



HAL
open science

Smoothing in Markov switching linear and gaussian models: Applications to commodity modelization

Thi Ngoc Minh Nguyen

► **To cite this version:**

Thi Ngoc Minh Nguyen. Smoothing in Markov switching linear and gaussian models: Applications to commodity modelization. Statistics [math.ST]. Télécom ParisTech, 2016. English. NNT: 2016ENST0069 . tel-03689917

HAL Id: tel-03689917

<https://pastel.hal.science/tel-03689917>

Submitted on 7 Jun 2022

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 Image»

présentée et soutenue publiquement par

Thi Ngoc Minh NGUYEN

le 23 Novembre 2016

Lissage de modèles linéaires et gaussiens à régimes markoviens Applications à la modélisation de marchés de matières premières

Directeur de thèse: **Éric MOULINES**

Co-encadrement de la thèse: **Olivier HIGELIN**

Jury

M. Frédéric ABERGEL, Professeur, MAS, Ecole Centrale Paris

M. Huyên PHAM, Professeur, LPMA, Université Paris 7

M. François ROUEFF, Professeur, LCTI, Télécom ParisTech

M. Éric MOULINES, Professeur, CMAP, Ecole Polytechnique

M. Sylvain LE CORFF, Chargé de recherches, CNRS, Université Paris Sud

M. Olivier HIGELIN, Docteur, ENGIE

Rapporteur

Rapporteur

Examinateur

Examinateur

Examinateur

Examinateur

TELECOM ParisTech

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

46 rue Barrault 75013 Paris - (+33) 1 45 81 77 77 - www.telecom-paristech.fr

Contents

1	Introduction	9
1.1	Contexte général	9
1.2	Structure du manuscrit	17
1.3	Contribution scientifique	30
2	Two-filter algorithms	33
2.1	Two-filter algorithms	35
2.2	Other Sequential Monte Carlo algorithms to approximate smoothing distributions	40
2.3	Contribution	44
2.3.1	Exponential Deviation Inequalities	44
2.3.2	Asymptotic normality	47
3	Commodity term structure modeling	49
3.1	Commodity Modeling	52
3.2	Contribution	59
3.2.1	Smoothing algorithm	59
3.2.2	Statistical inference applied to crude oil data	62
4	Convergence results for two-filter algorithms	63
4.1	Introduction	63
4.2	The two-filter algorithms	66
4.2.1	Forward filter	67
4.2.2	Backward filter	68

CONTENTS

4.2.3	Two-filter approximations of the marginal smoothing distributions	70
4.2.4	$\mathcal{O}(N)$ approximations of the marginal smoothing distributions	72
4.3	Exponential deviation inequality for the two-filter algorithms	73
4.4	Asymptotic normality of the two-filter algorithms	76
4.5	Proofs	80
4.5.1	Proof of Theorem 1	80
4.5.2	Proof of Theorem 2	81
4.5.3	Proof of Theorem 3	83
4.5.4	Proof of Theorem 4	84
4.5.5	Proof of Theorem 5	89
5	Regime Switching Hidden Markov models	93
5.1	Introduction	94
5.2	Rao Blackwellized smoothing algorithm	97
5.2.1	Forward filter	98
5.2.2	Algorithm based on the structural approximation of [56], [3] .	100
5.2.3	FFBS based algorithm of [67, 85, 65]	101
5.2.4	Rao Blackwellized Two-filter Smoother (RBTF)	105
5.3	Simulated data	111
5.4	Application to CME crude oil (WTI)	116
5.4.1	Model	116
5.4.2	Numerical results	119
5.5	Conclusions	121
A	Appendices	125
A.1	Appendix of Chapter 4	125
A.1.1	Exponential deviation inequalities for the forward filter and the backward information filter	125
A.1.2	Asymptotic normality of the forward filter and the backward information filter	126
A.2	Appendix of Chapter 5	127

A.2.1	Proofs of lemmas 1-3	127
A.2.2	SDE explicit integration	130

CONTENTS

Remerciements

Ma thèse s'est effectuée dans le cadre d'un contrat CIFRE à l'école Télécom Paris-Tech en collaboration avec la société Lunalogic. Cette thèse de doctorat constitue un travail de recherche scientifique, mais également de son application à un domaine concret - la finance. Ma thèse a constitué le fil conducteur d'une tranche de ma vie au crépuscule de la candeur étudiante, et j'espère, à l'aube d'une maturité scientifique. De nombreuses personnes ont croisé mon chemin pendant cette thèse, et ce sont certaines de ces personnes que j'aimerais remercier aujourd'hui.

Je souhaite tout d'abord exprimer ma plus profonde gratitude à mon directeur de thèse, *Eric Moulines* et mon co-encadrant, *Sylvain Le Corff*, pour leur intérêt, leur soutien et leur grande disponibilité durant ma thèse. J'ai vraiment apprécié leurs conseils, leur ouverture et leur patience qu'ils ont pu me donner.

Ce travail n'aurait pas été possible sans le soutien de la société Lunalogic qui m'a accueillie non seulement pour cette thèse, mais aussi lors de mon stage de fin d'études. J'adresse ici mes plus sincères remerciements à la PDG de l'entreprise *Fadila Palmer* ainsi qu'à mes co-encadrants depuis le début de mon stage jusqu'à la fin de ma thèse : *Nicolas Rouveyrollis*, *Olivier Higelin*, *Mathieu Zaradzki*, *Hervé Lambert*. Grâce aux échanges et à leur suivi incontournable, j'ai pu développer l'application de mes connaissances mathématiques aux problématiques réelles du monde de la finance. J'ai beaucoup progressé et appris de nombreuses astuces indispensables pour utiliser les modèles et méthodes numériques de façon efficace sur les vraies données du marché. Je voudrais également remercier mes collègues de Lunalogic que je ne peux pas tous citer dans cette petite page de remerciements pour la serviabilité dont a fait preuve chaque membre, pour leurs conseils et leur

CONTENTS

bienveillance. J'ai partagé de bons moments et des souvenirs inoubliables.

Je tiens à remercier chaleureusement *Frédéric Abergel* et *Huyên Pham* d'avoir accepté de rapporter ce manuscrit ainsi que *François Roueff*, le président du jury, pour leurs remarques pertinentes et leur intérêt porté à mon travail. Plus particulièrement, je voudrais remercier *Huyên Pham* grâce à qui j'ai pu commencer l'étude des mathématiques appliquées à la finance dans le cadre du Master DEA de Laure Elie, Modélisation Aléatoire (M2MO) de Paris 7. Sa présence dans mon jury constitue une grande fierté pour moi.

Je voudrais remercier vivement à *Florence Besnard*, *Laurence Zelmar* et l'ensemble du personnel de l'école doctorale Télécom ParisTech pour leur aide et leur dévouement lors des procédures administratives durant mes trois ans de thèse.

J'adresse mes plus sincères remerciements à mes amis grâce à qui j'ai passé trois merveilleuses années. Leurs encouragements et leur stimulation intellectuelle m'ont particulièrement aidé dans mon travail. Dans la vie quotidienne, je suis toujours heureuse grâce à eux.

Enfin, je souhaite remercier mes grands-parents, mes parents et toute ma famille. Je suis reconnaissante de leurs encouragements et de leur soutien illimité qu'ils m'ont réservés depuis que je suis née. Je les aime énormément !

Chapter 1

Introduction

1.1 Contexte général

Les méthodes de Monte Carlo séquentielles (aussi connues sous le nom de méthodes ou filtres particulières) constituent une famille importante d'algorithmes utilisés pour résoudre des problèmes d'estimation de modèles de Markov cachés. Ces modèles supposent l'existence d'un état caché markovien X_k , $k \geq 0$, observé partiellement, chaque observation Y_k , $k \geq 0$, étant une fonction de l'état caché courant. La présence d'états non observés permet à ces modèles de représenter des relations de dépendances complexes. Depuis leur introduction dans [5], ils ont fait l'objet de nombreuses études, cette popularité s'expliquant par leur nombre important d'applications dans des domaines aussi variés que la reconnaissance automatique de la parole, la poursuite d'objets en environnement complexe, la navigation et la cartographie simultanées, la restauration de signaux ou d'images, ou encore l'identification de systèmes non linéaires.

