.1.2 Convergence de l'estimateur du maximum de $L$ -vraisemblance

Les hypothèses suivantes peuvent être utilisées pour obtenir la convergence de l'estimateur du maximum de  $L$ -vraisemblance  $\hat{\theta}_{X,n}^L$  de la  $L$ -vraisemblance  $\chi \mathbf{L}^\theta \langle Z_{0:n-1} \rangle \mathbb{1}_X$ . Avant de lister ces hypothèses, on introduit la définition suivante, qui nous sera utile.

**Définition 3.4** (Ensemble de Doeblin  $L$ -local). Soit  $C \in \mathcal{X}$  un sous-ensemble non-vide de  $X$ . L'ensemble  $C$  est appelé *ensemble de Doeblin  $L$ -local* par rapport à la famille  $\{\mathbf{L}^\theta \langle z \rangle : (\theta, z) \in \Theta \times \mathbb{Z}\}$  s'il existe des fonctions positives  $\epsilon_C^- : \mathbb{Z} \rightarrow \mathbb{R}_+$ ,  $\epsilon_C^+ : \mathbb{Z} \rightarrow \mathbb{R}_+$ , une famille de la mesure de probabilités  $\{\mu_C^\theta \langle z \rangle : (\theta, z) \in \Theta \times \mathbb{Z}\}$  sur  $(X, \mathcal{X})$  et une famille de fonctions positives  $\{\varphi_C^\theta \langle z \rangle : (\theta, z) \in \Theta \times \mathbb{Z}\}$  de  $X$  à  $\mathbb{R}$ , telles que pour tout  $(\theta, z) \in \Theta \times \mathbb{Z}$ ,  $\mu_C^\theta \langle z \rangle(C) = 1$ , et pour toute fonction positive et mesurable  $f : (X, \mathcal{X}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  et  $x \in C$ ,

$$\epsilon_C^- \langle z \rangle \varphi_C^\theta \langle z \rangle(x) \mu_C^\theta \langle z \rangle(\mathbb{1}_C f) \leq \delta_x \mathbf{L}^\theta \langle z \rangle(\mathbb{1}_C f) \leq \epsilon_C^+ \langle z \rangle \varphi_C^\theta \langle z \rangle(x) \mu_C^\theta \langle z \rangle(\mathbb{1}_C f).$$

On remarque que l'ensemble de Doeblin  $L$ -local dans la définition 3.4 est similaire à l'ensemble dit de *Doeblin  $r$ -local* défini dans [Douc and Moulines, 2012, Définition 1], hormis le fait que le premier ensemble est défini par rapport à la famille générale de noyaux  $\{\mathbf{L}^\theta\langle z \rangle\}_{(\theta,z) \in \Theta \times \mathcal{Z}}$ , alors que le deuxième est défini par rapport à la famille de noyaux de transition spécifique qui spécifie les HMM.

(L-2) Il existe un ensemble  $K \in \mathcal{Z}$  tel que les conditions suivantes se tiennent.

- (i)  $\mathbb{P}_Z(Z_0 \in K) > 2/3$ .
- (ii) Pour tout  $\eta > 0$ , il existe un ensemble de Doeblin  $L$ -local  $C \in \mathcal{X}$  tel que pour tout  $\theta \in \Theta$  et  $z \in K$ ,

$$\sup_{x \in C^c} \delta_x \mathbf{L}^\theta\langle z \rangle \mathbb{1}_X \leq \eta \sup_{x \in X} \delta_x \mathbf{L}^\theta\langle z \rangle \mathbb{1}_X < \infty \quad (27)$$

et

$$\inf_{z \in K} \frac{\epsilon_C^-(z)}{\epsilon_C^+(z)} > 0, \quad (28)$$

où les fonctions  $\epsilon_C^+$  et  $\epsilon_C^-$  sont données dans la définition 3.4.

(L-3) Il existe un ensemble  $D \in \mathcal{X}$  tel que

$$\mathbb{E}_Z \left[ \ln^- \inf_{\theta \in \Theta} \inf_{x \in D} \mathbf{L}^\theta\langle Z_0 \rangle(x; D) \right] < \infty. \quad (29)$$

(L-4) On a

$$\mathbb{E}_Z \left[ \ln^+ \sup_{\theta \in \Theta} \sup_{x \in X} \mathbf{L}^\theta\langle Z_0 \rangle(x; X) \right] < \infty. \quad (30)$$

(L-5) Il existe un nombre entier positif  $p$  tel que pour tout  $n \geq p$  et  $x \in X$ , la fonction  $\theta \mapsto \mathbf{L}^\theta\langle Z_{0:n-1} \rangle(x; X)$  est  $\mathbb{P}_Z$ -p.s. continue sur  $\Theta$ .

Enfin, pour énoncer notre résultat, la définition ci-dessous est nécessaire.

**Définition 3.5.** Soit  $A \in \mathcal{X}$  un sous-ensemble mesurable de  $X$ . La collection  $\mathcal{M}^+(A)$  note l'ensemble de toutes les distributions de probabilité sur  $(X, \mathcal{X})$  qui ne s'annulent pas sur  $A$ , c'est-à-dire que si  $\chi \in \mathcal{P}(X, \mathcal{X})$  telle que  $\chi(A) > 0$ , alors  $\chi \in \mathcal{M}^+(A)$ .

Les hypothèses (L-1)–(L-4) impliquent que pour toute mesure de probabilité  $\chi \in \mathcal{M}^+(D)$ , il existe cette fonction  $\mathbb{P}_Z$ -intégrable  $\pi_Z^\theta : Z^{\mathbb{Z}^-} \rightarrow \mathbb{R}$  telle que  $|\chi \mathbf{L}^\theta\langle Z_k | Z_{0:k-1} \rangle \mathbb{1}_X - \pi_Z^\theta(Z_{-\infty:k-1})|$  converge  $\mathbb{P}_Z$ -p.s. vers zéro quand  $k \rightarrow \infty$ . En conséquence, on a le théorème suivant.

**Théorème 3.6.** *On suppose que (L-1)–(L-5) se tiennent. On suppose de plus que pour tout  $\chi \in \mathcal{M}^+(D)$ ,  $\mathbb{P}_Z$ -p.s.,*

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in \Theta} n^{-1} \left| \ln \chi \check{\mathbf{L}}^\theta \langle Z_{0:n-1} \rangle \mathbb{1}_X - \ln \chi \mathbf{L}^\theta \langle Z_{0:n-1} \rangle \mathbb{1}_X \right| = 0,$$

où pour tout  $n \in \mathbb{Z}_+^*$ ,  $z_{0:n-1} \in \mathcal{Z}^n$  et  $\theta \in \Theta$ ,  $\check{\mathbf{L}}^\theta \langle z_{0:n-1} \rangle$  est un noyau positif sur  $X \times X$  et tel que  $\theta \mapsto \chi \check{\mathbf{L}}^\theta \langle z_{0:n-1} \rangle \mathbb{1}_X$  est continue. Alors, pour tout  $\chi \in \mathcal{M}^+(D)$ ,

$$\lim_{n \rightarrow \infty} \Delta(\check{\theta}_{\chi,n}^L, \Theta_\star^L) = 0, \quad \mathbb{P}_Z\text{-p.s.},$$

où

$$\check{\theta}_{\chi,n}^L \in \operatorname{argmax}_{\theta \in \Theta} n^{-1} \ln \chi \check{\mathbf{L}}^\theta \langle Z_{0:n-1} \rangle \mathbb{1}_X$$

et

$$\Theta_\star^L := \operatorname{argmax}_{\theta \in \Theta} \mathbb{E}_Z [\ln \pi_Z^\theta(Z_{-\infty:0})].$$

Le résultat ci-dessous est une implication directe du théorème 3.6.

**Corollaire 3.7.** *On suppose que (L-1)–(L-5) se tiennent. Alors, pour tout  $\chi \in \mathcal{M}^+(D)$ ,  $\mathbb{P}_Z$ -p.s.,*

$$\lim_{n \rightarrow \infty} \Delta(\hat{\theta}_{\chi,n}^L, \Theta_\star^L) = 0, \quad \mathbb{P}_Z\text{-p.s.},$$

où  $\hat{\theta}_{\chi,n}^L$  est l'estimateur du maximum de  $L$ -vraisemblance défini dans (25).

### 3.2 Consistance de l'EMV dans la classe des PMM entièrement dominés

Maintenant, on considère la consistance de l'EMV pour une sous-classe des PMM considérés dans la section 2, à savoir la classe des PMM entièrement dominés, dans le cadre des modèles bien spécifiés. Cette classe contient beaucoup de modèles d'intérêt, par exemple des HMM, des modèles de Markov à changement de régime et beaucoup des chaînes de Markov en environnement aléatoire, dont les lois de distribution sont dominées par des mesures  $\sigma$ -finies fixées. Le résultat principal de cette présente section est obtenu par la combinaison du résultat de convergence de l'estimateur du maximum de  $L$ -vraisemblance dérivé dans la section 3.1, avec le résultat d'identifiabilité obtenu dans la section 2.

**Définition 3.8.** On dit que le modèle de Markov partiellement observé  $\{K^\theta : \theta \in \Theta\}$  donné dans la définition 2.1 de la section 2 est entièrement dominé

s'il existe de plus une mesure  $\sigma$ -finie positive  $\mu$  sur  $\mathsf{X}$  telle que pour tout  $(x, y), (x', y') \in \mathsf{X} \times \mathsf{Y}$ ,

$$Q^\theta((x, y); dx') = q^\theta((x, y); x') \mu(dx') \quad (31)$$

et si la fonction de densité conditionnelle  $q^\theta$  satisfait par ailleurs

$$q^\theta((x, y); x') > 0. \quad (32)$$

Il s'ensuit que  $\kappa^\theta$  dans (4) satisfait, pour tout  $x, x' \in \mathsf{X}$  et  $y, y' \in \mathsf{Y}$ ,

$$\kappa^\theta\langle y, y' \rangle(x; dx') := q^\theta((x, y); x') g^\theta((x, y, x'); y') \mu(dx').$$

Dans cette section, on restreint notre considération au cas où  $\mathsf{X}$  est un espace métrique complet et séparable muni d'un métrique  $d_X$  et où  $\mathcal{X}$  est la  $\sigma$ -algèbre borélienne associée. Pour tout nombre entier  $s \geq 1$ ,  $y_{0:s} \in \mathsf{Y}^{s+1}$  et  $\theta \in \Theta$ , on définit le noyau (possiblement non-normalisé)  $\ell^\theta\langle y_{0:s} \rangle$  sur  $\mathsf{X} \times \mathcal{X}$  par, pour tout  $(x_0, A) \in \mathsf{X} \times \mathcal{X}$ ,

$$\ell^\theta\langle y_{0:s} \rangle(x_0; A) := \int_{\mathsf{X}^s} \mathbb{1}_A(x_s) \prod_{i=0}^{s-1} \kappa^\theta\langle y_i, y_{i+1} \rangle(x_i; dx_{i+1}). \quad (33)$$

On remarque que pour tout  $\theta \in \Theta$ , le noyau  $\ell^\theta$  est positif et le terme de gauche de (33) est strictement positif quand  $A = \mathsf{X}$ . Par définition, la log-vraisemblance  $L_{\xi, n}(\theta)$  donnée par (6) peut être exprimée comme, pour toute mesure  $\chi \in \mathcal{P}(\mathsf{X}, \mathcal{X})$  et  $\xi = \chi \otimes \delta_{y_0}$ ,

$$L_{\xi, n}(\theta) = \ln \chi \ell^\theta\langle y_{0:n} \rangle \mathbb{1}_{\mathsf{X}}.$$

On présente les hypothèses supplémentaires qui produisent la consistance de l'EMV définie par (5).

(F-1) L'hypothèse (K-1) se tient. Par ailleurs, pour tout  $\theta_* \in \Theta$ , il existe un entier  $r \geq 1$  et un ensemble  $\mathsf{K} \in \mathcal{Y}^{\otimes(r+1)}$  tels que, en notant  $\mathsf{Z} := \mathsf{Y}^{r+1}$  et, pour  $k \in \mathbb{Z}$ ,

$$z_i := y_{kr:(k+1)r} \in \mathsf{Z}, \quad Z_k := Y_{kr:(k+1)r} \in \mathsf{Z},$$

chacune des conditions suivantes se tient.

- (i)  $\tilde{\mathbb{P}}^{\theta_*}(Z_0 \in \mathsf{K}) > 2/3$ .
- (ii) Pour tout  $\eta > 0$ , il existe un ensemble de Doeblin  $L$ -local  $C \in \mathcal{X}$  par rapport à la famille  $\{\ell^\theta\langle z \rangle\}_{(\theta, z) \in \Theta \times \mathsf{Z}}$  (donné dans la définition 3.4 avec  $\ell^\theta$  défini par (33)) tel que pour tout  $\theta \in \Theta$  and  $z \in \mathsf{K}$ ,

$$\sup_{x \in C^c} \ell^\theta\langle z \rangle(x; \mathsf{X}) \leq \eta \sup_{x \in \mathsf{X}} \ell^\theta\langle z \rangle(x; \mathsf{X}) < \infty \quad (34)$$

et

$$\inf_{z \in \mathcal{K}} \frac{\epsilon_C^-(z)}{\epsilon_C^+(z)} > 0, \quad (35)$$

où les fonctions  $\epsilon_C^+$  et  $\epsilon_C^-$  sont données dans la définition 3.4.

(iii) Il existe un ensemble  $D \in \mathcal{X}$  tel que

$$\tilde{\mathbb{E}}^{\theta_*} \left[ \ln^- \inf_{\theta \in \Theta} \inf_{x \in D} \ell^\theta \langle Z_0 \rangle (x; D) \right] < \infty. \quad (36)$$

(F-2) On a,

$$\tilde{\mathbb{E}}^{\theta_*} \left[ \ln^+ \sup_{\theta \in \Theta} \sup_{(x, x') \in \mathcal{X}^2} g^\theta((x, Y_0, x'); Y_1) \right] < \infty. \quad (37)$$

(F-3) Pour tout  $\theta_* \in \Theta$ , il existe un entier  $p \geq 1$  tel que pour tout  $n \geq p$  et  $x \in \mathcal{X}$ , la fonction  $\theta \mapsto \ell^\theta \langle Y_{0:n} \rangle (x; \mathcal{X})$  est  $\tilde{\mathbb{P}}^{\theta_*}$ -p.s. continue sur  $\Theta$ .

Ces hypothèses peuvent être utilisées pour dériver la convergence de l'EMV dans le cadre des modèles mal spécifiés, hormis le fait que dans (F-1), plutôt que de supposer que l'hypothèse (K-1) se tient, on suppose à la place que, avec le même  $r$ , le processus observé  $\{Y_{k:k+r} : k \in \mathbb{Z}\}$  est stationnaire et ergodique. Cela généralise alors le résultat obtenu dans la section 2.4 et élargit le résultat de Douc and Moulines [2012] au-delà des HMM.

Il est toujours souhaitable que la consistance soit montrée insensible par rapport à toute mesure de probabilité initiale  $\xi$  sur  $\mathbf{X} \times \mathbf{Y}$ . Malheureusement, toutes les mesures de probabilité  $\xi$  ne peuvent satisfaire ceci; et il est donc important de considérer une classe de mesures de probabilité initiales de telle sorte que le résultat se tienne. Pour l'entier  $r$  et l'ensemble  $D \in \mathcal{X}$  défini dans (F-1), soit  $\mathcal{M}^{\theta_*}(D, r)$  le sous-ensemble de  $\mathcal{P}(\mathbf{X}, \mathcal{X})$  des mesures de probabilité sur  $(\mathbf{X}, \mathcal{X})$  satisfaisant

$$\mathcal{M}^{\theta_*}(D, r) = \left\{ \chi \in \mathcal{P}(\mathbf{X}, \mathcal{X}) : \tilde{\mathbb{E}}^{\theta_*} \left[ \ln^- \inf_{\theta \in \Theta} \chi \ell^\theta \langle Y_{0:s} \rangle \mathbb{1}_D \right] < \infty, \text{ pour tout } s \in \{1, \dots, r\} \right\}.$$

On a maintenant tous les outils nécessaires pour dériver les résultats concernant la consistance de l'EMV pour des PMM entièrement dominés. Ces résultats sont énoncés comme suit.

**Théorème 3.9.** *On suppose (K-1), (F-1)–(F-3). Alors, pour tout  $\chi \in \mathcal{M}^{\theta_*}(D, r)$ , en notant  $\xi = \chi \otimes \delta_{Y_0}$ , on a*

$$\lim_{n \rightarrow \infty} \Delta \left( \hat{\theta}_{\xi, n}, [\theta_*] \right) = 0, \quad \tilde{\mathbb{P}}^{\theta_*}\text{-p.s.},$$

où  $\hat{\theta}_{\xi,n}$  est défini par (5) et  $[\theta_\star]$  est la classe d'équivalence de  $\theta_\star$  comme définie dans la définition 2.2.