Les méthodes de Monte Carlo séquentielles permettent de résoudre les problèmes de filtrage et de lissage qui sont au centre de nombreuses problématiques majeures dans le cadre des modèles de Markov cachés. Il s'agit d'estimer la loi de l'état courant (filtrage) ou d'une suite d'états (lissage) conditionnellement à un ensemble d'observations reçues. Nous nous intéressons particulièrement dans cette thèse aux algorithmes fournissant une approximation des lois de deux états con-

sécutifs (X_k, X_{k+1}) conditionnellement à toutes les observations Y_0, \dots, Y_T . Ces lois apparaissent de façon très naturelle lors de l'estimation au sens du maximum de vraisemblance de modèles de Markov cachés par la mise en oeuvre de l'algorithme Expectation Maximization introduit dans [23] ou d'algorithmes de type descente de gradient en utilisant l'identité de Fisher, voir par exemple [14, Chapitre 10]. A l'exception de certains cas simples tels que les modèles de Markov cachés linéaires et Gaussiens ou les modèles à espace d'état fini, les calculs exacts de ces lois de lissage ne sont pas disponibles explicitement. Les méthodes de Monte Carlo séquentielles approchent ces lois à l'aide de variables aléatoires à valeurs dans l'espace d'état, appelées *particules*, associées à des poids d'importance. Elles reposent sur une succession d'étapes de sélection éliminant les particules à faible poids et dupliquant celles ayant un poids élevé et d'étapes de mutation permettant de propager les particules dans l'espace d'état de façon aléatoire. Les premières méthodes particulières par [59] et [41] approchent la suite des lois de filtrage, c'est-à-dire les lois des états X_k conditionnellement aux observations (Y_0, \dots, Y_k) pour chaque valeur de $k \geq 0$. Lorsque toutes les observations ont été reçues, ces algorithmes permettent de façon immédiate d'approcher les lois de lissage jointe, c'est-à-dire les lois des trajectoires (X_0, \dots, X_k) sachant (Y_0, \dots, Y_k) en conservant la généalogie de chaque particule (voir par exemple [20, 21] pour une analyse détaillée). Il s'agit donc d'une première solution pour obtenir une approximation des lois marginales qui nous intéressent (celles de deux états consécutifs (X_k, X_{k+1}) sachant les observations (Y_0, \dots, Y_T)). Cette approche est cependant très peu efficace, en particulier lorsque l'on s'intéresse à des lois de ce type avec k petit devant T . A chaque instant $k \geq 0$, chaque trajectoire de particules entre les instants 0 et $k + 1$ est obtenue en adjoignant une nouvelle particule simulée à une trajectoire passée entre les instants 0 et k choisie aléatoirement parmi les N trajectoires disponibles. A chaque itération, si un chemin ancestral n'est sélectionné par aucune nouvelle particule, toute la trajectoire associée est détruite et n'est plus utile pour l'approximation particulière. Pour de grandes valeurs de T , toutes les trajectoires partagent les mêmes ancêtres et l'approximation de la loi de (X_k, X_{k+1}) sachant les observations $Y_{0:n}$ repose sur une seule trajectoire. Ce problème de dégénéres-

cence des trajectoires a été étudié dans de nombreux travaux, voir par exemple [59, 60, 34]. La dégénérescence due à la répétition des étapes de rééchantillonnage peut-être évitée à l'aide de nombreuses alternatives. Les méthodes de lissage à horizon fixe (*fixed-lag smoothing*) introduites dans [60] et étudiées de façon détaillée dans [76] et [77] localisent le lissage des états (X_k, X_{k+1}) en ne conservant que les observations proches de l'instant k et en ne considérant plus les observations obtenues après un horizon temporel choisi par l'utilisateur. D'autres algorithmes basés sur la décomposition directe-rétrograde (*forward-backward smoothing*) des lois de lissage ont également été proposés: l'algorithme FFBS (*forward filtering backward smoothing*) proposée par [57], [49] et [28] et l'algorithme FFBSi (*forward filtering backward simulation*) introduit par [40]. Ces méthodes reposent sur un premier filtre particulaire direct simulant des particules avec leur poids d'importance pour approcher les lois de filtrage. Toutes ces particules sont gardées en mémoire avec leur poids. Une fois la dernière observation reçue, un parcours arrièrèbasé sur le noyau rétrograde associé à la chaîne de Markov cachée permet de définir une nouvelle approximation des lois de lissage. L'algorithme FFBS conserve toutes les particules engendrées lors du filtre direct mais modifie les poids associés à ces particules. L'algorithme FFBSi simule de façon rétrograde des trajectoires parmi toutes les trajectoires possibles (indépendamment de la généalogie obtenue lors du filtre forward). Une version récursive de l'algorithme FFBS n'utilisant qu'un parcours forward des observations et définissant une approximation d'espérances conditionnelles de fonctionnelles additives appelée PaRis (*particle-based rapid incremental smoother*) a été récemment proposée dans [78].

Dans cette thèse, nous nous concentrons sur les algorithmes de type two-filter introduits dans [9] et [58]. Ces méthodes utilisent deux filtres indépendants: un filtre initialisé à l'instant 0 et évoluant de façon direct et un filtre rétrograde initialisé à l'instant T . Les approximations particulières associées à ces deux récursions sont ensuite recombinaées à chaque pas de temps pour obtenir une approximation des lois de lissage marginales. Le calcul du filtre direct est le même que pour les algorithmes FFBS, FFBSi ou PaRis alors que le filtre rétrograde se calcule en utilisant une suite de distributions instrumentales γ_k , $0 \leq k \leq T$, introduites dans

[10] pour construire une densité de probabilité auxiliaire pouvant être approchée par des méthodes de Monte Carlo séquentielles.

La convergence des méthodes FFBS et FFBSi pour l'approximation des lois de lissage a été étudiée dans [22] et [24]. Ces travaux établissent des inégalités de concentration exponentielles ainsi que la normalité asymptotique des approximations particulières issues de ces algorithmes. Ces résultats ont été étendus à l'approximation d'espérances conditionnelles de fonctionnelles additives des états cachés sachant les observations dans [31] et [78] : des inégalités L^q et des inégalités de concentration exponentielles ont été obtenues avec une dépendance explicite en fonction du nombre d'observations T et du nombre de particules N . À l'inverse, les résultats portant sur la convergence des approximations fournies par les différentes méthodes two-filter restent encore partiels.

Dans [79], les auteurs proposent une estimation de la vraisemblance des observations obtenue à l'aide de l'algorithme two-filter de [10]. Ils établissent un théorème central limite pour cet estimateur non biaisé lorsque le nombre de particules tend vers $+\infty$. Nous nous attachons à obtenir des résultats théoriques portant sur le contrôle de l'approximation des lois de lissage donnée par [10] et [34]. Nous proposons notamment des inégalités de concentration exponentielles non asymptotiques ainsi que des théorèmes central limite faisant explicitement intervenir les lois instrumentales γ_k , $0 \leq k \leq T$. Ce travail permet de compléter les connaissances actuelles sur les méthodes séquentielles de Monte Carlo et en particulier de mettre en avant les mêmes propriétés de convergence pour les méthodes two-filter que pour les méthodes FFBS, FFBSi et PaRis. Ces résultats nous semblent importants puisque la complexité numérique de certaines implémentations de ces méthodes les rend très attractives. D'autre part, la forme particulière de la variance asymptotique de l'estimateur des lois de lissage marginale a un intérêt particulier. Cette variance est la somme de deux contributions correspondant aux deux filtres indépendants et a donc une forme bien plus simple à interpréter que les variances associées aux méthodes FFBS, FFBSi et PaRis. Récemment, [64] ont proposé un algorithme d'estimation de la variance des filtres particuliers basé uniquement sur les particules et les poids simulés pour construire les estimateurs particuliers. Ces

techniques pourraient alors être étendues de façon à obtenir un estimateur *en ligne* de la variance asymptotique de méthodes de lissage two-filter, ce qui n'existe pour aucune autre méthode de lissage.

Dans un second temps, nous proposons dans cette thèse d'utiliser un algorithme de type two-filter pour l'estimation de modèles de Markov cachés à changements de régimes décrivant les marchés de matières premières. Nous nous intéressons particulièrement à la structure par terme des prix des matières premières, c'est-à-dire à l'ensemble des prix d'une matière première fixée, à une date donnée, pour différentes échéances de livraison appelées maturités. Les modèles statistiques introduits pour comprendre la structure par terme ont pour objectif d'interpréter les relations entre le prix au comptant d'un produit financier et différents prix à terme. Cette structure a été formulée pour la première fois dans [54]. Cependant, sa modélisation statistique reste encore un problème largement ouvert, une des difficultés provenant de la présence de régimes sur les marchés. Dans les modèles les plus simples, la structure par terme des matières premières présente deux comportements principaux qu'il s'agit d'identifier: le *backwardation* (à un instant donné, le prix future est une fonction décroissante des différentes maturités) et le *contango* (à un instant donné, le prix future est une fonction croissante en fonction des différentes maturités). Les travaux présentés dans cette thèse proposent une nouvelle approche permettant de modéliser la structure par terme des marchés des commodités en présence de changements de régimes latents. Le modèle que nous avons utilisé permet notamment de prendre en compte des changements aléatoires de dynamiques pouvant s'expliquer par des crises financières, des pénuries liées aux conditions climatiques ou à tout autre facteur aléatoire ayant un impact sur l'évolution du prix au comptant. Cette première partie a pour objectif de répondre en particulier aux questions suivantes :

- Quel modèle adopter pour décrire la structure par terme sur les marchés de matières premières afin d'expliquer les changements de régime sur une série de prix futures à différentes maturités ?
- Quelle méthode peut-on appliquer pour l'inférence du modèle à partir de ces don-

nées observées ?