## 4 Consistance de l'EMV dans la classe des ODM

Cette section a pour but d'établir des propriétés asymptotiques pour la classe générale des ODM de premier ordre considérée plus tôt dans la section 2.5. L'existence d'une solution ergodique et la consistance de l'EMV sont dérivées dans des conditions qui sont faciles à vérifier. On démontre nos résultats avec des ODM spécifiques, qui incluent les modèles NBIN-GARCH et NM-GARCH mentionnés dans la section 1.2, montrant la consistance de l'EMV pour ces deux modèles.

### 4.1 Consistance de classe d'équivalence de l'EMV

Dans cette partie, on restreint notre considération au cas où  $\mathsf{X}$  est un espace métrique complet et séparable muni d'un métrique  $d_{\mathsf{X}}$  et  $\mathcal{X}$  est la  $\sigma$ -algèbre borélienne associée. On suppose toujours que l'hypothèse (K-1) se tient. Voici une liste des hypothèses supplémentaires sur lesquelles notre résultat de convergence repose.

(A-1) Il existe une fonction  $\bar{V} : \mathsf{X} \rightarrow \mathbb{R}_+$  telle que, pour tout  $\theta \in \Theta$ ,  $\pi_1^\theta(\bar{V}) < \infty$ .

L'hypothèse (A-1) est souvent obtenue comme produit de la preuve de l'hypothèse (A-1); voir la section 4.2. Il est ici énoncé comme hypothèse par commodité. L'ensemble des conditions suivantes peut être aisément vérifié sur  $g^\theta$  et  $\psi^\theta$ .

(B-1) Pour tout  $y \in \mathsf{Y}$ , la fonction  $(\theta, x) \mapsto g^\theta(x; y)$  est continue sur  $\Theta \times \mathsf{X}$ .

(B-2) Pour tout  $y \in \mathsf{Y}$ , la fonction  $(\theta, x) \mapsto \psi_y^\theta(x)$  est continue on  $\Theta \times \mathsf{X}$ .

La fonction  $\bar{V}$  qui apparaît dans (B-3)(viii) ci-dessous est la même que celle dans l'hypothèse (A-1). Par ailleurs, dans cette condition et tout au long de cette section, on écrit  $f \lesssim V$  pour une fonction réelle  $f$  et pour une fonction positive  $V$  définie sur le même espace  $\mathsf{X}$ , lorsqu'il existe une constante positive  $c$  telle que  $|f(x)| \leq cV(x)$  pour tout  $x \in \mathsf{X}$ .

(B-3) Il existe  $x_1 \in \mathsf{X}$ , un ensemble fermé  $\mathsf{X}_1 \subseteq \mathsf{X}$ ,  $\varrho \in (0, 1)$ ,  $C \geq 0$  et des fonctions mesurables  $\bar{\psi} : \mathsf{X} \rightarrow \mathbb{R}_+$ ,  $H : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  et  $\bar{\phi} : \mathsf{Y} \rightarrow \mathbb{R}_+$  tels que les affirmations suivantes se tiennent.

(i) Pour tout  $\theta \in \Theta$  et  $(x, y) \in \mathsf{X} \times \mathsf{Y}$ ,  $\psi_y^\theta(x) \in \mathsf{X}_1$ .



(ii)  $\sup_{(\theta,x,y) \in \Theta \times X_1 \times Y} g^\theta(x; y) < \infty.$

(iii) Pour tout  $\theta \in \Theta$ ,  $n \in \mathbb{Z}_+$ ,  $x \in X$ , et  $y_{1:n} \in Y^n$ ,

$$d_X(\psi^\theta\langle y_{1:n} \rangle(x_1), \psi^\theta\langle y_{1:n} \rangle(x)) \leq \varrho^n \bar{\psi}(x). \quad (38)$$

(iv)  $\bar{\psi}$  est localement bornée.

(v) Pour tout  $\theta \in \Theta$  et  $y \in Y$ ,  $\bar{\psi}(\psi_y^\theta(x_1)) \leq \bar{\phi}(y).$

(vi) Pour tout  $\theta \in \Theta$  et  $(x, x', y) \in X_1 \times X_1 \times Y$ ,

$$\left| \ln \frac{g^\theta(x; y)}{g^\theta(x'; y)} \right| \leq H(d_X(x, x')) e^{C(d_X(x_1, x) \vee d_X(x_1, x'))} \bar{\phi}(y). \quad (39)$$

(vii)  $H(u) = O(u)$  quand  $u \rightarrow 0.$

(viii) Si  $C = 0$ , alors, pour tout  $\theta \in \Theta$ ,

$$G^\theta \ln^+ \bar{\phi} \lesssim \bar{V}; \quad (40)$$

sinon, pour tout  $\theta \in \Theta$ ,

$$G^\theta \bar{\phi} \lesssim \bar{V}. \quad (41)$$

On remarque que les hypothèses ci-dessus impliquent (C-1), (C-2) et (C-3). On énonce maintenant notre résultat principal comme suit.

**Théorème 4.1.** *On suppose que (K-1), (A-1), (B-1), (B-2) et (B-3) se tiennent. Alors, en fixant  $x = x_1 \in X$  comme dans (B-3),*

$$\lim_{n \rightarrow \infty} \Delta(\hat{\theta}_{x,n}, [\theta_\star]) = 0, \quad \tilde{\mathbb{P}}^{\theta_\star}\text{-p.s.},$$

où  $\hat{\theta}_{x,n}$  est défini par (17) et la classe d'équivalence  $[\theta_\star]$  est donnée par (21).

## 4.2 Ergodicité

Dans cette section, l'ODM est étudié pour prouver la condition (A-1). Puisque c'est une condition de type « pour tout  $\theta$  (...) », pour alléger la notation, on omet l'exposant  $\theta$  de, par exemple,  $G^\theta$ ,  $R^\theta$  (donné par (14)) et  $\psi^\theta$ , et on écrit  $G$ ,  $R$  et  $\psi$  respectivement à la place.

L'ergodicité de chaînes de Markov est habituellement étudiée en utilisant la  $\varphi$ -irréductibilité. Cette approche est connue pour être assez efficace quand on travaille sur des modèles entièrement dominés; voir [Meyn and Tweedie \[2009\]](#). Il n'en va pas de même pour les ODM, où d'autres outils sont nécessaires; voir [Fokianos and Tjøstheim \[2011\]](#), [Douc et al. \[2013\]](#). Puisque

l'ergodicité est étudiée pour un paramètre donné  $\theta$ , les résultats d'ergodicité de Douc et al. [2013] s'appliquent directement, même si les ODM sont restreints au cas où  $g$  ne dépend pas du paramètre inconnu  $\theta$  dans cet article. Notre contribution principale ici est de se concentrer sur une liste d'hypothèses faciles à vérifier produisant les conditions d'ergodicité (A-1) et (A-1). On fournit également un lemme (le lemme 4.3) qui engendre la construction de fonctions instrumentales  $\alpha$  et  $\phi$  utilisées dans la liste des hypothèses.

- (A-2) L'espace mesurable  $(X, d_X)$  est un espace métrique compact localement, complet et séparable, et sa  $\sigma$ -algèbre associée  $\mathcal{X}$  est la  $\sigma$ -algèbre borélienne.
- (A-3) Il existe  $(\lambda, \beta) \in (0, 1) \times \mathbb{R}_+$  et une fonction mesurable  $V : X \rightarrow \mathbb{R}_+$  tels que  $RV \leq \lambda V + \beta$ , et  $\{V \leq M\}$  est compact pour tout  $M > 0$ .
- (A-4) Le noyau de Markov  $R$  est *weak Feller*, c'est-à-dire que pour toute fonction continue et bornée  $f$  définie sur  $X$ ,  $Rf$  est continue et bornée sur  $X$ .
- (A-5) Le noyau de Markov  $R$  a un point accessible, c'est-à-dire qu'il existe  $x_0 \in X$  tel que, pour tout  $x \in X$  et tout voisinage  $\mathcal{N}$  de  $x_0$ ,  $R^m(x; \mathcal{N}) > 0$  pour au moins un entier positif  $m$ .
- (A-6) On a 
$$\sup_{\substack{(x, x', y) \in X^2 \times Y \\ x \neq x'}} \frac{d_X(\psi_y(x), \psi_y(x'))}{d_X(x, x')} < 1.$$
- (A-7) Il existe une fonction mesurable  $\alpha$  de  $X^2$  à  $[0, 1]$ , une fonction mesurable  $\phi : X^2 \rightarrow X$  et une fonction mesurable  $W : X^2 \rightarrow [1, \infty)$  telles que les affirmations suivantes se tiennent.

- (i) Pour tout  $(x, x') \in X^2$  et  $y \in Y$ ,

$$\min \{g(x; y), g(x'; y)\} \geq \alpha(x, x')g(\phi(x, x'); y). \quad (42)$$

- (ii) Pour tout  $x \in X$ ,  $W(x, \cdot)$  est bornée de manière finie dans un voisinage de  $x$ , c'est-à-dire qu'il existe  $\gamma_x > 0$  tel que  $\sup_{x' \in B(x, \gamma_x)} W(x, x') < \infty$ .
- (iii) Pour tout  $(x, x') \in X^2$ ,  $1 - \alpha(x, x') \leq d_X(x, x')W(x, x')$ .
- (iv)  $\sup (\int_Y W(\psi_y(x), \psi_y(x')) G(\phi(x, x'); dy) - W(x, x')) < \infty$ , où le supremum est pris sur tout  $(x, x') \in X^2$ .

On peut maintenant énoncer le résultat principal d'ergodicité.

**Théorème 4.2.** *Les conditions (A-2), (A-3), (A-4), (A-5), (A-6) et (A-7) impliquent que  $K$  admet une distribution stationnaire unique  $\pi$  sur  $X \times Y$ . De plus,  $\pi_1 \bar{V} < \infty$  pour toute  $\bar{V} : X \rightarrow \mathbb{R}_+$  telle que  $\bar{V} \lesssim V$ .*

On fournit maintenant un cadre général de construction des fonctions  $\alpha$  et  $\phi$  qui apparaissent dans (A-7).

**Lemme 4.3.** *On suppose que  $\mathsf{X} = \mathsf{C}^{\mathsf{S}}$  pour un espace mesurable  $(\mathsf{S}, \mathcal{S})$  et  $\mathsf{C} \subseteq \mathbb{R}$ . Alors, pour tout  $x \in \mathsf{X}$ , on écrit  $x = (x_s)_{s \in \mathsf{S}}$ , où  $x_s \in \mathsf{C}$  pour tout  $s \in \mathsf{S}$ . On suppose par ailleurs que pour tout  $x = (x_s)_{s \in \mathsf{S}} \in \mathsf{X}$ , on peut exprimer la densité conditionnelle  $g(x; \cdot)$  comme un mélange de densités de la forme  $j(x_s)h(x_s; \cdot)$  sur  $s \in \mathsf{S}$ . C'est-à-dire que pour tout  $t \in \mathsf{C}$ ,  $y \mapsto j(t)h(t; y)$  est une densité par rapport à  $\nu$  et qu'il existe une mesure de probabilité  $\mu$  sur  $(\mathsf{S}, \mathcal{S})$  telle que*

$$g(x; y) = \int_{\mathsf{S}} j(x_s)h(x_s; y)\mu(ds), \quad y \in \mathsf{Y}. \quad (43)$$

*On suppose de plus que  $h$  prend des valeurs positives et que l'une des deux hypothèses suivantes se tient.*

- (H-1) Pour tout  $y \in \mathsf{Y}$ , la fonction  $h(\cdot; y) : t \mapsto h(t; y)$  est non-décroissante.  
(H-2) Pour tout  $y \in \mathsf{Y}$ , la fonction  $h(\cdot; y) : t \mapsto h(t; y)$  est non-croissante.

*Pour tout  $(x, x') \in \mathsf{X}^2$ , en notant  $x \wedge x' := (\min\{x_s, x'_s\})_{s \in \mathsf{S}}$  et  $x \vee x' := (\max\{x_s, x'_s\})_{s \in \mathsf{S}}$ , on définit  $\alpha(x, x')$  et  $\phi(x, x')$  comme*

$$\begin{cases} \alpha(x, x') = \inf_{s \in \mathsf{S}} \left\{ \frac{j(x_s \vee x'_s)}{j(x_s \wedge x'_s)} \right\} & \text{et } \phi(x, x') = x \wedge x' & \text{sous (H-1);} \\ \alpha(x, x') = \inf_{s \in \mathsf{S}} \left\{ \frac{j(x_s \wedge x'_s)}{j(x_s \vee x'_s)} \right\} & \text{et } \phi(x, x') = x \vee x' & \text{sous (H-2).} \end{cases}$$

*Alors,  $\alpha$  et  $\phi$  définies ci-dessus satisfont (A-7)(i).*

### 4.3 Exemples

Appliquons maintenant ces résultats pour prouver la consistance de classe d'équivalence de l'EMV pour les modèles NBIN-GARCH, NM-GARCH et le modèle INGARCH à seuil.

#### 4.3.1 Modèle NBIN-GARCH

On considère la récursion suivante :

$$\begin{aligned} X_{k+1} &= \omega + aX_k + bY_k, \\ Y_{k+1} | X_{0:k+1}, Y_{0:k} &\sim \mathcal{NB} \left( r, \frac{X_{k+1}}{1 + X_{k+1}} \right), \end{aligned} \quad (44)$$

où  $X_k$  prend des valeurs dans  $\mathsf{X} = \mathbb{R}_+$ ;  $Y_k$  prend des valeurs dans  $\mathbb{Z}_+$ ; et  $\theta = (\omega, a, b, r) \in (0, \infty)^4$  est un paramètre inconnu. Dans (44),  $\mathcal{NB}(r, p)$  représente la distribution binomiale négative avec les paramètres  $r > 0$  et

$p \in (0, 1)$ , c'est-à-dire que si  $Y \sim \mathcal{NB}(r, p)$ , alors  $\mathbb{P}(Y = k) = \frac{\Gamma(k+r)}{k!\Gamma(r)}(1-p)^r p^k$  pour tout  $k \geq 0$ , où  $\Gamma$  représente la fonction Gamma. Bien que de nombreuses analyses aient été menées sur ce modèle, à notre connaissance, la preuve complète de la consistance de l'EMV n'a pas été obtenue (voir la fin des discussions de la section 6 dans [Zhu \[2011\]](#) ainsi que [Davis and Liu \[2012\]](#) pour le traitement partiel).

Le modèle NBIN-GARCH est un cas particulier de la définition 2.10 où  $\nu$  est la mesure de comptage sur  $\mathbf{Y} = \mathbb{N}$ ,

$$\psi_y^\theta(x) = \omega + ax + by, \quad g^\theta(x; y) = \frac{\Gamma(y+r)}{y!\Gamma(r)} \left( \frac{1}{1+x} \right)^r \left( \frac{x}{1+x} \right)^y,$$

avec  $\theta = (\omega, a, b, r)$  dans un sous-ensemble compact  $\Theta$  de  $(0, \infty)^4$  et  $\mathbf{X} = (0, \infty)$ .

Dans le théorème 1 de [Zhu \[2011\]](#), l'équation satisfaite par la moyenne des observations  $\mu_k = \mathbb{E}[Y_k]$  est dérivée et est prouvée admettre une solution constante si et seulement si

$$rb + a < 1. \tag{45}$$

Ceci implique clairement que cette condition est nécessaire pour avoir une solution stationnaire  $\{Y_k\}$  avec une moyenne finie. Cependant, cela n'implique pas l'existence d'une telle solution. En réalité, le résultat suivant montre que (45) est en effet une condition nécessaire et suffisante pour avoir une solution stationnaire  $\{Y_k\}$  avec une moyenne finie. Par ailleurs, par le théorème 4.1, la consistance de l'EMV  $\hat{\theta}_{x_1, n}$  est obtenue pour tout  $x_1 \in \mathbf{X}$ .

**Théorème 4.4.** *On suppose que tout  $\theta = (\omega, a, b, r)$  dans  $\Theta$  satisfait la condition (45). Alors, les hypothèses (K-1) et (A-1) se tiennent avec  $\bar{V}$  définie comme la fonction d'identité sur  $\mathbf{X}$ . De plus, la conclusion du théorème 4.1 se tient avec tout  $x_1 \in \mathbf{X}$ .*

### 4.3.2 NM-GARCH Model

Soit  $d \in \mathbb{Z}_+^*$ . On considère la récursion suivante :

$$\begin{aligned} \mathbf{X}_{k+1} &= \boldsymbol{\omega} + \mathbf{A}\mathbf{X}_k + Y_k^2 \mathbf{b}, \\ Y_{k+1} | \mathbf{X}_{0:k+1}, Y_{0:k} &\sim G^\theta(\mathbf{X}_{k+1}; \cdot), \\ \frac{dG^\theta(\mathbf{x}; \cdot)}{d\nu}(y) &= \sum_{\ell=1}^d \gamma_\ell \frac{e^{-y^2/2x_\ell}}{(2\pi x_\ell)^{1/2}}, \quad \mathbf{x} \in (0, \infty)^d, y \in \mathbb{R}, \end{aligned} \tag{46}$$

où  $\nu$  est la mesure de Lebesgue sur  $\mathbb{R}$ ;  $\mathbf{X}_k = [X_{1,k} \dots X_{d,k}]^T$  prend des valeurs dans  $\mathbf{X} = \mathbb{R}_+^d$ ;  $\boldsymbol{\gamma} = [\gamma_1 \dots \gamma_d]^T$  est un vecteur de dimension  $d$

de coefficients mélangés appartenant au simplexe de dimension  $d$  :  $\mathbf{P}_d = \left\{ \boldsymbol{\gamma} \in \mathbb{R}_+^d : \sum_{\ell=1}^d \gamma_\ell = 1 \right\}$ ;  $\boldsymbol{\omega}$ ,  $\mathbf{b}$  sont des paramètres vectoriels de dimension  $d$  avec des entrées strictement positives et positives, respectivement; et  $\mathbf{A}$  est un paramètre matriciel  $d \times d$  avec des entrées positives. Ici, on a  $\theta = (\boldsymbol{\gamma}, \boldsymbol{\omega}, \mathbf{A}, \mathbf{b})$ . On remarque que  $G^\theta$  dépend de  $\theta$  seulement par les coefficients mélangés  $\gamma_1, \dots, \gamma_d$ . Si  $d = 1$ , on obtient le processus GARCH(1, 1) conditionnellement gaussien habituel. Dans ce cas, puisque  $\boldsymbol{\gamma} = \gamma_1 = 1$ ,  $G^\theta$  ne dépend plus de  $\theta$ . À notre connaissance, la preuve usuelle de la consistance de l'EMV pour GARCH ne peut pas être directement adaptée à ce modèle.

Le modèle NM( $d$ )-GARCH(1, 1) est un cas particulier de la définition 2.10 où  $\mathbf{X} = \mathbb{R}_+^d$  et  $\nu$  est la mesure de Lebesgue sur  $\mathbf{Y} = \mathbb{R}$ ,

$$\psi_y^\theta(\mathbf{x}) = \boldsymbol{\omega} + \mathbf{A}\mathbf{x} + y^2\mathbf{b}, \quad g^\theta(\mathbf{x}; y) = \sum_{\ell=1}^d \gamma_\ell \frac{e^{-y^2/2x_\ell}}{(2\pi x_\ell)^{1/2}}, \quad (\mathbf{x}, y) \in \mathbf{X} \times \mathbf{Y},$$

et  $\theta = (\boldsymbol{\gamma}, \boldsymbol{\omega}, \mathbf{A}, \mathbf{b}) \in \Theta$ , un sous-ensemble compact de  $\mathbf{P}_d \times (0, \infty)^d \times \mathbb{R}_+^{d \times d} \times \mathbb{R}_+^d$ .

Dans Haas et al. [2004], il est montré que l'équation satisfaite par la variance du processus NM( $d$ )-GARCH(1, 1) univarié admet une solution constante si et seulement si

$$|\lambda|_{\max}(\mathbf{A} + \mathbf{b}\boldsymbol{\gamma}^T) < 1, \quad (47)$$

où, pour toute matrice carrée  $\mathbf{M}$ ,  $|\lambda|_{\max}(\mathbf{M})$  représente le rayon spectral de  $\mathbf{M}$ . Il s'ensuit que l'existence d'une solution faiblement stationnaire implique (47), mais qu'on ne sait rien de l'existence d'une solution stationnaire ou faiblement stationnaire. Le résultat ci-dessous montre que (47) est en effet une condition suffisante pour l'existence d'une solution stationnaire avec une variance finie. Par ailleurs, par le théorème 4.1, la consistance de l'EMV  $\hat{\theta}_{\mathbf{x}_1, n}$  est obtenue pour tout  $\mathbf{x}_1 \in \mathbf{X}$ .

**Théorème 4.5.** *On suppose que tout  $\theta = (\boldsymbol{\gamma}, \boldsymbol{\omega}, \mathbf{A}, \mathbf{b})$  dans  $\Theta$  satisfait (47). Alors, les hypothèses (K-1) et (A-1) se tiennent avec  $\bar{V}$  définie comme une norme sur  $\mathbf{X}$ . De plus, la conclusion du théorème 4.1 se tient avec tout  $\mathbf{x}_1 \in \mathbf{X}$ .*

### 4.3.3 Modèle IN-GARCH à seuil

On considère la récursion suivante :

$$X_{k+1} = \omega + aX_k + bY_k, \quad Y_{k+1}|X_{0:k+1}, Y_{0:k} \sim \mathcal{P}(X_{k+1} \wedge \tau),$$

où  $X_k$  prend des valeurs dans  $\mathbf{X} = (0, \infty)$ ;  $Y_k$  prend des valeurs dans  $\mathbb{Z}_+$  et  $\theta = (\omega, a, b, \tau) \in (0, \infty)^4$  est un paramètre inconnu. En comparaison avec

le modèle IN-GARCH habituel, un seuil  $\tau$  a été ajouté à la distribution conditionnelle d'observation. Cela correspond au cas pratique où la variable cachée a une influence sur l'observation jusqu'à ce seuil.

Le modèle IN-GARCH(1, 1) à seuil est un cas particulier de la définition 2.10 où  $\nu$  est la mesure de comptage sur  $\mathbf{Y} = \mathbb{Z}_+$ ,

$$\psi_y^\theta(x) = \omega + ax + by, \quad g^\theta(x; y) = e^{-(x \wedge \tau)} \frac{(x \wedge \tau)^y}{y!},$$

avec  $\theta = (\omega, a, b, \tau)$  dans un sous-ensemble compact  $\Theta$  de  $(0, \infty)^4$  et  $\mathbf{X} = (0, \infty)$ . Dans ce modèle, si  $a < 1$ , on a alors les résultats d'ergodicité et de consistance comme énoncés dans le théorème 4.6 suivant.

**Théorème 4.6.** *On suppose que tout  $\theta = (\omega, a, b, \tau)$  dans  $\Theta$  satisfait  $a < 1$ . Alors, les hypothèses (K-1) et (A-1) se tiennent avec  $\bar{V}$  définie comme la fonction d'identité sur  $\mathbf{X}$ . Par ailleurs, la conclusion du théorème 4.1 se tient avec tout  $x_1 \in \mathbf{X}$ .*

## 4.4 Simulation

Nous considérons deux modèles NBIN-GARCH(1, 1) avec les paramètres suivants :

- (M.1)  $\theta_\star = (\omega_\star, a_\star, b_\star, r_\star) = (3, .2, .2, 2)$  et  
 (M.2)  $\theta_\star = (\omega_\star, a_\star, b_\star, r_\star) = (3, .35, .1, 1.5)$ .

Nous avons simulé  $m = 200$  ensembles de données pour chaque taille d'échantillon  $n = 2^7, 2^8, 2^9$  et  $2^{10}$ . Dans la figure 4.4, nous affichons les boxplots de la différence des log-vraisemblances normalisées évaluées respectivement à l'EMV et à la vraie valeur  $\theta_\star$ . Comme la théorie le prévoyait, cette différence semble converger vers 0 quand le nombre d'observations  $n \rightarrow \infty$ . Pour le modèle NBIN-GARCH(1, 1), on peut montrer que  $\Theta_\star = \{\theta_\star\}$ , qui implique la convergence de l'EMV vers le vrai paramètre. On peut observer ce comportement pour chaque composante de l'EMV pour les deux modèles des figures 4.5 et 4.6. Nous présentons également la moyenne de Monte Carlo en même temps que l'erreur de déviation absolue moyenne ou « mean absolute deviation error (MADE) » :  $\text{MADE} = m^{-1} \sum_{j=1}^m |\hat{\theta}_{x,n}^j - \theta_\star^j|$  comme critère d'évaluation pour le paramètre estimé dans la table 4.1.

TABLE 4.1 – Moyenne des estimations, MADE (entre parenthèses) pour les modèles NBIN-GARCH(1, 1)

Modèle	Paramètre	Taille de l'échantillon $n$			
		$n = 2^7$	$n = 2^8$	$n = 2^9$	$n = 2^{10}$
(M.1)	$\hat{\omega}$	3,311(0,973)	3,212(0,719)	3,108(0,507)	3,062(0,372)
	$\hat{a}$	0,165(0,138)	0,173(0,113)	0,187(0,076)	0,193(0,055)
	$\hat{b}$	0,194(0,049)	0,195(0,034)	0,197(0,025)	0,200(0,018)
	$\hat{r}$	2,045(0,241)	2,035(0,166)	2,020(0,112)	2,011(0,074)
(M.2)	$\hat{\omega}$	3,525(1,325)	3,362(1,258)	3,326(1,041)	3,167(0,761)
	$\hat{a}$	0,252(0,227)	0,290(0,213)	0,296(0,170)	0,319(0,136)
	$\hat{b}$	0,092(0,056)	0,097(0,039)	0,098(0,028)	0,100(0,022)
	$\hat{r}$	1,563(0,175)	1,539(0,129)	1,520(0,093)	1,513(0,066)

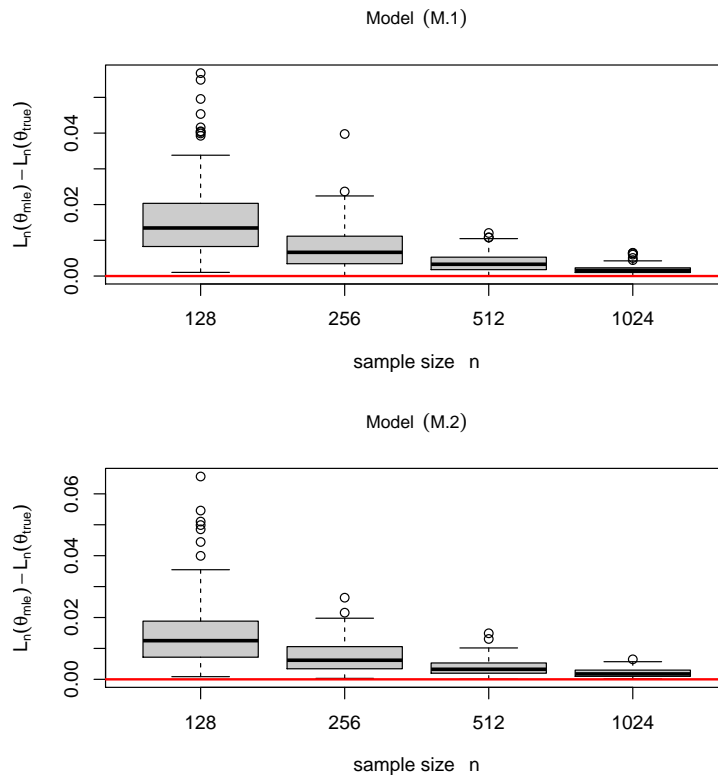


FIGURE 4.4 – Boxplots des différences des log-vraisemblances évaluées à l'EMV et à la vraie valeur pour les modèles (M.1) et (M.2) avec des tailles d'échantillon  $n = 2^7, 2^8, 2^9$  et  $n = 2^{10}$ , respectivement. La ligne "continue" rouge indique la position du zéro.

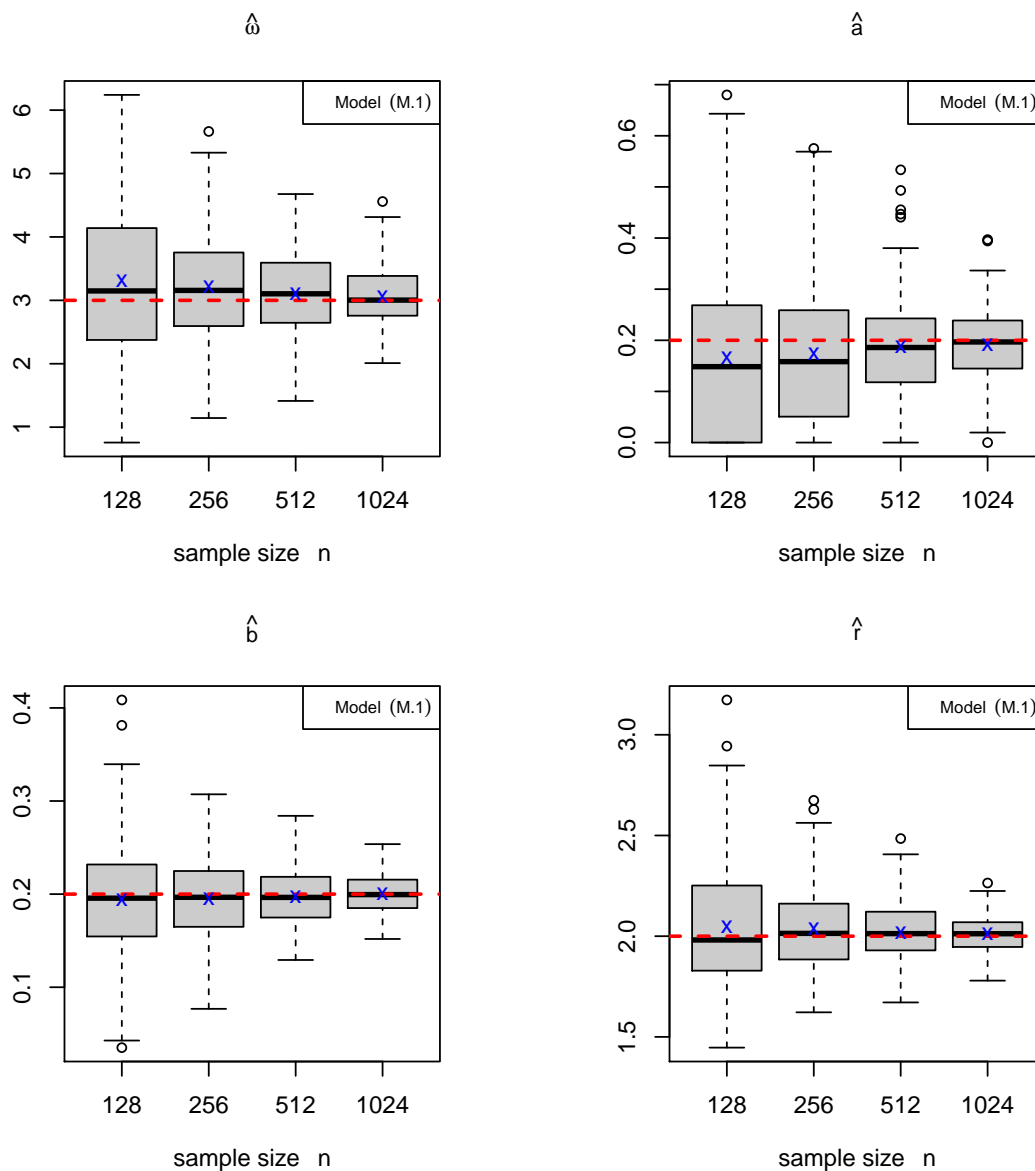


FIGURE 4.5 – Boxplots de l’EMV estimé pour le modèle (M.1) avec des tailles d’échantillon  $n = 2^7, 2^8, 2^9$  et  $n = 2^{10}$ , respectivement. La ligne “pointillée” rouge indique la vraie valeur du paramètre et le “x” bleu indique la moyenne de Monte Carlo de l’EMV.