Les premiers modèles stochastiques permettant de comprendre les marchés de commodités ont été introduits par [8]. Dans ces travaux, l'unique variable d'état permettant d'expliquer les prix à terme est le prix au comptant de la commodité étudiée modélisé par un mouvement Brownien géométrique. Ces modèles ont ensuite été étendus par de nombreux travaux faisant intervenir d'autres variables stochastiques pour expliquer la dynamique du prix au comptant. [39] ont introduit une seconde variable d'état, appelée *convenience yield*, dans la modélisation dynamique du prix au comptant. Ce second facteur spécifique aux marchés de matières premières représente le bénéfice associé à l'achat du produit sous-jacent et non à l'achat d'un contrat future et est donc très fortement lié aux coûts de transport ou de stockage. Dans le modèle de Gibson-Schwartz, le *convenience yield* a une influence sur l'évolution du prix au comptant similaire à celle du dividende sur l'évolution des prix dans les marchés d'actifs financiers. Ce facteur est modélisé par un processus d'Ornstein Uhlenbeck permettant de représenter plus fidèlement les différents comportements de la structure par terme. [87] a également suggéré de prendre en compte le caractère stochastique du taux d'intérêt. Ces modèles à deux ou trois facteurs mettent en évidence les contraintes particulières liées aux marchés des commodités, pour lesquels une vente à livraison différée permet de se protéger contre d'éventuelles catastrophes climatiques ou encore de minimiser des coûts de stockage ou de transport durant certaines périodes. [15] et [68] ont proposé des extensions de ces modèles faisant intervenir le prix au comptant et un terme de volatilité stochastique dans la dynamique du *convenience yield*. Il existe également d'autres approches permettant de modéliser le prix au comptant à l'aide de plusieurs facteurs sans utiliser le *convenience yield*. Dans [88], le logarithme du prix au comptant est décomposé en deux facteurs: un premier terme représentant un prix d'équilibre sur le long terme, l'autre décrivant une déviation de court terme par rapport à cet équilibre. Cette approche a ensuite été généralisée dans les travaux de [70], [69] et [1]. Ils ont été largement utilisés dans des contextes différents ces vingt dernières années.

Nous nous intéressons à l'approche de [39] car la structure par terme (backwar-

ation et contango) des prix futures est correctement expliquée par la présence du processus convenience yield dans la dynamique du prix au comptant. L'évolution du prix au comptant n'étant pas observée, l'inférence du modèle n'est possible qu'à l'aide des prix futures qui sont eux observés sur le marché. La discrétisation en temps conduit à un modèle d'état où la variable d'état à l'instant k est $Z_k = (\ln S_k, \delta_k)$ (où S_k est le prix comptant et δ_k est le convenience yield). En ajoutant l'hypothèse selon laquelle les prix sont observés dans un bruit additif Gaussien, nous obtenons un modèle de Markov caché linéaire et Gaussien où l'observation à un instant donné Y_k est un vecteur dans \mathbb{R}^p formé par les prix futures associés à p maturités différentes. Cette structure permet alors d'estimer le modèle au sens du maximum de vraisemblance, par exemple à l'aide de l'algorithme Expectation Maximization.

Le modèle que nous avons utilisé combine l'approche de [39] avec l'introduction de changements de régimes et donc avec l'évolution au cours du temps de la valeur des paramètres à des instants aléatoires. De manière à décrire plus précisément les différents régimes nous introduisons un processus Markovien additionnel prenant un nombre fini de valeurs. La valeur de ce processus à un instant donné identifie l'état dans lequel se trouve le marché, caractérisé par l'ensemble des valeurs des paramètres du modèle. La discrétisation de ce modèle conduit à un modèle de Markov cachés conditionnellement linéaire et Gaussien dans lequel l'espace d'état est augmenté d'une variable supplémentaire à dynamique markovienne et à espace d'état fini. Les changements entre différents régimes entre deux instants successifs sont contrôlés par la matrice de transition de cette chaîne. La variable d'état à l'instant k est constituée de $Z_k = (\ln S_k, \delta_k)$ et de l'indicateur du régime du marché a_k . L'observation à un instant donné Y_k est toujours un vecteur dans \mathbb{R}^p formé par les prix futures associés à p maturités différentes. Un premier modèle à changements de régimes basé sur le modèle de Gibson-Schwartz a été proposé par [2] pour les marchés de matières premières. L'auteur de cet article suppose que les prix futures sont observés sans erreurs pour certaines dates de livraison afin d'estimer le modèle. Nous avons privilégié une approche entièrement fondée sur un modèle de Markov cachés où le modèle d'observation suppose un bruit additif Gaussien pour toutes les maturités observées. Cette écriture plus souple pose cependant

de nombreuses difficultés lors de l'estimation des paramètres du modèle. En effet, conditionnellement aux valeurs des labels à tous les instants (c'est-à-dire à la connaissance du régime dans lequel se trouve le marché), les variables d'états sont décrites par un modèle linéaire et Gaussien pour lequel des méthodes d'estimation explicites sont disponibles. En présence de J régimes et T observations, le nombre de trajectoires de labels (toutes les suites de valeurs possibles pour définir les régimes traversés par le marché) est J^T . Ainsi, pour des valeurs raisonnables de J et T , les méthodes d'intégration explicites ne peuvent être mises en oeuvre. Les travaux présentés dans cette thèse proposent d'approcher ces étapes explicites par des méthodes séquentielles de Monte Carlo.

Puisque le modèle discrétisé est un modèle de Markov partiellement observé (par les prix futures disponibles sur le marché) nous proposons d'utiliser un algorithme de type Expectation Maximization. Cet algorithme itératif produit des estimations des paramètres en maximisant la vraisemblance des observations. Il nécessite cependant le calcul d'espérances de fonctions des états cachés conditionnellement à toutes les observations. Ces calculs requièrent les lois de toutes les paires d'états cachés (label indiquant le régime, logarithme du prix au comptant, convenience yield) consécutifs conditionnellement à toutes les observations qui ne sont pas calculables explicitement. Dans ce contexte de modèle Markov caché à changements de régimes et pour faire suite à la première partie de ce manuscrit, nous proposons un nouvel algorithme basé sur les méthodes de Monte Carlo séquentielles de type two-filter introduites par [10]. Dans cet article, les auteurs proposent une méthode d'estimation de ces lois d'intérêt dans le cas spécifique de modèles de Markov caché à changements de régimes. Ces filtres utilisent la structure particulière du modèle conditionnellement linéaire et gaussien: à l'aide de l'algorithme inspiré par le filtre de Kalman introduit dans [16], les particules simulées ne correspondent qu'à des suites de labels identifiant les régimes potentiellement traversés par le marché. Conditionnellement aux observations et à chaque suite de labels proposée, les autres états cachés suivent une loi gaussienne dont la variance et l'espérance peuvent être mises à jour directement (il n'y a donc aucun besoin de simuler ces autres états cachés). La densité de probabilité

cible aux instants $k-1$ et k est alors approchée dans [10] en combinant le filtre direct de l'instant $k-1$ avec le filtre rétrograde de l'instant k . L'étape de recombinaison des deux filtres conduit souvent à des lois dégénérées puisque les trajectoires forward et backward recombinaison sont totalement indépendantes. L'extension que nous avons proposée pour approcher la loi de lissage marginale des instants $k-1$ et k se base sur des filtres qui ont été mis à jour jusqu'aux instants $k-2$ pour le filtre forward et $k+2$ pour le filtre backward. La recombinaison des deux filtres s'effectue ensuite en intégrant sur tous les états intermédiaires possibles $z_{k-1}, a_{k-1}, z_k, a_k$ (ce qui est possible par le caractère conditionnellement linéaire et Gaussien du modèle). Cette méthode permet d'avoir une étape de combinaison plus flexible dans l'estimation de loi de lissage comme l'illustrent nos simulations numériques. Une fois l'algorithme testé sur les données simulées, nous l'appliquons aux données réelles du pétrole brut WTI sur le marché CME. Notre estimation des lois postérieures des régimes conditionnellement aux observations illustre le fait que les changements de régimes est approprié pour capter le changement de comportement de la structure par terme des marchés de matières premières.

1.2 Structure du manuscrit

Les résultats obtenus durant la thèse ont fait l'objet d'articles soumis à des revues internationales. Nous insérons ces articles dans les chapitres 4 et 5. Un préambule est associé à chacun de ces articles pour en expliquer le contenu et pour mettre en valeur notre contribution (voir chapitres 2 et 3).

Chapitre 4 On the two-filter approximation of marginal smoothing distribution in general state space models [N.M. Nguyen, S. Le Corff, E. Moulines], préambule: Chapitre 2. Article soumis pour publication dans la revue *Methodology and Computing in Applied Probability*. Dans cet article, nous étudions l'approximation des lois de lissage marginales obtenues par les méthodes de type two-filter proposées par [10] et [34] dans les modèles de Markov cachés généraux. Les approches à deux filtres sont des méthodes de Monte Carlo séquentielles qui com-

binent une passe forward et une passe backward pour approximer les distributions marginales de lissage $\phi_{\chi,t|T}$ pour $t = 0$ à $t = T$ conditionnellement à toutes les observations Y_0, \dots, Y_T . Le but de la passe forward est d'obtenir une approximation récursive des distributions de l'état courant t sachant un ensemble d'observations reçues $\phi_{\chi,t} = \phi_{\chi,t:t|t}$. [41] et [59] ont introduit un algorithme basé sur l'échantillonnage d'importance et les étapes de rééchantillonnage d'importance pour approximer $\phi_{\chi,t}$ en utilisant un ensemble de points, les *particules*, $\{\xi_t^\ell\}_{\ell=1}^N$ associées à des poids $\{\omega_t^\ell\}_{\ell=1}^N$. À $t = 0$, N particules $\{\xi_0^\ell\}_{\ell=1}^N$ sont échantillonnées indépendamment de la densité instrumentale ρ_0 . Alors, ξ_0^ℓ est associée au poids d'importance

$$\omega_0^\ell := \chi(\xi_0^\ell) g_0(\xi_0^\ell) / \rho_0(\xi_0^\ell).$$

Pour toute fonction bornée et mesurable h définie sur X , l'espérance $\phi_{\chi,0}[h]$ est approximée par

$$\phi_{\chi,0}^N[h] := \frac{1}{\Omega_0^N} \sum_{\ell=1}^N \omega_0^\ell h(\xi_0^\ell), \quad \Omega_0^N := \sum_{\ell=1}^N \omega_0^\ell.$$