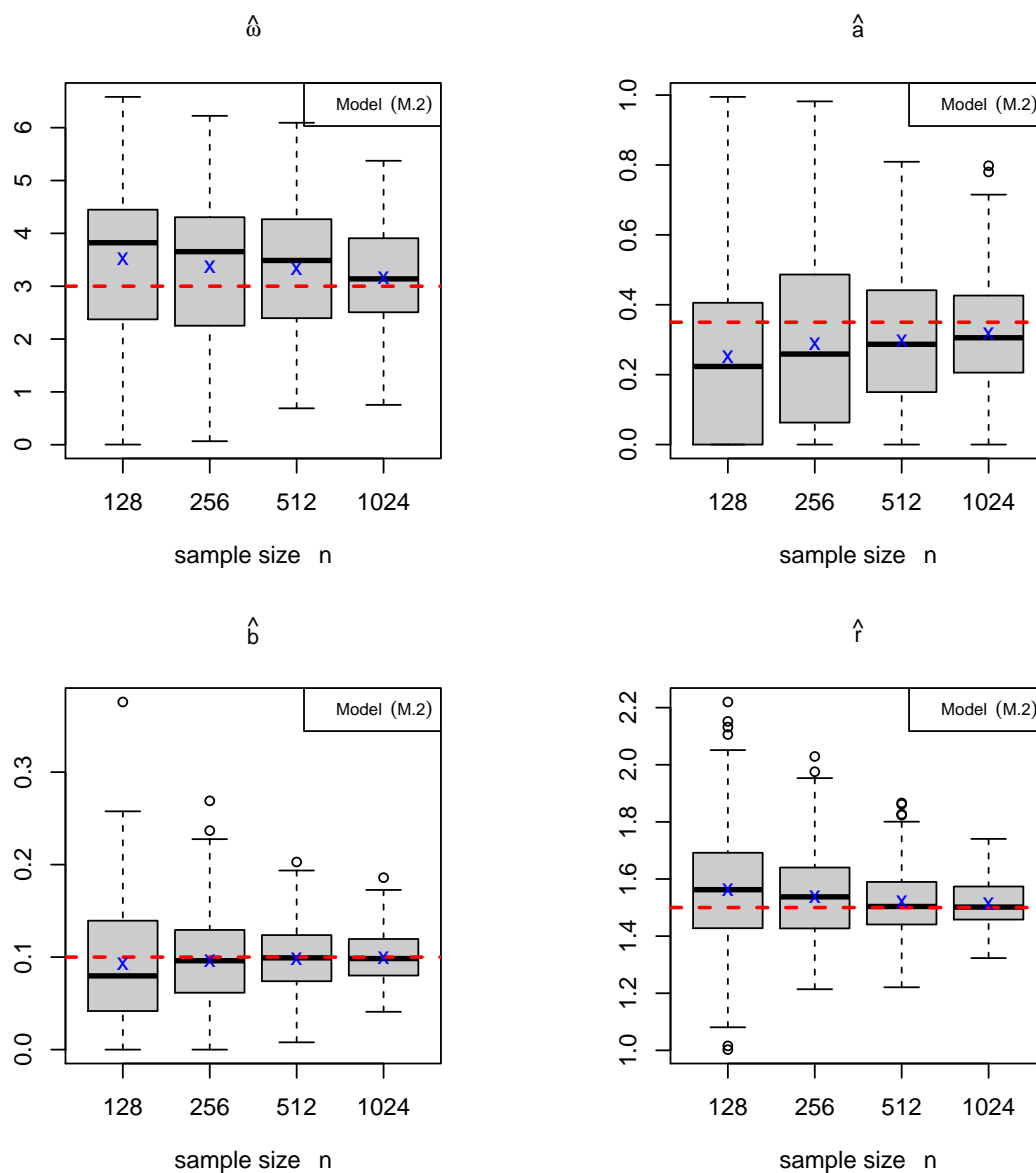


FIGURE 4.6 – Même légende que pour la figure 4.5 mais pour le modèle (M.2).

## 5 ODM d'ordre général : ergodicité, consistance et normalité asymptotique de l'EMV

Dans cette section, on permet à l'équation d'état du processus caché de l'ODM de dépendre également d'autres variables décalées des processus caché

et d'observation, ce qui en fait un ODM de type GARCH( $p, q$ ).

## 5.1 Définition et notation

**Définition 5.1.** Soient  $p, q \in \mathbb{Z}_+^*$ . Une série temporelle  $\{Y_k : k \geq -q + 1\}$  prenant des valeurs dans  $\mathsf{Y}$  est dite distribuée selon un ODM d'ordre  $(p, q)$  (ci-après dénommé, ODM( $p, q$ )) avec un paramètre  $\theta \in \Theta$  s'il existe une famille de fonctions mesurables  $\{(x_{1:p}, y_{1:q}) \mapsto \psi_{y_{1:q}}^\theta(x_{1:p}) : \theta \in \Theta\}$  de  $(\mathsf{X}^p \times \mathsf{Y}^q, \mathcal{X}^{\otimes p} \otimes \mathcal{Y}^{\otimes q})$  à  $(\mathsf{X}, \mathcal{X})$  et un processus  $\{X_k : k \geq -p + 1\}$  sur  $(\mathsf{X}, \mathcal{X})$  tels que pour tout  $k \in \mathbb{Z}_+$ ,

$$\begin{aligned} X_{k+1} &= \psi_{Y_{k-q+1:k}}^\theta(X_{k-p+1:k}), \\ Y_{k+1} \mid \mathcal{F}_k &\sim G^\theta(X_{k+1}; \cdot), \end{aligned} \tag{48}$$

où  $\mathcal{F}_k = \sigma(X_{-p+1:k+1}, Y_{-q+1:k})$ . De plus, on dit que le modèle est dominé par une mesure  $\sigma$ -finie  $\nu$  sur  $(\mathsf{Y}, \mathcal{Y})$  si pour tout  $x \in \mathsf{X}$ , le noyau de probabilité  $G^\theta(x; \cdot)$  est dominé par  $\nu$ . Dans ce cas, on note  $g^\theta(x; \cdot)$  sa dérivée de Radon–Nikodym,  $g^\theta(x; y) = \frac{dG^\theta(x; \cdot)}{d\nu}(y)$ , et on suppose toujours que pour tout  $(x, y) \in \mathsf{X} \times \mathsf{Y}$  et pour tout  $\theta \in \Theta$ ,

$$g^\theta(x; y) > 0.$$

Un des exemples les plus connus de cette classe est le modèle GARCH( $p, q$ ) général introduit par [Bollerslev \[1986\]](#), où  $\mathsf{X} = (0, \infty)$ ,  $\mathsf{Y} = \mathbb{R}$ ,  $G^\theta(x; \cdot)$  est la distribution gaussienne centrée de variance  $x$ , la fonction déterministe  $\psi_{y_{1:q}}^\theta(x_{1:p}) = \omega + \sum_{i=1}^p a_i x_i + \sum_{i=1}^q b_i y_i^2$  et  $\theta = (\omega, a_{1:p}, b_{1:q})$  avec  $\omega > 0$  et  $a_{1:p}, b_{1:q} \geq 0$ . Ce modèle a ensuite été amplement étudié, par exemple, par [Bougerol and Picard \[1992\]](#), [Francq et al. \[2004\]](#), [Francq and Zakoïan \[2009\]](#), [Lindner \[2009\]](#), [Francq and Zakoïan \[2011\]](#) et par les références qui y figurent. Pour d'autres exemples GARCH de cette classe, voir [Bollerslev \[2008\]](#).

Il est évident que, quand  $p = q = 1$ , alors l'ODM( $p, q$ ) défini par (48) se réduit à un ODM de premier ordre. On remarque qu'en fixant  $r := \max(p, q)$  et en définissant  $\varphi^\theta : \mathsf{X}^r \times \mathsf{Y}^r \rightarrow \mathsf{X}$  par, pour tout  $(x_{-r+1:0}, y_{-r+1:0}) \in \mathsf{X}^r \times \mathsf{Y}^r$ ,  $\varphi_{y_{-r+1:0}}^\theta(x_{-r+1:0}) := \psi_{y_{-q+1:0}}^\theta(x_{-p+1:0})$ , alors l'ODM( $p, q$ ) peut être inclus de manière générale dans un ODM( $r, r$ ). Par conséquent, sans perte de généralité, on peut toujours supposer que  $p = q$ .

Dans ce cadre, la vraisemblance conditionnelle des observations  $(Y_1, \dots, Y_n)$  sachant  $(X_{-p+1:0}, Y_{-q+1:0}) = (x_{-p+1:0}, y_{-q+1:0}) := z$ , pour une valeur  $z \in \mathsf{X}^p \times \mathsf{Y}^q$  arbitraire, par rapport à  $\nu^{\otimes n}$ , est, sous le paramètre  $\theta \in \Theta$ ,

$$y_{1:n} \mapsto \prod_{k=0}^{n-1} g^\theta(\psi_{y_{-q+1:k}}^\theta(x_{-p+1:0}); y_{k+1}),$$

où, pour tout vecteur  $y_{0:k}$ , la fonction  $\psi^\theta \langle y_{0:k} \rangle : \mathbf{X}^p \times \mathbf{Y}^q \rightarrow \mathbf{X}$  est successivement définie par, pour tout  $k \in \mathbb{Z}_+$ ,

$$\begin{aligned} \psi^\theta \langle y_{-q+1:k} \rangle (x_{-p+1:0}) \\ := \psi_{y_{-q+1:k}}^\theta \left( \psi^\theta \langle y_{-q+1:k-p} \rangle (x_{-p+1:0}), \dots, \psi^\theta \langle y_{-q+1:k-1} \rangle (x_{-p+1:0}) \right) \end{aligned}$$

avec la convention que, pour  $1 \leq j \leq p$ ,  $\psi^\theta \langle y_{-q+1:-j} \rangle (x_{-p+1:0}) := x_{-j+1}$ . Alors, l'EMV (conditionnel) correspondant  $\hat{\theta}_{z,n}$  du paramètre  $\theta$  est

$$\hat{\theta}_{z,n} \in \operatorname{argmax}_{\theta \in \Theta} \mathbb{L}_{z,n}^\theta, \quad (49)$$

où

$$\mathbb{L}_{z,n}^\theta := n^{-1} \sum_{k=0}^{n-1} \ln g^\theta \left( \psi^\theta \langle y_{-q+1:0}, Y_{1:k} \rangle (x_{-p+1:0}); Y_{k+1} \right). \quad (50)$$

Pour simplifier, on suppose que  $p = q$  et que  $p \geq 2$  et on pose  $\mathbf{Z} = \mathbf{X}^p \times \mathbf{Y}^{p-1}$  et  $\mathcal{Z} = \mathcal{X}^{\otimes p} \otimes \mathcal{Y}^{\otimes (p-1)}$ . Soit  $\Psi^\theta : \mathbf{Z} \times \mathbf{Y} \rightarrow \mathbf{Z}$  définie par, pour tout  $z = (z_{1:2p-1}) \in \mathbf{Z}$  et  $y \in \mathbf{Y}$ ,

$$\Psi_y^\theta(z) := \left( z_2, \dots, z_p, \psi_{z_{p+1:2p-1}, y}^\theta(z_{1:p}), z_{p+2}, \dots, z_{2p-1}, y \right).$$

La fonction  $\Psi^\theta$  est en effet mesurable sur  $(\mathbf{Z} \times \mathbf{Y}, \mathcal{Z} \otimes \mathcal{Y})$ . En posant  $Z_k = (X_{k-p+1:k}, Y_{k-p+1:k-1})$  et en observant que  $\mathcal{F}_k = \sigma(Z_{0:k+1}, Y_{0:k})$ , le modèle (48) peut être remplacé par : pour tout  $k \in \mathbb{Z}_+$ ,

$$\begin{aligned} Z_{k+1} &= \Psi_{Y_k}^\theta(Z_k), \\ Y_{k+1} \mid \mathcal{F}_k^{Z,Y} &\sim H^\theta(Z_{k+1}; \cdot), \end{aligned}$$

où, pour tout  $z = (z_{1:2p-1}) \in \mathbf{Z}$ ,

$$H^\theta(z; \cdot) := G^\theta(\Pi_p(z); \cdot),$$

et, pour tout  $j \in \{1, \dots, 2p-1\}$ ,  $\Pi_j(z) = z_j$ . Sous cette forme, l'ODM( $p, p$ ) est donc inclus dans un ODM( $1, 1$ ). En principe, ceci nous permet d'appliquer les mêmes résultats obtenus pour la classe des ODM( $1, 1$ ) à la classe plus générale des ODM( $p, p$ ). En tant qu'ODM( $1, 1$ ), le processus bivarié  $\{(Z_k, Y_k) : k \in \mathbb{Z}_+\}$  est une chaîne de Markov sur l'espace  $(\mathbf{Z} \times \mathbf{Y}, \mathcal{Z} \otimes \mathcal{Y})$  avec un noyau de transition  $K_g^\theta$  satisfaisant, pour tout  $(z, y) \in \mathbf{Z} \times \mathbf{Y}$ ,  $A \in \mathcal{Z}$  et  $B \in \mathcal{Y}$ ,

$$K_g^\theta((z, y); A \times B) = \iint \mathbb{1}_{A \times B}(z', y') \delta_{\Psi_y^\theta(z)}(dz') G^\theta(\Pi_p(z); dy').$$

On remarque aussi que, en lui-même, le processus  $\{Z_k : k \in \mathbb{Z}_+\}$  est une chaîne de Markov sur  $(Z, \mathcal{Z})$  avec un noyau de transition  $R_g^\theta$  satisfaisant, pour tout  $z \in Z$  et  $A \in \mathcal{Z}$ ,

$$R_g^\theta(z; A) = \int \mathbb{1}_A(\Psi_y^\theta(z)) H^\theta(z; dy) = \int \mathbb{1}_A(\Psi_y^\theta(z)) G^\theta(\Pi_p(z); dy).$$

Néanmoins, pour appliquer les résultats connus à cet ODM(1,1) intégré, des généralisations sont nécessaires. On suppose d'abord que  $(Y, d_Y)$  est un espace localement compact, complet et séparable muni d'une métrique  $d_Y$ . Puis, puisque  $(X, d_X)$  est un espace localement compact, complet et séparable, cela implique que l'espace  $(Z, d_Z)$  est aussi un espace localement compact, complet et séparable avec la métrique  $d_Z$  défini de manière appropriée comme une fonction de  $d_X$  et de  $d_Y$  de telle sorte que  $\mathcal{Z}$  est la  $\sigma$ -algèbre borélienne associée. On propose maintenant l'hypothèse d'ergodicité suivante.

(AG-1) Pour tout  $\theta \in \Theta$ , le noyau de transition  $K_g^\theta$  de la chaîne complète admet une distribution stationnaire unique  $\pi_g^\theta$  sur  $Z \times Y$ .

Dans l'hypothèse (AG-1), on note  $\pi_{g,1}^\theta$  et  $\pi_{g,2}^\theta$  les distributions marginales de  $\pi_g^\theta$  sur  $Z$  et sur  $Y$ , respectivement, et on définit les distributions de probabilité  $\mathbb{P}_g^\theta$ ,  $\tilde{\mathbb{P}}_g^\theta$  et la relation d'équivalence  $\sim$  de même manière que dans la définition 2.2.

## 5.2 Consistance de l'EMV

On liste ci-dessous les hypothèses supplémentaires suffisantes pour obtenir la convergence de l'EMV pour une classe d'ODM d'ordre général.

(AG-2) Il existe une fonction  $\bar{V} : Z \rightarrow \mathbb{R}_+$  telle que, pour tout  $\theta \in \Theta$ ,  $\pi_1^\theta(\bar{V}) < \infty$ .

(BG-1) Pour tout  $y \in Y$ , la fonction  $(\theta, x) \mapsto g^\theta(x; y)$  est continue sur  $\Theta \times X$ .

(BG-2) Pour tout  $y_{1:p} \in Y$ , la fonction  $(\theta, x_{1:p}) \mapsto \psi_{y_{1:p}}^\theta(x_{1:p})$  est continue sur  $\Theta \times X^p$ .

(BG-3) Il existe  $z_1 \in Z$ , un ensemble  $Z_1 \subseteq Z$  tel que  $\Pi_p(Z_1)$  est fermé,  $\varrho \in (0, 1)$ ,  $C \geq 0$  et des fonctions mesurables  $\bar{\Psi} : Z \rightarrow \mathbb{R}_+$ ,  $H : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  et  $\bar{\phi} : Y \rightarrow \mathbb{R}_+$ , tels que les affirmations suivantes se tiennent.

(i) Pour tout  $\theta \in \Theta$  et  $(z, y) \in Z \times Y$ ,  $\Pi_p(\Psi_y^\theta(z)) \in Z_1$ .

(ii)  $\sup_{(\theta, z, y) \in \Theta \times Z_1 \times Y} g^\theta(\Pi_p(z); y) < \infty$ .

(iii) Pour tout  $\theta \in \Theta$ ,  $n \in \mathbb{Z}_+$ ,  $z \in Z$ , et  $y_{1:n} \in Y^n$ ,

$$d_Z(\Psi^\theta\langle y_{1:n} \rangle(z_1), \Psi^\theta\langle y_{1:n} \rangle(z)) \leq \varrho^n \bar{\Psi}(z),$$

- (iv)  $\bar{\Psi}$  est localement bornée.
- (v) Pour tout  $\theta \in \Theta$  et  $y \in Y$ ,  $\bar{\Psi}(\Psi_y^\theta(z_1)) \leq \bar{\phi}(y)$ .
- (vi) Pour tout  $\theta \in \Theta$  et  $(z, z', y) \in Z_1 \times Z_1 \times Y$ ,

$$\left| \ln \frac{g^\theta(\Pi_p(z); y)}{g^\theta(\Pi_p(z'); y)} \right| \leq H(d_Z(z, z')) e^{C(d_Z(z_1, z) \vee d_Z(z_1, z'))} \bar{\phi}(y),$$

- (vii)  $H(u) = O(u)$  quand  $u \rightarrow 0$ .
- (viii) Si  $C = 0$ , alors, pour tout  $\theta \in \Theta$ ,

$$G^\theta \ln^+ \bar{\phi} \lesssim \bar{V};$$

sinon, pour tout  $\theta \in \Theta$ ,

$$G^\theta \bar{\phi} \lesssim \bar{V}.$$

On énonce maintenant la consistance de classe d'équivalence de l'EMV pour les ODM d'ordre général.

**Théorème 5.2.** *On suppose que (AG-1), (AG-2), (BG-1), (BG-2) et (BG-3) se tiennent. Alors, en posant  $z_1 \in Z$  comme dans (BG-3),*

$$\lim_{n \rightarrow \infty} \Delta(\hat{\theta}_{z_1, n}, [\theta_*]) = 0, \quad \tilde{\mathbb{P}}_g^{\theta_*}\text{-p.s.},$$

Avec les hypothèses ci-dessus, on peut appliquer le résultat de Douc et al. [2015] pour montrer que pour tout  $\theta, \theta_* \in \Theta$ , il existe une fonction finie  $\tilde{\mathbb{P}}_g^{\theta_*}$ -p.s. mesurable  $\Psi^\theta \langle \cdot \rangle : Z^{\mathbb{Z}^-} \rightarrow Z$  telle que pour une valeur appropriée  $z \in Z$ ,

$$\lim_{m \rightarrow \infty} \Psi^\theta \langle Y_{-m:0} \rangle(z) = \Psi^\theta \langle Y_{-\infty:0} \rangle, \quad \tilde{\mathbb{P}}_g^{\theta_*}\text{-p.s.} \quad (51)$$

$$Z_1 = \Psi^{\theta_*} \langle Y_{-\infty:0} \rangle, \quad \mathbb{P}_g^{\theta_*}\text{-p.s.} \quad (52)$$

La consistance forte de l'EMV découle alors du théorème suivant.

**Théorème 5.3.** *On suppose que (AG-1), (AG-2), (BG-1), (BG-2) et (BG-3) se tiennent. On suppose de plus que*

- (a) pour tout  $\theta = (\vartheta, r), \theta' = (\vartheta', r') \in \Theta$  et  $x, x' \in X$ ,

$$G^\theta(x; \cdot) = G^{\theta'}(x'; \cdot) \quad \text{implique} \quad r = r' \quad \text{et} \quad x = x',$$

- (b) pour tout  $\theta = (\vartheta, r), \theta_* = (\vartheta_*, r_*) \in \Theta$ ,  $\tilde{\mathbb{P}}_g^{\theta_*}$ -p.s.,

$$\Pi_p(\Psi^{(\vartheta, r)} \langle Y_{-\infty:0} \rangle) = \Pi_p(\Psi^{(\vartheta_*, r_*)} \langle Y_{-\infty:0} \rangle) \quad \text{implique} \quad \vartheta = \vartheta_*,$$

où  $\Psi^\theta \langle Y_{-\infty:0} \rangle$  est définie par (51) et (52). Alors, en posant  $z_1 \in Z$  comme dans (BG-3),

$$\lim_{n \rightarrow \infty} \Delta(\hat{\theta}_{z_1, n}, \theta_*) = 0, \quad \tilde{\mathbb{P}}_g^{\theta_*}\text{-p.s.} \quad (53)$$

### 5.3 Normalité asymptotique de l'EMV

Dans cette section, on traite d'un cas particulier mais important d'ODM où l'espace  $Z$  est un sous-ensemble d'un espace vectoriel de dimension finie ;  $\Theta$  est un sous-ensemble compact de  $\mathbb{R}^d$ , pour un entier positif  $d$  ; et la fonction  $\Psi_y^\theta(z)$  définie dans (5.6) admet la forme suivante : pour tout  $(z, y, \theta) \in Z \times Y \times \Theta$ ,

$$\Psi_y^\theta(z) = a(\theta, y)z + b(\theta, y), \quad (54)$$

où  $a(\theta, y)$  et  $b(\theta, y)$  sont deux matrices de dimensions appropriées, et pour tout  $\theta \in \Theta$ ,  $y \mapsto a(\theta, y)$  et  $y \mapsto b(\theta, y)$  sont des applications mesurables sur  $(Y, \mathcal{Y})$ . Dans ce qui suit, on note  $|\cdot|$  une norme vectorielle sur  $\mathbb{R}^d$  et  $\|\cdot\|$  une norme matricielle. On considère ensuite les hypothèses suivantes.

(CG-1) On a :

- (i) Pour tout  $y \in Y$ ,  $\theta \mapsto a(\theta, y)$  et  $\theta \mapsto b(\theta, y)$  sont de classe  $\mathcal{C}^2$  sur  $\Theta$ .
- (ii) Il existe des constantes  $C > 0$  et  $\rho \in (0, 1)$  telles que pour tout  $\theta \in \Theta$  et  $y_{1:n} \in Y$ ,

$$\left\| \prod_{\ell=1}^n a(\theta, y_\ell) \right\| \leq C\rho^n.$$

- (iii)  $\tilde{\mathbb{E}}^{\theta_*}[\sup_{\theta \in \Theta} \|b(\theta, Y_1)\|] < \infty$ .
- (iv) Pour tout  $i \in \{1, \dots, d\}$ ,

$$\tilde{\mathbb{E}}_g^{\theta_*} \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial a(\theta, Y_1)}{\partial \theta_i} \right\| \right] + \tilde{\mathbb{E}}_g^{\theta_*} \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial b(\theta, Y_1)}{\partial \theta_i} \right\| \right] < \infty.$$

- (v) Pour tout  $i, j \in \{1, \dots, d\}$ ,

$$\tilde{\mathbb{E}}_g^{\theta_*} \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial^2 a(\theta, Y_1)}{\partial \theta_i \partial \theta_j} \right\| \right] + \tilde{\mathbb{E}}_g^{\theta_*} \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial^2 b(\theta, Y_1)}{\partial \theta_i \partial \theta_j} \right\| \right] < \infty.$$

**Lemme 5.4.** *On suppose (AG-1) et (CG-1). Alors, pour tout  $\theta \in \Theta$ , il existe une fonction  $\tilde{\mathbb{P}}_g^{\theta_*}$ -p.s. mesurable finie  $\Psi^\theta(\cdot) : Z^{\mathbb{Z}^-} \rightarrow Z$  satisfaisant (51) et (52) pour tout  $z \in Z$ . La fonction  $\theta \mapsto \Psi^\theta\langle Y_{-\infty:0} \rangle$  est de classe  $\mathcal{C}^2$  et uniformément différentiable sur  $\Theta$ , et, uniformément sur  $\Theta$ , on a,  $\tilde{\mathbb{P}}_g^{\theta_*}$ -p.s.,*

$$\begin{aligned} \frac{\partial \Psi^\theta\langle Y_{-\infty:0} \rangle}{\partial \theta} &= \lim_{m \rightarrow \infty} \frac{\partial \Psi^\theta\langle Y_{-m:0} \rangle(z)}{\partial \theta} \\ \frac{\partial^2 \Psi^\theta\langle Y_{-\infty:0} \rangle}{\partial \theta^T \partial \theta} &= \lim_{m \rightarrow \infty} \frac{\partial^2 \Psi^\theta\langle Y_{-m:0} \rangle(z)}{\partial \theta^T \partial \theta}. \end{aligned}$$

En outre,

$$\begin{aligned} \tilde{\mathbb{E}}_g^{\theta_\star} \left[ \sup_{\theta \in \Theta} \left\| \Psi^\theta \langle Y_{-\infty:0} \rangle \right\| \right] &< \infty, \\ \tilde{\mathbb{E}}_g^{\theta_\star} \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial \Psi^\theta \langle Y_{-\infty:0} \rangle}{\partial \theta} \right\| \right] &< \infty, \\ \tilde{\mathbb{E}}_g^{\theta_\star} \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial^2 \Psi^\theta \langle Y_{-\infty:0} \rangle}{\partial \theta^T \partial \theta} \right\| \right] &< \infty. \end{aligned}$$

(CG-2) Pour tout  $y \in Y$ , la fonction  $(\theta, x) \mapsto g^\theta(x; y)$  de classe  $\mathcal{C}^2$  sur  $\Theta \times X$ .

Pour présenter les hypothèses qui vont suivre, quelques notations sont nécessaires. Pour tout  $k \in \mathbb{Z}_+^*$ ,  $\theta \in \theta$  et  $z \in Z$ , on pose

$$\ell_{z,k}^\theta := \ln g^\theta \left( \Pi_p \left( \Psi^\theta \langle Y_{1:k-1} \rangle (z) \right); Y_k \right).$$

Alors, pour tout  $n \in \mathbb{Z}_+^*$ , il découle que  $L_{z,n}^\theta$  définie par (50) peut être réécrite

$$L_{z,n}^\theta = n^{-1} \sum_{k=1}^n \ell_{z,k}^\theta.$$

Pour tout  $k \in \mathbb{Z}_+^*$ ,  $\theta \in \theta$  et  $z \in Z$ , on pose

$$\ell_k^\theta := \ln g^\theta \left( \Pi_p \left( \Psi^\theta \langle Y_{-\infty:k-1} \rangle \right); Y_k \right),$$

qui est bien définie  $\tilde{\mathbb{P}}_g^{\theta_\star}$ -p.s. par le lemme 5.4. Pour tout  $n \in \mathbb{Z}_+^*$  et  $\theta \in \Theta$ , on définit

$$L_n^\theta := n^{-1} \sum_{k=1}^n \ell_k^\theta.$$

On remarque que le lemme 5.4, (AG-1), (CG-1) et (CG-2) impliquent que les fonctions  $\theta \mapsto \ell_{z,k}^\theta$  et  $\theta \mapsto \ell_k^\theta$  sont de classe  $\mathcal{C}^2$  sur un voisinage  $\mathcal{V}(\theta_\star)$  de  $\theta_\star$ ,  $\tilde{\mathbb{P}}_g^{\theta_\star}$ -p.s.

Les hypothèses (AG-1), (CG-1) et (CG-2) ainsi que l'ensemble des hypothèses suivantes, (CG-3)–(CG-7), sont suffisantes pour obtenir la normalité asymptotique de l'EMV  $\hat{\theta}_{z,n}$  pour une valeur arbitraire  $z \in Z$ .

(CG-3) Le vrai paramètre  $\theta_\star$  réside à l'intérieur de  $\Theta$ .

(CG-4) La consistance forte se tient, c'est-à-dire que  $\lim_{n \rightarrow \infty} \hat{\theta}_{z,n} = \theta_\star$ ,  $\tilde{\mathbb{P}}_g^{\theta_\star}$ -p.s.

(CG-5) Il existe une constante  $\alpha > 1/2$  et une variable aléatoire  $\tilde{\mathbb{P}}_g^{\theta_\star}$ -p.s. finie  $C > 0$  telles que pour tout  $k \geq 1$ ,

$$\sup_{\theta \in \mathcal{V}(\theta_\star)} \left| \frac{\partial \ell_{z,k}^\theta}{\partial \theta} - \frac{\partial \ell_k^\theta}{\partial \theta} \right| \leq C k^{-\alpha}, \quad \tilde{\mathbb{P}}_g^{\theta_\star}\text{-p.s.},$$

où  $\mathcal{V}(\theta_\star)$  est un voisinage de  $\theta_\star$ .

(CG-6) Il existe un voisinage  $\mathcal{V}(\theta_*)$  de  $\theta_*$  tel que

$$(i) \quad \tilde{\mathbb{E}}_g^{\theta_*} \left[ \sup_{\theta \in \mathcal{V}(\theta_*)} \left| \frac{\partial \ell_1^\theta}{\partial \theta} \right|^2 \right] < \infty.$$

$$(ii) \quad \tilde{\mathbb{E}}_g^{\theta_*} \left[ \sup_{\theta \in \mathcal{V}(\theta_*)} \left\| \frac{\partial^2 \ell_1^\theta}{\partial \theta^T \partial \theta} \right\| \right] < \infty.$$

$$(iii) \quad \tilde{\mathbb{E}}_g^{\theta_*} \left[ \sup_{\theta \in \mathcal{V}(\theta_*)} \left| \frac{\partial}{\partial \theta} g^\theta (\Pi_p (\Psi^\theta \langle Y_{-\infty:0} \rangle); Y_1) \right| \right] < \infty.$$

$$(iv) \quad \tilde{\mathbb{E}}_g^{\theta_*} \left[ \sup_{\theta \in \mathcal{V}(\theta_*)} \left\| \frac{\partial^2}{\partial \theta^T \partial \theta} g^\theta (\Pi_p (\Psi^\theta \langle Y_{-\infty:0} \rangle); Y_1) \right\| \right] < \infty.$$

(CG-7) La matrice  $\Sigma_* = -\tilde{\mathbb{E}}_g^{\theta_*} \left[ \frac{\partial^2 \ell_1^{\theta_*}}{\partial \theta^T \partial \theta} \right]$  est inversible.

**Théorème 5.5.** *On suppose que (AG-1) et (CG-1)–(CG-7) se tiennent. Alors, pour tout  $z \in \mathbb{Z}$ ,*

$$\sqrt{n}(\hat{\theta}_{z,n} - \theta_*) \rightsquigarrow_{\theta_*} \mathcal{N}(0, \Sigma_*^{-1}),$$

où le symbole  $\rightsquigarrow_{\theta_*}$  représente la convergence faible sous  $\tilde{\mathbb{P}}_g^{\theta_*}$ , et  $\mathcal{N}(0, \Sigma_*^{-1})$  représente la distribution gaussienne centrée de matrice de covariance  $\Sigma_*^{-1}$ .

## 5.4 Ergodicité

On fournit maintenant les conditions d'obtention de la stationnarité et de l'ergodicité de la chaîne de Markov  $\{(Z_k, Y_k) : k \in \mathbb{Z}_+\}$ . Comme dans la section 4.2, on fixe  $\theta$  une valeur arbitraire de  $\Theta$  et puisque c'est une expression « pour tout  $\theta$  (...) », pour alléger la notation, on retire l'exposant  $\theta$ , par exemple, de  $G^\theta$ ,  $R_g^\theta$  et  $\psi^\theta$  et on écrit respectivement  $G$ ,  $R_g$  et  $\psi$  à la place.

(AG-3) L'espace mesurable  $(\mathbb{Z}, d_Z)$  est un espace métrique localement compact, complet et séparable, et sa  $\sigma$ -algèbre associée  $\mathcal{Z}$  est la  $\sigma$ -algèbre borélienne.

(AG-4) Il existe un entier positif  $q$  tel que le noyau de Markov  $(R_g)^q$  est *weak Feller*, c'est-à-dire que pour toute fonction continue et bornée  $f$  définie sur  $\mathbb{X}$ ,  $(R_g)^q f$  est continue et bornée sur  $\mathbb{X}$ . De plus, il existe  $(\lambda, \beta) \in (0, 1) \times \mathbb{R}_+$  et une fonction mesurable  $V : \mathbb{Z} \rightarrow \mathbb{R}_+$  tels que  $(R_g)^q V \leq \lambda V + \beta$  et  $\{V \leq M\}$  est un ensemble compact pour tout  $M > 0$ .

(AG-5) Le noyau de Markov  $R_g$  admet un point accessible, c'est-à-dire qu'il existe un point  $z_0 \in \mathbb{Z}$  tel que, pour tout  $z \in \mathbb{Z}$  et tout voisinage  $\mathcal{N}$  de  $z_0$ ,  $(R_g)^m(z; \mathcal{N}) > 0$  pour au moins un entier positif  $m$ .

(AG-6) Il existe un entier positif  $\ell$  tel que

$$\sup_{\substack{(z, z', y_{1:\ell}) \in \mathbb{Z}^2 \times \mathbb{Y}^\ell \\ z \neq z'}} \frac{d_Z(\Psi \langle y_{1:\ell} \rangle(z), \Psi \langle y_{1:\ell} \rangle(z'))}{d_Z(z, z')} < 1,$$



et

$$\sup_{\substack{(z, z', y) \in Z^2 \times Y \\ z \neq z'}} \frac{d_Z(\Psi_y(z), \Psi_y(z'))}{d_Z(z, z')} < \infty.$$

(AG-7) Il existe une fonction mesurable  $\alpha$  de  $Z^2$  à  $[0, 1]$ , une fonction mesurable  $\phi : Z^2 \rightarrow X = \Pi_p(Z)$  et une fonction mesurable  $W : Z^2 \rightarrow [1, \infty)$  telles qu'on a les affirmations suivantes.

(i) Pour tout  $(z, z') \in Z^2$  et tout  $y \in Y$ ,

$$\min \{g(\Pi_p(z); y), g(\Pi_p(z'); y)\} \geq \alpha(z, z')g(\phi(z, z'); y).$$

(ii) Pour tout  $z \in Z$ ,  $W(z, \cdot)$  est bornée de manière finie dans un voisinage de  $z$ , c'est-à-dire qu'il existe  $\gamma_z > 0$  tel que  $\sup_{z' \in B(z, \gamma_z)} W(z, z') < \infty$ .