Puis, en utilisant $\{(\xi_{t-1}^\ell, \omega_{t-1}^\ell)\}_{\ell=1}^N$, les indices et les particules sont échantillonnés à partir de la distribution instrumentale:

$$\pi_{t|t}(\ell, x) \propto \omega_{t-1}^\ell \vartheta_t(\xi_{t-1}^\ell) p_t(\xi_{t-1}^\ell, x)$$

qui est définie sur $\{1, \dots, N\} \times X$, où $\{\vartheta_t(\xi_{t-1}^\ell)\}_{\ell=1}^N$ sont des fonctions d'ajustement aux poids ω_{t-1}^ℓ et p_t une densité de transition choisie par l'utilisateur. Cela signifie que chaque nouvelle particule ξ_t^ℓ au temps t est échantillonnée en deux étapes:

- choisir un indice de particule I_t^ℓ au temps $t-1$ dans $\{1, \dots, N\}$ avec des probabilités proportionnelles à $\omega_{t-1}^j \vartheta_t(\xi_{t-1}^j)$, pour j appartenant à $\{1, \dots, N\}$;
- simuler ξ_t^ℓ en utilisant la densité de transition de la particule choisie: $\xi_t^\ell \sim p_t(\xi_{t-1}^{I_t^\ell}, \cdot)$

Pour $\ell \in \{1, \dots, N\}$, ξ_t^ℓ est associée au poids d'importance suivant

$$\omega_t^\ell := \frac{q(\xi_{t-1}^{I_t^\ell}, \xi_t^\ell) g_t(\xi_t^\ell)}{\vartheta_t(\xi_{t-1}^{I_t^\ell}) p_t(\xi_{t-1}^{I_t^\ell}, \xi_t^\ell)},$$

et l'espérance $\phi_{\chi,t}[h]$ est approchée par

$$\phi_{\chi,t}^N[h] := \frac{1}{\Omega_t^N} \sum_{\ell=1}^N \omega_t^\ell h(\xi_t^\ell), \quad \Omega_t^N := \sum_{\ell=1}^N \omega_t^\ell.$$

La passe backward est basée sur une famille de fonctions mesurables non négatives $\{\gamma_t\}_{t \geq 0}$ définies sur X choisies par l'utilisateur (les hypothèses requises sont données dans le chapitre 4). Ces fonctions sont utilisées pour définir une distribution auxiliaire appelée *backward information filter* qui peut être calculée récursivement en arrière dans le temps. Le filtre backward est défini, pour toute fonction bornée et mesurable h sur X , par

$$\psi_{\gamma,t|T}[h] := \frac{\int \gamma_t(x_t) \left[\prod_{u=t+1}^T g_{u-1}(x_{u-1}) q(x_{u-1}, x_u) \right] g_T(x_T) h(x_t) \nu(dx_{0:T})}{\int \gamma_t(x_t) \left[\prod_{u=t+1}^T g_{u-1}(x_{u-1}) q(x_{u-1}, x_u) \right] g_T(x_T) \nu(dx_{0:T})}.$$

Le filtre backward peut ne pas être calculé explicitement mais être approximé numériquement en utilisant des méthodes séquentielles de type Monte Carlo. À $t = T$, nous échantillons indépendamment N particules $\{\check{\xi}_{T|T}^i\}_{i=1}^N$ à partir de la densité instrumentale $\check{\rho}_T$. Chaque particule $\check{\xi}_{T|T}^i$ est associée au poids d'importance suivant

$$\check{\omega}_{T|T}^i := g_T(\check{\xi}_{T|T}^i) \gamma_T(\check{\xi}_{T|T}^i) / \check{\rho}_T(\check{\xi}_{T|T}^i).$$

$\psi_{\gamma,T|T}[h]$ est ensuite approximé par

$$\psi_{\gamma,T|T}^N[h] := \frac{1}{\check{\Omega}_{T|T}} \sum_{i=1}^N \check{\omega}_{T|T}^i h(\check{\xi}_{T|T}^i), \quad \check{\Omega}_{T|T} = \sum_{i=1}^N \check{\omega}_{T|T}^i.$$

Le filtre backward au temps t est approximé en choisissant un poids d'ajustement $\vartheta_{t|T}$ et une densité instrumentale $r_{t|T}$ comme pour le passage forward. L'ensemble des particules $\{(\check{I}_t^i, \check{\xi}_{t|T}^i)\}_{i=1}^N$ sont simulées par une distribution instrumentale définie par

$$\pi_{t|T}(i, x_t) \propto \frac{\check{\omega}_{t+1|T}^i \vartheta_{t|T}(\check{\xi}_{t+1|T}^i)}{\gamma_{t+1}(\check{\xi}_{t+1|T}^i)} r_{t|T}(\check{\xi}_{t+1|T}^i, x_t).$$

Chaque nouvelle particule $\check{\xi}_{t|T}^i$ est échantillonnée en deux étapes au temps t :

- choisir un indice \check{I}_t^i à $t + 1$ parmi $\{1, \dots, N\}$ avec une probabilité proportionnelle à $\check{\omega}_{t+1|T}^i \vartheta_{t|T}(\check{\xi}_{t+1|T}^i) / \gamma_{t+1}(\check{\xi}_{t+1|T}^i)$;

- simuler $\check{\xi}_{t|T}^i$ par $\check{\xi}_{t|T}^i \sim r_{t|T}(\check{\xi}_{t+1|T}^i, \cdot)$.

Chaque particule parmi $\{\check{\xi}_{t|T}^i\}_{i=1}^N$ est ensuite associée à :

$$\check{\omega}_{t|T}^i := \frac{\gamma_t(\check{\xi}_{t|T}^i) g_t(\check{\xi}_{t|T}^i) q(\check{\xi}_{t|T}^i, \check{\xi}_{t+1|T}^i)}{\vartheta_{t|T}(\check{\xi}_{t+1|T}^i) r_{t|T}(\check{\xi}_{t+1|T}^i, \check{\xi}_{t|T}^i)}.$$

En combinant la définition du filtre forward à l'instant $t - 1$ et le filtre backward à l'instant $t + 1$, la distribution de lissage marginale au temps t peut être exprimée comme suit :

$$\phi_{\chi, t|T}[h] \propto \int \phi_{\chi, t-1}(dx) \psi_{\gamma, t+1|T}(dx') \left[q(x, x_t) g_t(x_t) \frac{q(x_t, x')}{\gamma_{t+1}(x')} \right] h(x_t) \nu(dx_t). \quad (1.1)$$

En utilisant cette combinaison pour la distribution de lissage marginale $\phi_{\chi, t|T}$, on peut utiliser plusieurs algorithmes pour combiner les particules produites dans la passe forward et backward. Cette article se concentre sur les propriétés de convergence des algorithmes à deux filtres dans les modèles d'états généraux. Ces résultats sont prouvés dans l'article présenté dans le chapitre 4. Nous donnons dans cet article des inégalités de déviation exponentielles et des théorèmes de limites centrales pour l'erreur d'approximation. Nous montrons également des contrôles uniformes en temps lorsque des conditions de mélange fort sont satisfaites. Nous obtenons enfin la normalité asymptotique (lorsque le nombre de particules tend vers $+\infty$) pour des estimateurs associés aux différentes méthodes.

Des inégalités de déviation exponentielles ont été établies ces dernières années pour les algorithmes FFBS, FFBSi et PaRis. Les auteurs fournissent des inégalités de déviation exponentielles pour $\phi_{0:T|T}^{N, \text{FFBS}}$ et $\phi_{0:T|T}^{N, \text{FFBSi}}$. Sous des hypothèses légères sur le modèle, il est prouvé dans [24, Théorème 5] et [24, Corollaire 6], qu'il existe des constantes B_T et C_T telles que pour tous $N \geq 1$, $\varepsilon > 0$ et toute fonction mesurable et bornée h définie sur X^{T+1} ,

$$\begin{aligned} \mathbb{P}\left(\left|\phi_{0:T|T}[h] - \phi_{0:T|T}^{N, \text{FFBS}}[h]\right| \geq \varepsilon\right) &\leq B_T \exp\left\{-\frac{C_T N \varepsilon^2}{\text{osc}(h)^2}\right\}, \\ \mathbb{P}\left(\left|\phi_{0:T|T}[h] - \phi_{0:T|T}^{N, \text{FFBSi}}[h]\right| \geq \varepsilon\right) &\leq B_T \exp\left\{-\frac{C_T N \varepsilon^2}{\text{osc}(h)^2}\right\}, \end{aligned}$$