(iii) Pour tout  $(z, z') \in Z^2$ ,  $1 - \alpha(z, z') \leq d_Z(z, z')W(z, z')$ .

(iv) Il existe un entier  $m > 0$  et un nombre réel  $D > 0$  tels que

$$\sup_{(z, z') \in Z^2} (J_m((z, z'); W) - W(z, z')) < \infty$$

et

$$\sup_{(z, z') \in Z^2} (J_1((z, z'); W) - DW(z, z')) < \infty,$$

où pour tout entier positif  $n$ ,  $(z, z') \in Z^2$  et toute fonction mesurable  $f : Z^2 \rightarrow \mathbb{R}$ ,

$$J_n((z, z'); f) := \int f(\Psi\langle y_{1:n} \rangle(z), \Psi\langle y_{1:n} \rangle(z')) \times \prod_{k=1}^n G(\phi(\Psi\langle y_{1:k-1} \rangle(z), \Psi\langle y_{1:k-1} \rangle(z'))); dy_k).$$

On peut maintenant énoncer notre résultat principal d'ergodicité.

**Théorème 5.6.** *Les conditions (AG-3), (AG-4), (AG-5), (AG-6) et (AG-7) impliquent que  $K_g$  admet une distribution stationnaire unique  $\pi_g$  sur  $Z \times Y$ . De plus,  $\pi_{g,1}V < \infty$  pour toute  $\bar{V} : X \rightarrow \mathbb{R}_+$  telle que  $\bar{V} \lesssim V$ .*

Pour construire les fonctions instrumentales  $\alpha$  et  $\phi$  apparaissant dans (AG-7), le lemme suivant, qui est une généralisation du théorème 4.3, est utile.

**Lemme 5.7.** *On suppose que  $\Pi_p(\mathcal{Z}) = \mathcal{X} = \mathcal{C}^{\mathcal{S}}$  pour un espace mesurable  $(\mathcal{S}, \mathcal{S})$  et  $\mathcal{C} \subseteq \mathbb{R}$ . Alors, pour tout  $x \in \mathcal{X}$ , on écrit  $x = (x_s)_{s \in \mathcal{S}}$ , où  $x_s \in \mathcal{C}$  pour tout  $s \in \mathcal{S}$ . On suppose par ailleurs que pour tout  $x = (x_s)_{s \in \mathcal{S}} \in \mathcal{X}$ , on peut exprimer la densité conditionnelle  $g(x; \cdot)$  comme un mélange de densités de la forme  $j(x_s)h(x_s; \cdot)$  sur  $s \in \mathcal{S}$ . Cela signifie que pour tout  $t \in \mathcal{C}$ ,  $y \mapsto j(t)h(t; y)$  est une densité par rapport à  $\nu$ , et qu'il existe une mesure de probabilité  $\mu$  sur  $(\mathcal{S}, \mathcal{S})$  telle que*

$$g(x; y) = \int_{\mathcal{S}} j(x_s)h(x_s; y)\mu(ds), \quad y \in \mathcal{Y}.$$

*On suppose de plus que  $h$  prend des valeurs positives et que l'une des deux hypothèses suivantes se tient :*

(H'-3) Pour tout  $y \in \mathcal{Y}$ , la fonction  $h(\cdot; y) : t \mapsto h(t; y)$  est non-décroissante.

(H'-4) Pour tout  $y \in \mathcal{Y}$ , la fonction  $h(\cdot; y) : t \mapsto h(t; y)$  est non-croissante.

*Pour tout  $(x, x') \in \mathcal{X}^2$ , on note  $x \wedge x' := (\min\{x_s, x'_s\})_{s \in \mathcal{S}}$  et  $x \vee x' := (\max\{x_s, x'_s\})_{s \in \mathcal{S}}$  et on définit  $\alpha(z, z')$  et  $\phi(z, z')$  par : pour tout  $(z, z') \in \mathcal{Z}^2$ , en posant  $x = \Pi_p(z)$  et  $x' = \Pi_p(z')$ ,*

$$\begin{cases} \alpha(z, z') = \inf_{s \in \mathcal{S}} \left\{ \frac{j(x_s \vee x'_s)}{j(x_s \wedge x'_s)} \right\} & \text{et } \phi(z, z') = x \wedge x' & \text{sous (H'-3);} \\ \alpha(z, z') = \inf_{s \in \mathcal{S}} \left\{ \frac{j(x_s \wedge x'_s)}{j(x_s \vee x'_s)} \right\} & \text{et } \phi(z, z') = x \vee x' & \text{sous (H'-4).} \end{cases}$$

*Alors,  $\alpha$  et  $\phi$  définies ci-dessus satisfont (AG-7)(i).*

## 5.5 Exemples

Dans cette section, on applique les résultats principaux établis ci-dessus à trois modèles d'intérêt, à savoir, le modèle GARCH( $p, p$ ) classique, le modèle Poisson log-linéaire GARCH( $p, p$ ) et le modèle NBIN-GARCH( $p, p$ ). Pour le modèle GARCH( $p, p$ ), cependant, on démontre seulement le résultat d'ergodicité; pour le résultat asymptotique de l'EMV pour ce modèle, on se réfère à [Francq et al. \[2004\]](#). À notre connaissance, la stationnarité et l'ergodicité, aussi bien que les propriétés asymptotiques de l'EMV pour les modèles Poisson log-linéaire GARCH( $p, p$ ) et de NBIN-GARCH( $p, p$ ) d'ordre général, n'ont pas été obtenues jusqu'ici. Par la suite,  $\mathbf{A}$ ,  $\mathbf{b}$  et  $\boldsymbol{\omega}$  représentent toujours,

respectivement, la matrice

$$\begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 \\ a_p & a_{p-1} & a_{p-2} & a_{p-3} & \cdots & a_1 & b_p & b_{p-1} & \cdots & b_3 & b_2 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix},$$

le vecteur  $b_1\boldsymbol{\delta}_p + \boldsymbol{\delta}_{2p-1}$ , avec pour tout  $j \in \{1, \dots, 2p-1\}$ ,  $\boldsymbol{\delta}_j = (\delta_{ij})_{i \in \{1, \dots, 2p-1\}}$ ,  $\delta_{ij}$  étant le delta de Kronecker, et le vecteur  $\omega\boldsymbol{\delta}_p$ . On note aussi  $\mathbf{A}_*$ ,  $\mathbf{b}_*$  et  $\boldsymbol{\omega}_*$  les valeurs correspondantes de  $\mathbf{A}$ ,  $\mathbf{b}$  et  $\boldsymbol{\omega}$  au point  $(\omega, a_{1:p}, b_{1:p}) = (\omega^*, a_{1:p}^*, b_{1:p}^*)$ , respectivement. Pour toute matrice  $\mathbf{M}$ , on note  $\mathbf{M}^T$  sa transposée. Pour tous nombres réels  $a$  et  $b$ , on note  $a \vee b$  ou  $\max\{a, b\}$  la valeur maximum, et  $a \wedge b$  ou  $\min\{a, b\}$  la valeur minimum.

## 5.6 Modèle GARCH( $p, p$ )

On considère le modèle GARCH( $p, p$ ),  $p \in \mathbb{Z}_+^*$ , défini comme suit : pour tout  $k \in \mathbb{Z}_+$ ,

$$X_{k+1} = \omega + \sum_{i=1}^p a_i X_{k-i+1} + \sum_{i=1}^p b_i Y_{k-i+1}^2,$$

$$Y_{k+1} | \mathcal{F}_k = X_{k+1}^{1/2} \varepsilon_{k+1},$$

où

- l'espace de paramètres

$$\Theta \subset \{\theta = (\omega, a_{1:p}, b_{1:p}) \in \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^p : \omega > 0, a_{1:p} \geq 0, b_{1:p} \geq 0\},$$

- $\mathcal{F}_k = \sigma(X_{-p+1:k+1}, Y_{-p+1:k})$ ,
- le processus  $\{\varepsilon_k : k \in \mathbb{Z}_+\}$  est une séquence de variables i.i.d. de moyenne zéro et de variance unitaire sur  $\mathbb{R}$ , et sont indépendantes de  $\mathcal{F}_k$ . Pour chaque valeur  $k \in \mathbb{Z}_+$ , la distribution de  $\varepsilon_k$  admet une densité  $\gamma$  par rapport à la mesure de Lebesgue sur  $\mathbb{R}$ . On suppose que la densité  $\gamma$  est positive et continue sur  $\mathbb{R}$ , et satisfait :  $\gamma(u) + \gamma(-u)$  est monotone sur  $u \in \mathbb{R}_+$ . De plus, on suppose que pour tout  $\epsilon > 0$ ,  $\mathbb{P}(\varepsilon_0^2 < \epsilon) > 0$ .

- $\mathbf{X} = (0, \infty)$ ,  $\mathcal{X} = \mathcal{B}(\mathbf{X})$ ,  $\mathbf{Y} = \mathbb{R}$ ,  $\mathcal{Y} = \mathcal{B}(\mathbf{Y})$  et les métriques sur  $\mathbf{X}$  et sur  $\mathbf{Y}$  sont des métriques sur  $\mathbb{R}$ .

Le modèle GARCH( $p, p$ ) est un ODM( $p, p$ ) et le résultat d'ergodicité suivant est établi.

**Proposition 5.8.** *On suppose que  $\theta \in \Theta$  satisfait  $\sum_{i=1}^p (a_i + b_i) < 1$ . Alors, la conclusion du théorème 5.6 se tient avec  $\bar{V}(z) = \mathbf{1}^T z$  pour tout  $z \in \mathbf{Z}$ .*

### 5.6.1 Modèle Poisson log-linéaire GARCH( $p, p$ )

On considère le modèle Poisson log-linéaire GARCH d'ordre ( $p, p$ ),  $p \in \mathbb{Z}_+^*$ , qui est défini comme suit : pour tout  $k \in \mathbb{Z}_+$ ,

$$X_{k+1} = \omega + \sum_{i=1}^p a_i X_{k-i+1} + \sum_{i=1}^p b_i \ln(1 + Y_{k-i+1}),$$

$$Y_{k+1} \mid \mathcal{F}_k \sim \mathcal{P}(e^{X_{k+1}}),$$

où

- l'espace de paramètres

$$\theta = (\omega, a_{1:p}, b_{1:p}) \in \Theta \subset \mathbb{R}^{2p+1},$$

- $\mathcal{F}_k = \sigma(X_{-p+1:k+1}, Y_{-p+1:k})$ ,
- $\mathbf{X} = \mathbb{R}$ ,  $\mathcal{X} = \mathcal{B}(\mathbf{X})$ ,  $\mathbf{Y} = \mathbb{Z}_+$ ,  $\mathcal{Y} = \mathcal{B}(\mathbf{Y})$  et les métriques sur  $\mathbf{X}$  et sur  $\mathbf{Y}$  sont des métriques sur  $\mathbb{R}$ .

On utilise la norme vectorielle et la norme matricielle définies respectivement sur  $\mathbb{R}^{2p-1}$  et sur l'espace de matrices de taille  $(2p-1) \times (2p-1)$  comme suit :

$$|z|_\infty = \bigvee_{i=1}^{2p-1} |\delta_i^T z| = \bigvee_{i=1}^{2p-1} |z_i|,$$

$$\|\mathbf{M}\|_\infty = \bigvee_{i=1}^{2p-1} \sum_{j=1}^{2p-1} |\mathbf{M}_{i,j}| = \sup_{|z|_\infty \leq 1} |\mathbf{M}z|_\infty.$$

Pour simplifier la notation, on note également  $\mathbf{A}_1 = \mathbf{A}$  et  $\mathbf{A}_2 = \mathbf{A} + \mathbf{b}\delta_p^T$ . Le modèle Poisson log-linéaire GARCH d'ordre ( $p, p$ ) est en effet un ODM( $p, p$ ). Pour ce modèle, chacun des résultats suivants se tient.

**Proposition 5.9.** *On suppose que  $\theta \in \Theta$  satisfait*

$$\bigvee_{(i_1, \dots, i_p) \in \{1,2\}^p} \left\| \prod_{\ell=1}^p \mathbf{A}_{i_\ell} \right\|_\infty < 1. \quad (55)$$

Alors, on a :

(i) Le théorème 5.6 se tient avec la fonction  $\bar{V} : \mathbb{Z} \rightarrow \mathbb{R}_+$  définie par, pour tout  $z \in \mathbb{Z}$ ,

$$\bar{V}(z) = e^{\tau|z|^\infty}, \quad \tau > 0.$$

(ii) Le théorème 5.2 se tient avec tout  $z_1 \in \mathbb{Z}$ . Si le vrai paramètre  $\theta_\star = (\omega^\star, a_{1:p}^\star, b_{1:p}^\star)$  satisfait de plus : pour tout  $k \in \mathbb{Z}_+$ ,

$$(\mathbf{A}^k - \mathbf{A}_\star^k)\mathbf{b}_\star = 0 \quad \text{implique} \quad \mathbf{A} = \mathbf{A}_\star, \quad (56)$$

alors, le théorème 5.3 se tient également.

(iii) Si le vrai paramètre  $\theta_\star$  réside à l'intérieur de  $\Theta$  et satisfait (56), alors, le théorème 5.5 se tient.

Notre exemple suivant est le modèle NBIN-GARCH( $p, p$ ),  $p \in \mathbb{Z}_+^*$ , qui est défini comme suit.

### 5.6.2 Modèle NBIN-GARCH( $p, p$ )

On considère un modèle statistique défini récursivement par : pour tout  $k \in \mathbb{Z}_+$ ,

$$X_{k+1} = \omega + \sum_{i=1}^p a_i X_{k-i+1} + \sum_{i=1}^p b_i Y_{k-i+1},$$

$$Y_{k+1} | \mathcal{F}_k \sim \mathcal{NB} \left( r, \frac{X_{k+1}}{1 + X_{k+1}} \right),$$

où

– l'espace de paramètres

$$\Theta \subset \{ \theta = (\omega, a_{1:p}, b_{1:p}, r) \in \mathbb{R}_+^* \times \mathbb{R}_+^p \times \mathbb{R}_+^p \times \mathbb{R}_+^* \},$$

–  $\mathcal{F}_k = \sigma(X_{-p+1:k+1}, Y_{-p+1:k})$ ,

–  $\mathbf{X} = (0, \infty)$ ,  $\mathcal{X} = \mathcal{B}(\mathbf{X})$ ,  $\mathbf{Y} = \mathbb{Z}_+$ ,  $\mathcal{Y} = \mathcal{B}(\mathbf{Y})$  et les métriques sur  $\mathbf{X}$  et sur  $\mathbf{Y}$  sont des métriques sur  $\mathbb{R}$  et  $\mathbb{Z}$ , respectivement.

Le modèle NBIN-GARCH( $p, p$ ) est un ODM( $p, p$ ). Chacun des résultats suivants se tient.

**Proposition 5.10.** *On suppose que  $\theta \in \Theta$  satisfait*

$$|\lambda|_{\max}(\mathbf{A} + r\mathbf{b}\boldsymbol{\delta}_p^T) < 1.$$

(i) Alors, la conclusion du théorème 5.6 se tient avec  $\bar{V}(z) = \mathbf{1}^T z$  pour tout  $z \in \mathbb{Z}$ .

(ii) Le théorème 5.2 se tient avec tout  $z_1 \in \mathcal{Z}$ . De plus, si le vrai paramètre  $\theta_\star = (\omega_\star, a_{1:p}^\star, b_{1:p}^\star)$  satisfait : pour tout  $k \in \mathbb{Z}_+^\star$ ,

$$(\mathbf{A}^k - \mathbf{A}_\star^k)\mathbf{b}_\star = 0 \quad \text{implique} \quad \mathbf{A} = \mathbf{A}_\star, \quad (57)$$

alors, le théorème 5.3 se tient également.

La normalité asymptotique de l'EMV pour le modèle NBIN-GARCH( $p, p$ ) découle de la proposition suivante.

**Proposition 5.11.** *On suppose que  $\theta \in \Theta$  satisfait*

$$|\lambda|_{\max} (\mathbf{A}^T \mathbf{A} + r (\mathbf{A}^T \mathbf{b} \delta_p^T + \delta_p \mathbf{b}^T \mathbf{A}) + r(r+1) \delta_p \mathbf{b}^T \mathbf{b} \delta_p^T) < 1.$$

De plus, on suppose que le vrai paramètre  $\theta_\star$  réside à l'intérieur de  $\Theta$  et satisfait (57). Alors, le théorème 5.5 se tient.