Des résultats similaires ont été obtenus simultanément dans [22]. Dans des conditions de mélange fort, en supposant en particulier que q est uniformément borné, il est prouvé dans [24, Théorème 6] que ces résultats sont uniformes en temps (avec les constantes qui ne dépendent pas de T) pour les distributions de lissage marginales $\phi_{t|T}^{N,FFBS}[h]$ et $\phi_{t|T}^{N,FFBSi}[h]$ pour toute fonction h définie sur X . Des inégalités de déviation similaires ont été démontrées dans [78, Théorème 1 et Corollaire 2] pour la méthode de PaRis $\phi_{\chi,0:T|T}^{N,PaRis}[S_T]$, où S_T est une fonctionnelle additive définie dans (2.2). En utilisant les conditions de mélange fort, [31] ont démontré que l'approximation pour $\phi_{0:T|T}^{N,FFBS}[S_T]$ et $\phi_{0:T|T}^{N,FFBSi}[S_T]$ est bornée en norme L^q . Et pour $q \geq 2$, il est prouvé dans [31, Théorème 1] qu'il existe $C > 0$ tel que pour tout $T < \infty$ et toute suite de fonctions mesurables bornées $\{h_t\}_{t=1}^T$,

$$\begin{aligned} \left\| \phi_{0:T|T}[S_T] - \phi_{0:T|T}^{N,FFBS}[S_T] \right\|_q &\leq C \sqrt{\frac{T}{N}} \left(1 + \sqrt{\frac{T}{N}} \right), \\ \left\| \phi_{0:T|T}[S_T] - \phi_{0:T|T}^{N,FFBSi}[S_T] \right\|_q &\leq C \sqrt{\frac{T}{N}} \left(1 + \sqrt{\frac{T}{N}} \right). \end{aligned}$$

De plus, grâce à [31, Théorème 2], il existe $C > 0$ tel que pour tout $T < \infty$ et toute suite de fonctions mesurables bornées $\{h_t\}_{t=1}^T$,

$$\begin{aligned} \mathbb{P} \left(\left| \phi_{0:T|T}[S_T] - \phi_{0:T|T}^{N,FFBS}[S_T] \right| \geq \varepsilon \right) &\leq B \exp \left\{ -CN\varepsilon^2/T \right\}, \\ \mathbb{P} \left(\left| \phi_{0:T|T}[S_T] - \phi_{0:T|T}^{N,FFBSi}[S_T] \right| \geq \varepsilon \right) &\leq B \exp \left\{ -CN\varepsilon^2/T \right\}. \end{aligned}$$

Tous ces résultats sont cruciaux pour évaluer l'efficacité statistique des méthodes de Monte Carlo séquentielles pour approcher les distributions de lissage, ce qui est une étape clé pour assurer la cohérence et la normalité asymptotique des procédures de déduction de maximum de vraisemblance de Monte Carlo. Cependant, il existe très peu de résultats dans la littérature pour prouver les propriétés de convergence des algorithmes à deux filtres. Les inégalités de déviation suivantes sont prouvées dans le chapitre 4.

Inégalités de déviation pour les méthodes à deux filtres $\text{TwoFilt}_{f_{wt}}$ et $\text{TwoFilt}_{b_{dm}}$: il est démontré dans le Chapitre 4, qu'il existe des constantes $B_{t|T}$ et $C_{t|T} > 0$ telles

que, pour tout $1 \leq t \leq T - 1$ et pour toute fonction h mesurable définie sur X ,

$$\begin{aligned} \mathbb{P} \left[\left| \frac{\sum_{i=1}^N \tilde{\omega}_{t|T}^i h(\tilde{\xi}_{t|T}^i)}{\sum_{i=1}^N \tilde{\omega}_{t|T}^i} - \phi_{\chi, t|T}[h] \right| \geq \epsilon \right] &\leq B_{t|T} \exp \left\{ -C_{t|T} N \epsilon^2 / \text{osc}(h)^2 \right\}, \\ \mathbb{P} \left[\left| \frac{\sum_{i=1}^N \tilde{\omega}_{t|T}^{i,f} h(\xi_t^i)}{\sum_{i=1}^N \tilde{\omega}_{t|T}^{i,f}} - \phi_{\chi, t|T}[h] \right| \geq \epsilon \right] &\leq B_{t|T} \exp \left\{ -C_{t|T} N \epsilon^2 / \text{osc}(h)^2 \right\}, \\ \mathbb{P} \left[\left| \frac{\sum_{i=1}^N \tilde{\omega}_{t|T}^{i,b} h(\check{\xi}_t^i)}{\sum_{i=1}^N \tilde{\omega}_{t|T}^{i,f}} - \phi_{\chi, t|T}[h] \right| \geq \epsilon \right] &\leq B_{t|T} \exp \left\{ -C_{t|T} N \epsilon^2 / \text{osc}(h)^2 \right\}. \end{aligned}$$

Cependant, ces bornes ne sont guère intéressantes dans la pratique puisqu'il n'y a pas de dépendance explicite en T . Pour surmonter ce problème, nous utilisons les conditions de mélange fort, voir 4.3 pour obtenir les inégalités de déviation uniformes en temps. Sous les conditions de mélange fort, similaires à [24], [31], [78], les inégalités de déviation exponentielles pour les approximations par TwoFilt_{fwt} et TwoFilt_{bdm} sont uniformes en temps (avec des constantes ne dépendant pas de T).

Normalité asymptotique : [24] ont établi un théorème de la limite centrale pour les algorithmes FFBS et FFBSi. Sous certaines hypothèses légères sur le modèle, les auteurs prouvent que les quantités $\sqrt{N}(\phi_{0:T|T}[h] - \phi_{0:T|T}^{N, \text{FFBS}}[h])$ et $\sqrt{N}(\phi_{0:T|T}[h] - \phi_{0:T|T}^{N, \text{FFBSi}}[h])$ sont asymptotiquement normales pour une fonction h mesurable définie sur X^{T+1} avec une variance asymptotique explicite quand N tend vers $+\infty$. En ajoutant les conditions de mélange fort, [24, Théorème 12] ont proposé une borne uniforme en temps pour la variance asymptotique des approximations $\phi_{t|T}^{N, \text{FFBS}}[h]$ et $\phi_{t|T}^{N, \text{FFBSi}}[h]$ quelque soit la fonction h mesurable sur X . Quant à la méthode PaRis, [78, Corollaire 5] ont prouvé la normalité asymptotique de la quantité $\sqrt{N}(\phi_{0:T|T}[S_T] - \phi_{0:T|T}^{N, \text{PaRis}}[S_T])$ quand N tend vers $+\infty$ pour les fonctionnelles additives définies par (2.2). Cette variance est représentée comme une somme entre la variance asymptotique de l'estimateur donné par FFBSi et un terme supplémentaire en raison des étapes d'échantillonnage spécifiques de la méthode. Nous fournissons dans le chapitre 4 un nouveau théorème de la limite centrale pour les méthodes à deux filtres en utilisant le théorème pour des échantillons pondérés dans [25]. Notons d'abord que la normalité asymptotique du filtre forward a été établie dans [20]. Pour

tout $0 \leq t \leq T$ et pour toute fonction mesurable h sur X ,

$$\sqrt{N} \left[\sum_{i=1}^N \frac{\omega_t^i}{\Omega_t} h(\xi_t^i) - \phi_{\chi,t}[h] \right] \xrightarrow[N \rightarrow +\infty]{\mathcal{D}} \mathcal{N}(0, \Gamma_{\chi,s}[h - \phi_{\chi,t}[h]]),$$

où $\Gamma_{\chi,s}[h]$ est calculée par une formule récursive qui est détaillée dans [20]. Quant au filtre backward, en suivant les mêmes étapes, nous prouvons le même résultat pour toute fonction mesurable h sur X

$$\sqrt{N} \left[\sum_{j=1}^N \frac{\check{\omega}_{t|T}^j}{\check{\Omega}_{t|T}} h(\check{\xi}_{t|T}^j) - \psi_{\gamma,t|T}[h] \right] \xrightarrow[N \rightarrow +\infty]{\mathcal{D}} \mathcal{N}(0, \check{\Gamma}_{t|T}[h - \psi_{\gamma,t|T}[h]]),$$

où $\check{\Gamma}_{t|T}[h]$ est calculée par une récursion backward en temps. La combinaison de ces deux résultats donne, pour tous $0 \leq s < t \leq T$ et pour toute fonction mesurable h sur X^2 ,

$$\sqrt{N} \left[\sum_{i,j=1}^N \frac{\omega_s^i \check{\omega}_{t|T}^j}{\Omega_s \check{\Omega}_{t|T}} h(\xi_s^i, \check{\xi}_{t|T}^j) - \phi_{\chi,s} \otimes \psi_{\gamma,t|T}[h] \right] \xrightarrow[N \rightarrow +\infty]{\mathcal{D}} \mathcal{N}(0, \tilde{\Gamma}_{s,t|T}[h - \phi_{\chi,s} \otimes \psi_{\gamma,t|T}[h]]),$$

où $\tilde{\Gamma}_{s,t|T}[h]$ est définie par:

$$\tilde{\Gamma}_{s,t|T}[h] := \Gamma_{\chi,s} \left[\int \psi_{\gamma,t|T}(dx_t) h(\cdot, x_t) \right] + \check{\Gamma}_{\gamma,t|T} \left[\int \phi_{\chi,s}(dx_s) h(x_s, \cdot) \right].$$

Ce résultat est crucial car il permet d'obtenir une normalité asymptotique des algorithmes TwoFilt_{fwt} et TwoFilt_{bdm} . D'autre part, sa forme spécifique est telle que toutes les variances asymptotiques des algorithmes à deux filtres reposent uniquement sur les variances asymptotiques du filtre forward et du filtre backward. Ces variances sont beaucoup plus faciles à interpréter que les variances asymptotiques des algorithmes FFBS, FFBSi et PaRis. Ainsi, avec les conditions de mélange fort, nous obtenons une variance asymptotique des méthodes à deux filtres qui est uniforme en temps.

Chapitre 5 A survey of Rao-Blackwellized Sequential Monte Carlo smoothers for Conditionally Linear and Gaussian models [N.M. Nguyen, S. Le Corff, E. Moulines], préambule: Chapitre 3. Article soumis pour publication dans la revue EURASIP Journal on Advances in Signal Processing. Dans cet article, nous étudions un modèle de Markov caché à changement de régimes inspiré par le modèle de Gibson-Schwartz utilisant des observations de prix futures sur les marchés. Les modèles de régime switching appliqués aux marchés des matières premières se concentrent généralement sur une maturité fixe spécifique, mais pas sur la structure par terme des commodités. La modélisation conjointe des séries temporelles futures de plusieurs échéances est complexe car elle nécessite de tenir compte de la dépendance statistique entre les prix futures de différentes échéances. [2] a proposé une dynamique basée sur Gibson-Schwartz avec des paramètres pilotés par un processus de Markov à deux états $\{a_t\}_{t \geq 0}$ sous la probabilité risque-neutre \mathbb{Q} comme suit:

Les modèles de régime switching appliqués aux marchés des matières premières se concentrent généralement sur une maturité fixe spécifique, mais pas sur la structure par terme des commodités. La modélisation conjointe des séries temporelles futures de plusieurs échéances est complexe car elle nécessite de tenir compte de la dépendance statistique entre les prix futures de différentes échéances. [2] a proposé une dynamique basée sur Gibson-Schwartz avec des paramètres pilotés par un processus de Markov à deux états $\{a_t\}_{t \geq 0}$ sous la probabilité risque-neutre \mathbb{Q} comme suit:

$$\begin{aligned} dS_t &= (r - \delta_t)S_t dt + \sigma_{a_t} S_t dW_t^1, \\ d\delta_t &= \kappa(\alpha_{a_t} - \delta_t)dt + \eta_{a_t} dW_t^2, \quad d\langle W_t^1, W_t^2 \rangle = \rho_{a_t} dt. \end{aligned}$$

L'estimation statistique est effectuée en utilisant des observations du marché et ne peut donc être effectuée directement sous la probabilité risque-neutre. Supposons que chaque facteur (spot, convenience yield) a sa propre prime de risque de marché, $\lambda_{a_t}^S$ pour S_t et $\lambda_{a_t}^\delta$ pour δ_t . La dynamique historique du modèle de Gibson-Schwartz est alors décrite comme suit:

$$\begin{aligned} dS_t &= (\mu_{a_t} - \delta_t)S_t dt + \sigma_{a_t} S_t d\widetilde{W}_t^1, \\ d\delta_t &= \kappa(\widetilde{\alpha}_{a_t} - \delta_t)dt + \eta_{a_t} d\widetilde{W}_t^2, \quad d\langle \widetilde{W}_t^1, \widetilde{W}_t^2 \rangle = \rho_{a_t} dt, \end{aligned}$$

où $((\widetilde{W}_t^1, \widetilde{W}_t^2), t \geq 0)$ sont des mouvements browniens sous la probabilité historique \mathbb{P} avec une corrélation ρ_{a_t} , $\mu_{a_t} := r - \lambda_{a_t}^S$ et $\widetilde{\alpha}_{a_t} := \alpha_{a_t} - \lambda_{a_t}^\delta / \kappa$. Dans ce modèle, le convenience yield est censé revenir à sa valeur d'équilibre $\widetilde{\alpha}_{a_t}$ à la vitesse κ . Basé sur la technique de [83], l'auteur a supposé que les prix futures de deux échéances sont observés sans erreur pour filtrer les variables d'état associées à ces prix et pour estimer les paramètres du modèle. Le modèle que nous avons utilisé pour

analyser la structure par terme des matières premières est étroitement lié à celui en [2]. Sous la probabilité historique \mathbb{P} notre modèle est donné par :

$$\begin{aligned} dS_t &= (\mu - \delta_t)S_t dt + \sigma_{a_t} S_t d\widetilde{W}_t^1, \\ d\delta_t &= \kappa(\widetilde{\alpha}_{a_t} - \delta_t)dt + \eta_{a_t} d\widetilde{W}_t^2, \quad d\langle \widetilde{W}_t^1, \widetilde{W}_t^2 \rangle = \rho_{a_t} dt, \end{aligned}$$

où $((\widetilde{W}_t^1, \widetilde{W}_t^2), t \geq 0)$ sont des mouvements brownien sous la probabilité historique \mathbb{P} . Contrairement à l'approche de [2], tous les prix futures pour toutes les maturités sont observés avec des erreurs gaussiennes. La discrétisation temporelle de ce modèle avec $Z_k = (\ln S_k, \delta_k)$ et les vecteurs de prix futures bruités $(Y_k)_{k \geq 0}$ conduit au modèle suivant (voir section 5.3 pour une expression détaillée de chaque paramètre) :

$$Z_k = d_{a_{k-1}} + T_{a_{k-1}} Z_{k-1} + H_{a_{k-1}} \varepsilon_k, \quad (1.2)$$

$$Y_k = c_{a_k} + B_{a_k} Z_k + G_{a_k} \eta_k, \quad (1.3)$$

où

- $(\varepsilon_k)_{k \geq 0}$ est une séquence de variables gaussiennes indépendantes et identiques (i.i.d.).
- $(a_k)_{k \geq 0}$ est une chaîne de Markov homogène qui prend des valeurs dans $\{1, \dots, J\}$. Notons Q la matrice de transition de cette chaîne de Markov; π sa distribution initiale sous la probabilité de risque neutre \mathbb{Q} .
- $(H_j)_{1 \leq j \leq J}$ sont des matrices définies positives.
- Z_1 est un vecteur aléatoire gaussien de taille m de moyenne μ_1 et de variance Σ_1 indépendant de $(\varepsilon_k)_{k \geq 0}$.
- $(\eta_k)_{k \geq 0}$ est une séquence de variables gaussiennes i.i.d., indépendante de $(\varepsilon_k)_{k \geq 0}$ et Z_1 .
- $(G_j)_{1 \leq j \leq J}$ sont des matrices définies positives de taille $p \times p$ et $(B_j)_{1 \leq j \leq J}$ sont des matrices de tailles $m \times p$.

Nous définissons

$$\bar{G}_j := G_j G_j', \bar{H}_j := H_j H_j'$$

et notons les paramètres inconnus θ , donné par

$$\theta := \left\{ Q, \mu_1, \Sigma_1, (d_j, c_j, T_j, B_j, \bar{H}_j, \bar{G}_j)_{1 \leq j \leq J} \right\}.$$

Notre objectif est d'estimer θ en utilisant la séquence d'observations bruitées des prix futures Y_k , $1 \leq k \leq n$. Ensuite, une fois que le modèle est calibré, il peut être utilisé pour prédire les prix futures ou pour détecter les changements de régime.

Dans cet article, nous présentons nos principales contributions pour aborder les problèmes de modélisation, d'inférence et de prévision dans les marchés des produits de commodité. Nous développons l'algorithme à deux filtres basé sur le changement de régime proposé dans [10]. Dans notre extension, nous proposons une combinaison plus flexible entre les filtres forward et backward pour rapprocher la distribution postérieure de deux états consécutifs $(a_{k-1}, Z_{k-1}, a_k, Z_k)$ sachant les observations. La performance de cette méthode a été numériquement comparée à l'algorithme de [10], voir plus de détails dans 5.3. Nous avons appliqué notre algorithme à deux filtres pour approcher l'étape E d'un algorithme EM (Expectation Maximization) afin d'estimer les paramètres du modèle (3.1) et (3.2) pour les données du pétrole brut de CME. Ensuite, l'étape M de l'algorithme EM est approchée à l'aide de la méthode d'optimisation Covariance Matrix Adaptation Evolution Strategy introduite dans [47] (voir également [46] et [45]), voir dans la section 5.4. Par ailleurs, récemment, [66] a proposé une approche forward backward pour approcher ces distributions de lissage sans introduire de distributions artificielles. Ils ont également profité de la sous-structure du modèle pour introduire une nouvelle couche Rao-Blackwellized plus lisse. Bien qu'il ne repose pas sur des distributions artificielles, leur méthode n'exige aucune approximation du modèle.

Algorithme de lissage : Nous proposons une extension des méthodes de type two-filter en intégrant explicitement certains états intermédiaires entre les filtres forward et backward. Cette étape est possible par la structure conditionnellement

linéaire et gaussienne de notre modèle et permet de combiner les trajectoires forward et backward de façon plus souple. Estimer les paramètres des modèles à changement de régime en se basant uniquement sur la séquence des observations (Y_1, Y_2, \dots, Y_n) est un problème statistique difficile. Notre algorithme repose sur l'algorithme Expectation Maximization (EM) qui produit itérativement des estimations de paramètres qui augmentent la vraisemblance des observations. Comme expliqué dans le Chapitre 2, le calcul de la quantité intermédiaire $\theta \mapsto Q(\theta, \theta_p)$ de l'algorithme EM pour un paramètre courant θ_p nécessite les distributions conditionnelles $\phi_i^n(a_{i-1:i}, z_{i-1:i})$ des états $(a_{i-1}, a_i, z_{i-1}, z_i)$ dans (3.1) et (3.2) sachant des observations (Y_1, \dots, Y_n) pour tout $1 \leq i \leq n$.