## 5.7 Exemple empirique

Nous donnons une preuve empirique pour des données temporelles spécifiques pour lesquelles des ODM d'ordre plus élevé peuvent être plus appropriés que ceux de premier ordre. Notre ensemble de données, fourni par *U.S. Geology Survey (USGS)*, téléchargeable sur <http://earthquake.usgs.gov/earthquakes/search/>, comprend une série temporelle de nombres de tremblements de terre se produisant dans le monde (par heure), de magnitudes supérieures à 4 sur l'échelle de Richter (un niveau détectable par l'homme). Il consiste en 8 759 observations faites pendant l'année 2014, du 1<sup>er</sup> janvier au 31 décembre. La série temporelle et sa corrélation sérielle sont montrées dans les figures 5.7 et 5.8, respectivement, montrant que les données possèdent une mémoire à long terme. Dans notre analyse, nous utilisons les 8 559 premières observations pour établir l'estimation du paramètre et les 200 dernières observations pour évaluer la performance de prédiction des modèles ajustés. Nous ajustons les données par plusieurs modèles différents, les modèles NBIN-GARCH et de Poisson log-linéaire GARCH avec des ordres différents : (1, 1), (1, 2), (2, 1), (2, 2), (3, 1), (3, 2) et (3, 3). Pour estimer les paramètres, nous utilisons la fonction d'optimisation non-linéaire contrainte `auglag` (*Augmented Lagrangian Minimization Algorithm*) du package `alabama` (*Augmented Lagrangian Adaptive Barrier Minimization Algorithm*) dans R. Les valeurs d'AIC, de BIC et l'erreur de prédiction carrée moyenne ASPE (*average squared prediction error*),  $\text{ASPE} = m^{-1} \sum_{k=1}^m (Y_k - \hat{Y}_k)^2$ , où  $\hat{Y}_k$  est le prédicteur de l'observation  $Y_k$ —et donc le processus de moyennes évalué à l'EMV—, pour ces modèles, sont reportées dans la table 5.2. Qu'on utilise le critère

AIC ou le critère BIC, cette table montre que les modèles d'ordre plus élevé sont préférés pour à la fois le modèle NBIN-GARCH et le modèle de Poisson log-linéaire GARCH, qui s'opposent à leurs homologues de premier ordre. La table 5.2 montre également que les valeurs ASPE des modèles sélectionnés par le critère AIC autant que par le critère BIC, sont plus basses que celles des modèles de premier ordre. Ceci suggère que les modèles d'ordre plus élevé sont plus performants dans les deux cas, celui du modèle NBIN-GARCH et celui du modèle de Poisson log-linéaire GARCH. Cependant, il n'y a pratiquement pas de différences entre ces valeurs pour les modèles sélectionnés par l'AIC et le BIC, respectivement. En les comparant, le modèle de Poisson log-linéaire GARCH sélectionné semble avoir une performance de prédiction aussi bonne que celle du modèle NBIN-GARCH sélectionné, dans nos données. Les figures 5.9 et 5.10, respectivement, représentent les séries de premier ordre prédites et celles choisies par l'AIC et le BIC, comparées aux observations réelles, pour les modèles NBIN-GARCH et de Poisson log-linéaire GARCH.

TABLE 5.2 – Valeurs d'AIC, de BIC et d'ASPE pour les modèles NBIN-GARCH et de Poisson log-linéaire GARCH d'ordres différents : (1, 1), (1, 2), (2, 1), (2, 2), (3, 1), (3, 2) et (3, 3).

	NBIN-GARCH						
	(1, 1)	(1, 2)	(2, 1)	(2, 2)	(3, 1)	(3, 2)	(3, 3)
AIC	30102	30097	30083	30078	30074	30070	<b>30065</b>
BIC	30130	30132	30118	30120	<b>30117</b>	30119	30122
ASPE	1,633	1,633	1,618	1,633	1,614	1,614	1,613
	Poisson log-linéaire GARCH						
	(1, 1)	(1, 2)	(2, 1)	(2, 2)	(3, 1)	(3, 2)	(3, 3)
AIC	30632	30621	30616	30614	30610	30608	<b>30607</b>
BIC	30653	30649	<b>30644,5</b>	30649	30645.4	30651	30656
ASPE	1,654	1,658	1,616	1,616	1,613	1,615	1,617

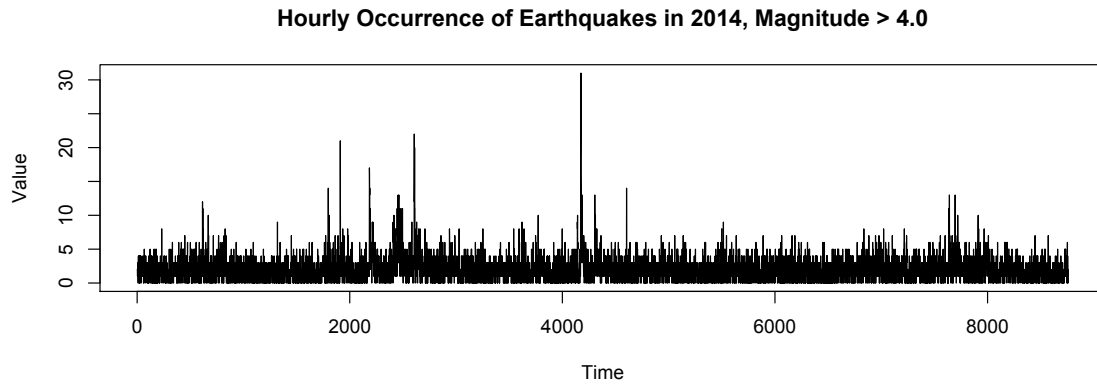


FIGURE 5.7 – Série temporelle de nombres de tremblements de terre en 2014, du 1<sup>er</sup> janvier au 31 décembre (par heure).

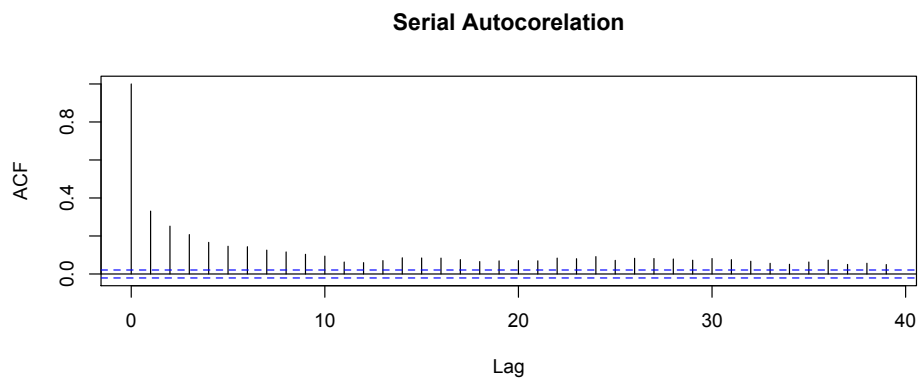


FIGURE 5.8 – Fonction d'autocorrélation.



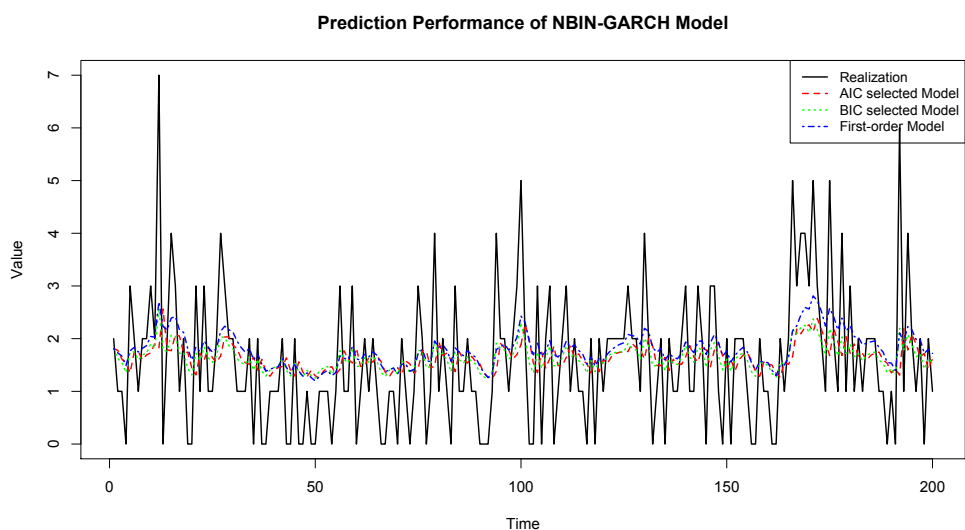


FIGURE 5.9 – Séries estimées par le modèle NBIN-GARCH d'ordre (1, 1) (pointillés bleus), le modèle sélectionné par l'AIC (longs pointillés rouges) et le modèle sélectionné par le BIC (petits pointillés verts), comparées à la vraie série (noire).

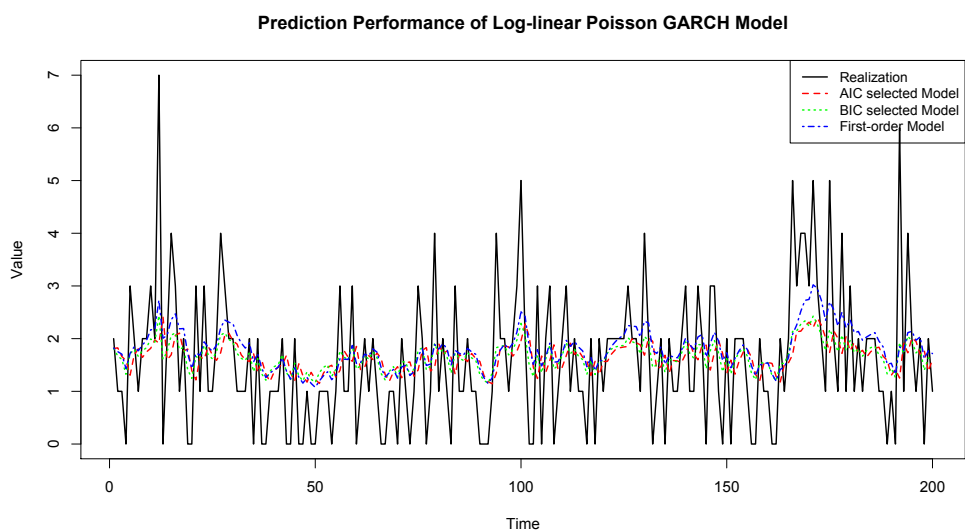


FIGURE 5.10 – Séries estimées par le modèle de Poisson log-linéaire GARCH d'ordre (1, 1) (pointillés bleus), le modèle sélectionné par l'AIC (longs pointillés rouges) et le modèle sélectionné par le BIC (petits pointillés verts), comparées à la vraie série (noire).

## 6 Conclusions et perspectives

Cette thèse se concentre sur le développement de la théorie d'estimation du paramètre pour une classe générale de modèles de séries temporelles, à savoir la classe de chaînes de Markov bivariées partiellement observées. Des modèles de séries temporelles appartenant ou compris dans cette classe qui a des applications potentielles aux séries temporelles de comptage sont particulièrement considérés.

Nous établissons dans des hypothèses assez générales que la limite de la log-vraisemblance normalisée est maximisée quand les paramètres appartiennent à la classe d'équivalence du vrai paramètre, ce qui constitue une étape importante pour l'obtention de la consistance de l'EMV dans le cadre de modèles bien spécifiés. Ce résultat est obtenu dans le cadre général de modèles de Markov partiellement observés et est abordé en utilisant principalement l'unicité de la distribution invariante de la chaîne de Markov associée aux données complètes, sans tenir compte de son taux de convergence vers l'équilibre. En raison de sa généralité, ce résultat peut être appliqué à de nombreuses classes d'intérêt, dont la classe des chaînes de Markov cachées (HMM) et celle des modèles « observation-driven » (ODM), une classe importante pour traiter des séries temporelles de comptage.

La consistance de l'EMV dans le sens de [Leroux \[1992\]](#) est développée indépendamment pour la classe des chaînes de Markov totalement dominées et partiellement observées d'une part, et pour la classe des ODM d'autre part. Le résultat obtenu dans la première classe est dérivé dans des conditions suffisantes assez générales, qui recouvrent beaucoup de modèles intéressants, comme des HMM, des modèles de Markov à changement de régime et certaines versions de chaînes de Markov en environnement aléatoire. Pour les ODM, à la fois de premier ordre et d'ordre général, le résultat est obtenu sous des ensembles d'hypothèses faciles à vérifier dans des exemples spécifiques. Par ailleurs, la consistance forte et la normalité asymptotique de l'EMV sont également traitées pour certains cas spéciaux mais importants d'ODM.

Cette thèse peut être enrichie et mener à de futures recherches dans plusieurs directions.

1. Dans notre preuve d'identification de l'ensemble des maxima de la log-vraisemblance normalisée asymptotique, la structure markovienne des modèles est nécessaire. Ce résultat tiendrait-il toujours dans le cadre de modèles plus généraux que ceux de Markov ?
2. Une approche unifiée pour l'identification de l'ensemble des maxima de la log-vraisemblance normalisée asymptotique pour la classe des modèles de Markov partiellement observés est maintenant obtenue. Ce-

pendant, à notre connaissance, montrer que l'EMV converge pour cette classe est réalisé séparément selon les cas. Un résultat unifié répondant à ce problème pourrait-il être obtenu également pour la classe des modèles de Markov partiellement observés ?

3. La preuve de la normalité asymptotique de l'EMV requiert généralement la consistance forte de l'EMV. Cependant, on ne sait pas encore si la consistance de classe d'équivalence est suffisante pour la preuve. Par conséquent, il serait intéressant de pouvoir montrer que c'est en effet le cas.
4. La preuve de la normalité asymptotique de l'EMV pour les ODM d'ordre général est obtenue sous une forme spécifique de l'équation d'état caché. Ce résultat peut être étendu vers plus de généralité par des approches similaires.

# Bibliography

- Pierre Ailliot and Valérie Monbet. Markov-switching autoregressive models for wind time series. *Environmental Modelling & Software*, 30:92–101, 2012.
- Pierre Ailliot and Françoise Pene. Consistency of the maximum likelihood estimate for non-homogeneous markov-switching models. *arXiv preprint arXiv:1306.2116*, 2013.
- Pierre Ailliot, Julie Bessac, Valérie Monbet, and Françoise Pene. Non-homogeneous hidden markov-switching models for wind time series. *Journal of Statistical Planning and Inference*, 160:75–88, 2015.
- Carol Alexander and Emese Lazar. Normal mixture garch (1, 1): Applications to exchange rate modelling. *Journal of Applied Econometrics*, 21(3): 307–336, 2006.
- Elizabeth S Allman, Catherine Matias, and John A Rhodes. Identifiability of parameters in latent structure models with many observed variables. *Ann. Statist.*, pages 3099–3132, 2009.
- A. Barron. The strong ergodic theorem for densities; generalized Shannon-McMillan-Breiman theorem. *Ann. Probab.*, 13:1292–1303, 1985.
- L. E. Baum and J. A. Eagon. An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology. *Bull. Am. Math. Soc.*, 73:360–363, 1967.
- L. E. Baum and T. P. Petrie. Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Statist.*, 37:1554–1563, 1966.
- L. E. Baum, T. P. Petrie, G. Soules, and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Statist.*, 41(1):164–171, 1970.

## BIBLIOGRAPHY

---

- Leonard E Baum. An equality and associated maximization technique in statistical estimation for probabilistic functions of markov processes. *Inequalities*, 3:1–8, 1972.
- Leonard E Baum and GR Sell. Growth functions for transformations on manifolds. *Pacific J. Math.*, 27(2):211–227, 1968.
- Luc Bauwens, Arie Preminger, and Jeroen VK Rombouts. Theory and inference for a markov switching garch model. *The Econometrics Journal*, 13(2):218–244, 2010.
- Krishnan Bhaskaran, Antonio Gasparrini, Shakoor Hajat, Liam Smeeth, and Ben Armstrong. Time series regression studies in environmental epidemiology. *International journal of epidemiology*, page dyt092, 2013.
- P. J. Bickel, Y. Ritov, and T. Rydén. Asymptotic normality of the maximum likelihood estimator for general hidden Markov models. *Ann. Statist.*, 26:1614–1635, 1998.
- P. J. Bickel, Y. Ritov, and T. Rydén. Hidden Markov model likelihoods and their derivatives behave like i.i.d. ones. *Ann. Inst. H. Poincaré Probab. Statist.*, 38:825–846, 2002.
- T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *J. Econometrics*, 31:307–327, 1986.
- Tim Bollerslev. Glossary to arch (garch). Technical report, CREATES Research Paper, September 2008.
- P. Bougerol and N. Picard. Stationarity of garch processes and of some nonnegative time series. *J. Econometrics*, 52(1992):115 – 127, 1992. ISSN 0304-4076. doi: 10.1016/0304-4076(92)90067-2.
- Richard D Bourgin and Robert Cogburn. On determining absorption probabilities for markov chains in random environments. *Advances in Applied Probability*, pages 369–387, 1981.
- James K Brooks. On the vitali-hahn-saks and nikodým theorems. *Proceedings of the National Academy of Sciences of the United States of America*, 64(2):468, 1969.
- O. Cappé, E. Moulines, and T. Rydén. *Inference in Hidden Markov Models*. Springer, 2005.