Supposons que $m(a_{i-1}, z_{i-1}; z_i)$ est la densité de la distribution conditionnelle de Z_i sachant (a_{i-1}, Z_{i-1}) et que $g(a_i, z_i; y_i)$ est la densité de la distribution conditionnelle de Y_i sachant (a_i, Z_i) lorsque la valeur du paramètre est θ_p (θ_p est supprimé dans ces notations pour une meilleure clarté). Conditionnellement à la séquence $(a_i)_{1 \leq i \leq n}$, (3.1) et (3.2) définissent un modèle linéaire et gaussien standard. En utilisant cette structure spécifique, [10] ont proposé d'utiliser Rao-Blackwellization combiné avec un algorithme à deux filtres pour fournir une procédure séquentielle de Monte-Carlo pour approximer $\phi_i^n(a_{i-1:i}, z_{i-1:i})$. Les auteurs ont introduit une famille de densités de probabilité artificielle $(\gamma_i)_{1 \leq i \leq n}$ de sorte que pour tout $1 \leq i \leq n$,

$$\prod_{k=i+1}^n Q(a_{k-1}, a_k) p(y_{i:n} | a_{i:n}, z_i) \propto \gamma_i^{-1}(a_i, z_i) \tilde{p}(a_{i:n}, z_i | y_{i:n}),$$

où $\tilde{p}(a_{i:n}, z_i | y_{i:n})$ est une densité artificielle sur $\{1, \dots, J\}^{n-i+1} \times \mathbb{R}^m$ bien que $p(y_{i:n} | a_{i:n}, z_i)$ ne le soit pas. Ensuite, $\phi_i^n(a_{i-1:i}, z_{i-1:i})$ peut être écrit comme :

$$\begin{aligned} \phi_i^n(a_{i-1:i}, z_{i-1:i}) &\propto \gamma_i^{-1}(a_i, z_i) Q(a_{i-1}, a_i) m(a_{i-1}, z_{i-1}; z_i) \\ &\quad \times p(a_{i-1}, z_{i-1} | y_{1:i-1}) \tilde{p}(a_i, z_i | y_{i:n}). \end{aligned} \quad (1.4)$$

L'algorithme de [10] se déroule comme suit :

- (i) Un filtre forward se rapproche de $p(a_{i-1} | y_{1:i-1})$ pour tout $i \in \{2, \dots, n\}$ avec des particules pondérées $\{(a_{i-1}^k, \omega_{i-1}^k)\}_{k=1}^N$ en utilisant la technique d'un mélange

de filtre de Kalman introduit dans [16]. Conditionnellement à $\{(a_{i-1}^k, \omega_{i-1}^k)\}_{k=1}^N$ et à $y_{1:i-1}$, Z_{i-1} a une distribution gaussienne de moyenne μ_i^k et de matrice de covariance P_i^k qui peuvent être mises à jour à la volée. Ensuite, la densité conditionnelle $p(a_{i-1}, z_{i-1} | y_{1:i-1})$ à l'instant $i - 1$ est approximée par :

$$p^N(a_{i-1}, z_{i-1} | y_{1:i-1}) := \sum_{k=1}^N \omega_{i-1}^k \delta_{a_{i-1}^k}(a_{i-1}) \varphi_{\mu_i^k, P_i^k}(z_{i-1}),$$

où $\varphi_{\mu_i^k, P_i^k}$ est la densité d'une distribution gaussienne de moyenne μ_i^k et de matrice de covariance P_i^k .

- (ii) Un filtre backward produit des échantillons pondérés $\{(\tilde{a}_{i:n}^\ell, \tilde{\omega}_i^\ell)\}_{k=1}^N$ pour tout $i \in \{1, \dots, n\}$ afin d'approximer $\tilde{p}(a_i, z_i | y_{i:n})$ en utilisant la technique d'un mélange de filtre de Kalman.

Dans [10], l'approximation de la distribution de lissage combine directement les deux filtres à particules dans (3.3). Si ces probabilités ont leurs régions de masses de haute probabilité significativement disjointes, on ne peut s'attendre à ce que l'approximation résultante soit bonne, car l'approximation de la répartition du lissage est telle que la plupart des particules ont des poids négligeables. Pour surmonter cette difficulté, nous avons introduit la marginalisation suivante des distributions de lissage bivariées :

$$\phi_i^n(a_{i-1:i}, z_{i-1:i}) = \sum_{a_{i-2}} \sum_{a_{i+1}} \int_{z_{i-1}} \int_{z_{i+1}} \psi_i^n(a_{i-2:i+1}, z_{i-2:i+1}) dz_{i-2} dz_{i+1}, \quad (1.5)$$

où $\psi_i^n(a_{i-2:i+1}, z_{i-2:i+1})$ est la distribution conditionnelle des états et des régimes cachés de $i - 2$ à $i + 1$ sachant toutes les observations. Notons

$$r_i(a_{i-1}, z_{i-1}, y_{1:i-2}) = \sum_{a_{i-2}} \int_{z_{i-2}} p(a_{i-2}, z_{i-2} | y_{1:i-2}) m(a_{i-2}, z_{i-2}; z_{i-1}) Q(a_{i-2}, a_{i-1}) dz_{i-2}$$

et

$$t_i(a_i, z_i, y_{i:n}) = \sum_{a_{i+1}} \int_{z_{i+1}} m(a_i, z_i; z_{i+1}) Q(a_i, a_{i+1}) g(a_i, z_i; y_i) p(y_{i+1:n} | a_{i+1}, z_{i+1}) dz_{i+1}.$$

En utilisant (3.4), $\phi_i^n(a_{i-1:i}, z_{i-1:i})$ devient :

$$\phi_i^n(a_{i-1:i}, z_{i-1:i}) \propto r_i(a_{i-1}, z_{i-1}, y_{1:i-2}) [m(a_{i-1}, z_{i-1}; z_i) Q(a_{i-1}, a_i) g(a_{i-1}, z_{i-1}; y_{i-1})] \times t_i(a_i, z_i, y_{i:n}). \quad (1.6)$$

Nous proposons d'approcher les distributions de lissage marginales en fusionnant les trajectoires forward et backward correspondant aux états cachés jusqu'à $i - 2$ pour le filtre forward et $i + 1$ pour le filtre backward. Ensuite, le calcul de r_{i-1} et t_i nécessite d'intégrer tous les chemins possibles entre $i - 2$ et $i - 1$ et entre i et $i + 1$, respectivement. Cette intégration est calculée explicitement dans notre cas linéaire et gaussien et se révèle produire une approximation plus robuste que l'algorithme de [10]. Suivant [10] notre algorithme est décomposé en un passage forward pour se rapprocher de r_i et un passage backward pour se rapprocher de t_i en utilisant la technique de deux filtres. Les deux chemins sont basés sur des méthodes de Monte Carlo séquentielles : le passage forward fournit une séquence d'états $a_{1:i-2}^k$ associés à des poids d'importance $\omega_{i-2}^k, 1 \leq k \leq N$ pour estimer $p(a_{i-2}|y_{1:i-2})$. Cela fournit une approximation de r_i qui remplace l'étape (i) de [10]. Un autre filtre à particules est utilisé dans le passage backward en utilisant des distributions artificielles pour produire une séquence de trajectoires arrières $\tilde{a}_{i+1:n}^k$ associés à des poids d'importance $\tilde{\omega}_{i-2}^k, 1 \leq k \leq N$ pour estimer $p(y_{i+1:n}|a_{i+1}, z_{i+1})$. Cela fournit une estimation de t_i qui remplace l'étape (ii) de [10]. Ces approximations sont ensuite insérées dans (3.5) pour obtenir une approximation basée sur les particules de $\phi_i^n(a_{i-1:i}, z_{i-1:i})$. La performance des deux méthodes de filtrage dépend fortement du choix des distributions artificielles. Nous avons proposé une distribution artificielle définie comme un mélange de distributions gaussiennes basées sur les particules produites lors du passage forward.

Application aux données du pétrole brut : L'algorithme présenté dans la section précédente a été appliqué pour estimer les paramètres du régime en utilisant les modèles cachés de Markov décrits par (3.1) et (3.1) à l'aide d'un algorithme EM avec des données du pétrole brut. Dans [2], l'auteur a supposé que deux de ces séries chronologiques futures sont observées avec des erreurs et les autres

sans. Ces hypothèses créent des difficultés pour choisir des prix bruités ou non bruités et implique également des résultats dépendant de ce choix. Contrairement à [2], nous avons considéré que des observations bruitées sont disponibles pour tous les contrats futures. La performance de notre algorithme a été comparée à celle d'une méthode two-filter traditionnelle en utilisant ces données réelles. Notre modèle de regime switching est approprié pour comprendre les changements de comportement de structure par terme du marché. De plus, notre algorithme est aussi capable de détecter ces changements. Les illustrations de la convergence de la procédure d'estimation sont données dans le chapitre 5.

1.3 Contribution scientifique

Articles dans les revues internationales

1. On the two-filter approximation of marginal smoothing distribution in general state space models, [N.M. Nguyen, S. Le Corff, E. Moulines]. *Article soumis pour publication dans la revue Methodology and Computing in Applied Probability.*
2. A survey of Rao-Blackwellized Sequential Monte Carlo smoothers for Conditionally Linear and Gaussian models, [N.M. Nguyen, S. Le Corff, E. Moulines]. *Article soumis pour publication dans la revue EURASIP Journal on Advances in Signal Processing.*

Communication dans les conférences internationales

1. Inference in Regime Switching models for Commodity Term Structure with applications to crude oil market, [N.M. Nguyen, S. Le Corff, E. Moulines]. *9th World Congress of The Bachelier Financial Society, New York, July 15-19, 2016.*
2. Regime switching Hidden Markov Models applied to Commodity Term Structure, [N.M. Nguyen, S. Le Corff, E. Moulines]. *Analysis, Probability and their*

Applications, Quy Nhon, Vietnam, December 11-17, 2016.