## BIBLIOGRAPHY

---

- AA Chernov. Replication of a multicomponent chain by lightning mechanism. *BIOPHYSICS-USSR*, 12(2):336, 1967.
- Vasiliki Christou and Konstantinos Fokianos. Quasi-likelihood inference for negative binomial time series models. *Journal of Time Series Analysis*, 35(1):55–78, 2014.
- Vasiliki Christou and Konstantinos Fokianos. On count time series prediction. *Journal of Statistical Computation and Simulation*, 85(2):357–373, 2015.
- G. Churchill. Hidden Markov chains and the analysis of genome structure. *Computers & Chemistry*, 16(2):107–115, 1992.
- Robert Cogburn. Markov chains in random environments: the case of markovian environments. *The Annals of Probability*, pages 908–916, 1980.
- Robert Cogburn. The ergodic theory of markov chains in random environments. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 66(1):109–128, 1984.
- Robert Cogburn. On direct convergence and periodicity for transition probabilities of markov chains in random environments. *The Annals of probability*, pages 642–654, 1990.
- Robert Cogburn. On the central limit theorem for markov chains in random environments. *The Annals of Probability*, pages 587–604, 1991.
- Francis Comets, Mikael Falconnet, Oleg Loukianov, and Dasha Loukianova. Maximum likelihood estimator consistency for recurrent random walk in a parametric random environment with finite support. *arXiv preprint arXiv:1404.2551*, 2014a.
- Francis Comets, Mikael Falconnet, Oleg Loukianov, Dasha Loukianova, and Catherine Matias. Maximum likelihood estimator consistency for a ballistic random walk in a parametric random environment. *Stochastic Processes and their Applications*, 124(1):268–288, 2014b.
- Fabienne Comte and Offer Lieberman. Asymptotic theory for multivariate garch processes. *Journal of Multivariate Analysis*, 84(1):61–84, 2003.
- DR Cox. Statistical analysis of time-series: some recent developments. *Scand. J. Statist.*, 8(2):93–115, 1981. ISSN 0303-6898.

## BIBLIOGRAPHY

---

- Matthew S Crouse, Robert D Nowak, and Richard G Baraniuk. Wavelet-based statistical signal processing using hidden markov models. *Signal Processing, IEEE Transactions on*, 46(4):886–902, 1998.
- R.A. Davis and H. Liu. Theory and inference for a class of observation-driven models with application to time series of counts. *Preprint, arXiv:1204.3915*, 2012.
- RA Davis, WTM Dunsmuir, and SB Streett. Observation-driven models for Poisson counts. *Biometrika*, 90(4):777–790, DEC 2003. ISSN 0006-3444. doi: {10.1093/biomet/90.4.777}.
- Richard A Davis, WTM Dunsmuir, and Y Wang. Modeling time series of count data. *STATISTICS TEXTBOOKS AND MONOGRAPHS*, 158:63–114, 1999.
- R. Douc and C. Matias. Asymptotics of the maximum likelihood estimator for general hidden Markov models. *Bernoulli*, 7(3):381–420, 2001.
- R. Douc, E. Moulines, and T. Rydén. Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime. *Ann. Statist.*, 32(5):2254–2304, 2004.
- R. Douc, G. Fort, E. Moulines, and P. Priouret. Forgetting the initial distribution for hidden Markov models. *Stochastic Processes and their Applications*, 119(4):1235–1256, 2009.
- R. Douc, E. Moulines, J. Olsson, and R. van Handel. Consistency of the maximum likelihood estimator for general hidden Markov models. *Ann. Statist.*, 39(1):474–513, 2011. ISSN 0090-5364.
- R. Douc, P. Doukhan, and E. Moulines. Ergodicity of observation-driven time series models and consistency of the maximum likelihood estimator. *Stoch. Proc. App.*, 123(7):2620–2647, 2013.
- Randal Douc and Eric Moulines. Asymptotic properties of the maximum likelihood estimation in misspecified hidden markov models. *Ann. Statist.*, 40(5):2697–2732, 2012.
- Randal Douc, Francois Roueff, and Tepmony Sim. The maximizing set of the asymptotic normalized log-likelihood for partially observed markov chains. 2014. URL <https://hal.archives-ouvertes.fr/hal-01080955>.

## BIBLIOGRAPHY

---

- Randal Douc, François Roueff, and Tepmony Sim. Handy sufficient conditions for the convergence of the maximum likelihood estimator in observation-driven models. *Lithuanian Mathematical Journal*, 55(3):367–392, 2015. doi: 10.1007/s10986-015-9286-8.
- Paul Doukhan, Konstantinos Fokianos, and Dag Tjøstheim. On weak dependence conditions for Poisson autoregressions. *Statist. Probab. Lett.*, 82(5): 942–948, 2012. ISSN 0167-7152. doi: {10.1016/j.spl.2012.01.015}.
- Y. Ephraim and N. Merhav. Hidden Markov processes. *IEEE Trans. Inform. Theory*, 48:1518–1569, June 2002.
- Yariv Ephraim and Brian L Mark. Bivariate markov processes and their estimation. *Foundations and Trends in Signal Processing*, 6(1):1–95, 2012.
- René Ferland, Alain Latour, and Driss Oraichi. Integer-valued GARCH process. *J. Time Ser. Anal.*, 27(6):923–942, 2006. ISSN 0143-9782. doi: 10.1111/j.1467-9892.2006.00496.x.
- L. Finesso. *Consistent estimation of the order for Markov and hidden Markov Chains*. PhD thesis, Maryland University, 1991.
- K. Fokianos and D. Tjøstheim. Log-linear poisson autoregression. *J. of Multivariate Analysis*, 102(3):563–578, 2011.
- Konstantinos Fokianos and Dag Tjøstheim. Nonlinear poisson autoregression. *Annals of the Institute of Statistical Mathematics*, 64(6):1205–1225, 2012.
- Konstantinos Fokianos, Anders Rahbek, and Dag Tjøstheim. Poisson autoregression. *J. Am. Statist. Assoc.*, 104(488):1430–1439, 2009. ISSN 0162-1459. doi: 10.1198/jasa.2009.tm08270. URL <http://dx.doi.org/10.1198/jasa.2009.tm08270>. With electronic supplementary materials available online.
- Konstantinos Fokianos, Michael H Neumann, et al. A goodness-of-fit test for poisson count processes. *Electronic Journal of Statistics*, 7:793–819, 2013.
- Wai Mun Fong and Kim Hock See. A markov switching model of the conditional volatility of crude oil futures prices. *Energy Economics*, 24(1):71–95, 2002.
- Christian Francq and Jean-Michel Zakoïan. A tour in the asymptotic theory of garch estimation. *Handbook of Financial Time Series*, pages 85–111, 2009.



## BIBLIOGRAPHY

---

- Christian Francq and Jean-Michel Zakoian. *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons, 2011.
- Christian Francq, Jean-Michel Zakoian, et al. Maximum likelihood estimation of pure garch and arma-garch processes. *Bernoulli*, 10(4):605–637, 2004.
- Jürgen Franke. *Weak dependence of functional INGARCH processes*. Technische Universität Kaiserslautern, Fachbereich Mathematik, 2010.
- D. Fredkin and J. Rice. Correlation functions of a function of a finite-state Markov process with application to channel kinetics. *Math. Biosci.*, 87: 161–172, 1987.
- Sylvia Frühwirth-Schnatter. *Finite mixture and Markov switching models*. Springer Science & Business Media, 2006.
- E. Gassiat, A. Cleynen, and S. Robin. Inference in finite state space non parametric hidden markov models and applications. *Statistics and Computing*, 26(1-2):61–71, 2016. URL <http://dx.doi.org/10.1007/s11222-014-9523-8>.
- V. Genon-Catalot and C. Laredo. Leroux’s method for general hidden Markov models. *Stochastic Process. Appl.*, 116(2):222–243, 2006. ISSN 0304-4149.
- Markus Haas, Stefan Mittnik, and Marc S Paoletta. Mixed normal conditional heteroskedasticity. *Journal of Financial Econometrics*, 2(2):211–250, 2004.
- P. Hall and C. C. Heyde. *Martingale Limit Theory and its Applications*. Academic Press, New York, London, 1980.
- J. D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57:357–384, 1989.
- James D Hamilton. Rational-expectations econometric analysis of changes in regime: An investigation of the term structure of interest rates. *Journal of Economic Dynamics and Control*, 12(2):385–423, 1988.
- James D Hamilton. Specification testing in markov-switching time-series models. *Journal of Econometrics*, 70(1):127–157, 1996.
- James D Hamilton. Regime-switching models. *The new palgrave dictionary of economics*, 2, 2008.

## BIBLIOGRAPHY

---

- Andréas Heinen. Modelling time series count data: an autoregressive conditional poisson model. *Available at SSRN 1117187*, 2003.
- S. G. Henderson, D.S. Matteson, and D.B. Woodard. Stationarity of generalized autoregressive moving average models. *Electronic Journal of Statistics*, 5:800–828, 2011.
- Xuedong D Huang, Yasuo Ariki, and Mervyn A Jack. *Hidden Markov models for speech recognition*, volume 2004. Edinburgh university press Edinburgh, 1990.
- J. Hull and A. White. The pricing of options on assets with stochastic volatilities. *J. Finance*, 42:281–300, 1987.
- AR Ives, B Dennis, KL Cottingham, and SR Carpenter. Estimating community stability and ecological interactions from time-series data. *Ecological monographs*, 73(2):301–330, 2003.
- Søren F. Jarner and Gareth O. Roberts. Polynomial convergence rates of Markov chains. *Ann. Appl. Probab.*, 12(1):224–247, 2002. ISSN 1050-5164.
- J. L. Jensen and N. V. Petersen. Asymptotic normality of the maximum likelihood estimator in state space models. *Ann. Statist.*, 27:514–535, 1999.
- B. Juang and L. Rabiner. Hidden Markov models for speech recognition. *Technometrics*, 33:251–272, 1991.
- Olav Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002. ISBN 0-387-95313-2. doi: 10.1007/978-1-4757-4015-8. URL <http://dx.doi.org/10.1007/978-1-4757-4015-8>.
- Benjamin Kedem and Konstantinos Fokianos. *Regression models for time series analysis*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2002. ISBN 0-471-36355-3. doi: 10.1002/0471266981.
- M. L. Kleptsyna and A. Yu. Veretennikov. On discrete time ergodic filters with wrong initial data. *Probab. Theory Related Fields*, 141(3-4):411–444, 2008. ISSN 0178-8051.
- Vikram Krishnamurthy and Shin-Ho Chung. Signal processing based on hidden markov models for extracting small channel currents. In *Biological Membrane Ion Channels*, pages 623–650. Springer, 2007.

## BIBLIOGRAPHY

---

- A. Krogh, I. S. Mian, and D. Haussler. A hidden Markov model that finds genes in E. coli DNA. *Nucleic Acids Res.*, 22:4768–4778, 1994.
- F. Le Gland and L. Mevel. Exponential forgetting and geometric ergodicity in hidden Markov models. *Math. Control Signals Systems*, 13:63–93, 2000.
- F. Le Gland and L. Mevel. Basic properties of the projective product with application to products of column-allowable nonnegative matrices. *Math. Control Signals Systems*, 13(1):41–62, 2000. ISSN 0932-4194.
- B. G. Leroux. Maximum-likelihood estimation for hidden Markov models. *Stoch. Proc. Appl.*, 40:127–143, 1992.
- Roman Liesenfeld and Jean-Francois Richard. Univariate and multivariate stochastic volatility models: estimation and diagnostics. *Journal of empirical finance*, 10(4):505–531, 2003.
- Alexander M Lindner. Stationarity, mixing, distributional properties and moments of garch (p, q)–processes. In *Handbook of financial time series*, pages 43–69. Springer, 2009.
- Iain L MacDonald and Walter Zucchini. *Hidden Markov and other models for discrete-valued time series*, volume 110. CRC Press, 1997.
- Rogemar S. Mamon and Robert J. Elliott. *Hidden Markov Models in Finance*, volume 104 of *International Series in Operations Research & Management Science*. Springer, Berlin, 2007.
- L. Mevel. *Statistique Asymptotique Pour Les Modèles de Markov Cachés*. PhD thesis, Université de Rennes, 1997.
- L. Mevel and L. Finesso. Asymptotical statistics of misspecified hidden Markov models. *IEEE Trans. Automat. Control*, 49(7):1123–1132, 2004. ISSN 0018-9286. doi: 10.1109/TAC.2004.831156. URL <http://dx.doi.org/10.1109/TAC.2004.831156>.
- S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Springer, London, 1993.
- S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, London, 2009.
- E. Moulines, P. Priouret, and F. Roueff. On recursive estimation for time varying autoregressive processes. *Ann. Statist.*, 33(6):2610–2654, 2005. ISSN 0090-5364. doi: 10.1214/009053605000000624. available at [[arXiv](#)].

## BIBLIOGRAPHY

---

- Kurt Nawrotzki. Finite markov chains in stationary random environments. *The Annals of Probability*, pages 1041–1046, 1982.
- Michael H. Neumann. Absolute regularity and ergodicity of Poisson count processes. *Bernoulli*, 17(4):1268–1284, NOV 2011. ISSN 1350-7265. doi: {10.3150/10-BEJ313}.
- Steven Orey. Markov chains with stochastically stationary transition probabilities. *The Annals of Probability*, pages 907–928, 1991.
- K. R. Parthasarathy. *Probability measures on metric spaces*. AMS Chelsea Publishing, Providence, RI, 2005. ISBN 0-8218-3889-X. Reprint of the 1967 original.
- T. Petrie. Probabilistic functions of finite state Markov chains. *Ann. Math. Statist.*, 40:97–115, 1969.
- J. Pfanzagl. On the measurability and consistency of minimum contrast estimates. *Metrika*, 14:249–272, 1969.
- Wojciech Pieczynski. Pairwise markov chains. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(5):634–639, 2003.
- Wojciech Pieczynski and Abdel-Nasser Tebbache. Pairwise markov random fields and segmentation of textured images. *Machine graphics and vision*, 9(3):705–718, 2000.
- Robert S Pindyck and Daniel L Rubinfeld. *Econometric models and economic forecasts*, volume 4. Irwin/McGraw-Hill Boston, 1998.
- L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*, 77(2):257–285, February 1989.
- Tina Hviid Rydberg and Neil Shephard. Dynamics of trade-by-trade price movements: decomposition and models. *Journal of Financial Econometrics*, 1(1):2–25, 2003.
- S. Streett. *Some observation driven models for time series of counts*. PhD thesis, Colorado State University, Department of Statistics, 2000.
- Allan Timmermann. Moments of markov switching models. *Journal of Econometrics*, 96(1):75–111, 2000.
- Dag Tjøstheim. Some recent theory for autoregressive count time series. *TEST*, 21(3):413–438, 2012. ISSN 1133-0686. doi: {10.1007/s11749-012-0296-0}.

## BIBLIOGRAPHY

---

- Dag Tjøstheim. *Count Time Series with Observation-Driven Autoregressive Parameter Dynamics*. Handbook of Discrete-Valued Time Series. London: Chapman & Hall, 2015.
- P. Tuominen and R. Tweedie. Subgeometric rates of convergence of  $f$ -ergodic Markov Chains. *Advances in Applied Probability*, 26:775–798, 1994.
- Ramon Van Handel. The stability of conditional markov processes and markov chains in random environments. *The Annals of Probability*, pages 1876–1925, 2009.
- Chao Wang, Heng Liu, Jian-Feng Yao, Richard A Davis, and Wai Keung Li. Self-excited threshold poisson autoregression. *Journal of the American Statistical Association*, 109(506):777–787, 2014.
- Dawn Woodard, David Matteson, and Shane Henderson. Stationarity of count-valued and nonlinear time series models. 2010.
- Ryo Yoshida, Seiya Imoto, and Tomoyuki Higuchi. Estimating time-dependent gene networks from time series microarray data by dynamic linear models with markov switching. In *Computational Systems Bioinformatics Conference, 2005. Proceedings. 2005 IEEE*, pages 289–298. IEEE, 2005.
- Scott L Zeger. A regression model for time series of counts. *Biometrika*, 75(4):621–629, 1988.
- Yingjian Zhang. *Prediction of financial time series with Hidden Markov Models*. PhD thesis, Simon Fraser University, 2004.
- Fukang Zhu. A negative binomial integer-valued GARCH model. *J. Time Series Anal.*, 32(1):54–67, 2011. ISSN 0143-9782. doi: 10.1111/j.1467-9892.2010.00684.x. URL <http://dx.doi.org/10.1111/j.1467-9892.2010.00684.x>.