Chapter 2

Two-filter algorithms

Let (X, \mathcal{X}) and (Y, \mathcal{Y}) be two general spaces endowed with countably generated σ -fields and consider $\{(X_t, Y_t)\}_{t \geq 0}$ a discrete-time process defined on $X \times Y$ such that:

- $\{X_t\}_{t \geq 0}$ is a Markov chain with Markov transition kernel Q . It is assumed that there exists a given reference measure ν on (X, \mathcal{X}) such that the distribution of X_0 has probability density χ with respect to ν and for all $x \in X$, $Q(x, \cdot)$ has probability density $q(x, \cdot)$ with respect to ν ;
- conditional on $\{X_t\}_{t \geq 0}$, $\{Y_t\}_{t \geq 0}$ are independent. In addition, for all $\ell \geq 0$, the conditional distribution of Y_ℓ given $\{X_t\}_{t \geq 0}$ is defined, for any bounded measurable function h on (Y, \mathcal{Y}) , by:

$$\mathbb{E}[h(Y_\ell) | \{X_t\}_{t \geq 0}] = \int h(y) g(X_\ell, y) \lambda(dy),$$

where λ is a given reference measure on (Y, \mathcal{Y}) . In the following, as we always condition upon the same set of observations, we write $g(x, Y_\ell) = g_\ell(x)$.

It is assumed that the states $\{X_t\}_{t \geq 0}$ are not observed so that only the observations $\{Y_t\}_{t \geq 0}$ may be used to design statistical inference procedures. Fixed interval smoothing distributions are defined, for any bounded measurable function h on X^{t-s+1} , any $T \geq 0$ and $0 \leq s \leq t \leq T$, by:

$$\phi_{\chi, s:t|T}[h] := \frac{\int \chi(x_0) g_0(x_0) \prod_{u=1}^T q(x_{u-1}, x_u) g_u(x_u) h(x_{s:t}) \nu(dx_{0:T})}{\int \chi(x_0) g_0(x_0) \prod_{u=1}^T q(x_{u-1}, x_u) g_u(x_u) \nu(dx_{0:T})}.$$

CHAPTER 2. TWO-FILTER ALGORITHMS

In many cases, Q and the functions $\{g_t\}_{t \geq 0}$ depend on a parameter θ . In this case, we write $\chi = \chi^\theta$, $Q = Q^\theta$, $q = q^\theta$, $g_t = g_t^\theta$ for all $t \geq 0$, and

$$\phi_{\chi, s:t|T}^\theta[h] := \frac{\int \chi(x_0) g_0^\theta(x_0) \prod_{u=1}^T q^\theta(x_{u-1}, x_u) g_u^\theta(x_u) h(x_{s:t}) \nu(dx_{0:T})}{\int \chi(x_0) g_0^\theta(x_0) \prod_{u=1}^T q^\theta(x_{u-1}, x_u) g_u^\theta(x_u) \nu(dx_{0:T})}.$$

Maximum likelihood parameter estimates of θ may be then obtained for instance by applying the EM algorithm introduced in [23]. Given a current parameter estimate θ_p , the pivotal idea of this algorithm is to replace the loglikelihood of the observations by the surrogate quantity:

$$\theta \mapsto Q(\theta, \theta_p) = \mathbb{E}_{\theta_p} [\log p_\theta(X_{0:T}, Y_{0:T}) | Y_{0:T}],$$

where $\mathbb{E}_{\theta_p} [\cdot | Y_{0:T}]$ denotes the expectation conditional on $Y_{0:T}$ when the parameter is θ_p and p_θ is the complete likelihood of the states and observations:

$$p_\theta(x_{0:T}, Y_{0:T}) = \chi^\theta(x_0) g_0^\theta(x_0) \prod_{u=1}^T q^\theta(x_{u-1}, x_u) g_u^\theta(x_u).$$

The new parameter estimate is then obtained following the two steps:

- (i) compute $\theta \mapsto Q(\theta, \theta_p)$;
- (ii) set θ_{p+1} as one of the maximizers of $\theta \mapsto Q(\theta, \theta_p)$.

Forgetting the initial state and observation, the intermediate quantity may be written as:

$$Q(\theta, \theta_p) = \sum_{t=1}^T \mathbb{E}_{\theta_p} \left[\log \left\{ q^\theta(X_{t-1}, X_t) g_t^\theta(X_t) \right\} \middle| Y_{0:T} \right].$$

This algorithm is used in Chapter 5 to estimate Markov switching linear and Gaussian models applied to commodity markets term structure. Step (i) requires to compute at each iteration smoothed expectations of all couples of consecutive states (X_{t-1}, X_t) from $t = 1$ to $t = T$. Nevertheless, the conditional distributions $\phi_{\chi, t-1:t|T}^{\theta_p}$ is available explicitly only in very few models such as linear and Gaussian hidden Markov models or models with finite state space.

In this chapter, we are interested in a particular class of Sequential Monte Carlo methods known as two-filter algorithms introduced in [10] et [34] and used to approximate $\phi_{\chi, t-1:t|T}^{\theta_p}$ for $t = 1$ to $t = T$. Chapter 4 displays several new results on the control of these approximations such as nonasymptotic exponential deviation inequalities and central limit theorems. These controls have several practical interests. First, nonasymptotic deviation inequalities with explicit bounds depending on the number of observations T and the number of particles (Monte Carlo samples) N allow to tune N as a function of T to ensure a given accuracy. This is for instance fundamental to prove consistency results of maximum likelihood estimation procedures based on these approximations, see for instance [37] and [62]. On the other hand, the asymptotic variance of such estimators when the number of particles grows to infinity is available thanks to the central limit theorems proved in Chapter 4. In the particular case of two-filter algorithms, we prove that the asymptotic variance only involves terms related to variance of filtering distributions. This asymptotic variance may then be estimated using the recent results of [64], which is not the case for other known algorithms to estimate the marginal smoothing distributions $\phi_{\chi, t-1:t|T}^{\theta_p}$ for $t = 1$ to $t = T$.

In the remainder of this chapter, we describe the algorithms studied in Chapter 4 and highlight our contributions in comparison to existing results on the approximations of smoothing distributions using Sequential Monte Carlo methods. All the methods presented in this chapter focus on the approximation of $\phi_{\chi, t|T}^{\theta_p}$ for $t = 0$ to $t = T$ as this is the usual setting of two-filter based algorithms. This may be extended to bivariate smoothing distributions with no particular technicalities.

2.1 Two-filter algorithms

Two-filter approaches are Sequential Monte Carlo methods that combine a forward pass and a backward pass to approximate the marginal smoothing distributions $\phi_{\chi, t|T}$ for $t = 0$ to $t = T$. The aim of the forward pass is to obtain an approximation of the forward filtering distributions $\phi_{\chi, t} = \phi_{\chi, t:t|t}$ recursively for $t = 0$ to $t = T$. [41] and [59] introduced an algorithm based on importance sampling and importance

CHAPTER 2. TWO-FILTER ALGORITHMS

resampling steps to approximate $\phi_{\chi,t}$ using a set of points, the *particles*, $\{\xi_t^\ell\}_{\ell=1}^N$ associated with weights $\{\omega_t^\ell\}_{\ell=1}^N$. At $t = 0$, N particles $\{\xi_0^\ell\}_{\ell=1}^N$ are sampled independently from the instrumental density ρ_0 . Then, ξ_0^ℓ is associated with the importance weights

$$\omega_0^\ell := \chi(\xi_0^\ell) g_0(\xi_0^\ell) / \rho_0(\xi_0^\ell).$$

For any bounded and measurable function h defined on X , the expectation $\phi_{\chi,0}[h]$ is approximated by

$$\phi_{\chi,0}^N[h] := \frac{1}{\Omega_0^N} \sum_{\ell=1}^N \omega_0^\ell h(\xi_0^\ell), \quad \Omega_0^N := \sum_{\ell=1}^N \omega_0^\ell.$$

Then, using $\{(\xi_{t-1}^\ell, \omega_{t-1}^\ell)\}_{\ell=1}^N$, pairs $\{(I_t^\ell, \xi_t^\ell)\}_{\ell=1}^N$ of indices and particles are sampled from the instrumental distribution:

$$\pi_{t|t}(\ell, x) \propto \omega_{t-1}^\ell \vartheta_t(\xi_{t-1}^\ell) p_t(\xi_{t-1}^\ell, x)$$

defined on $\{1, \dots, N\} \times X$, where $\{\vartheta_t(\xi_{t-1}^\ell)\}_{\ell=1}^N$ are adjustment multiplier weights and p_t is a transition density chosen by the user. This means that each new particle ξ_t^ℓ at time t is sampled in two steps:

- choose a particle index I_t^ℓ at time $t - 1$ in $\{1, \dots, N\}$ with probabilities proportional to $\omega_{t-1}^j \vartheta_t(\xi_{t-1}^j)$, for j in $\{1, \dots, N\}$;
- sample ξ_t^ℓ using this chosen particle according to $\xi_t^\ell \sim p_t(\xi_{t-1}^{I_t^\ell}, \cdot)$.

For $\ell \in \{1, \dots, N\}$, ξ_t^ℓ is associated with the following importance weight:

$$\omega_t^\ell := \frac{q(\xi_{t-1}^{I_t^\ell}, \xi_t^\ell) g_t(\xi_t^\ell)}{\vartheta_t(\xi_{t-1}^{I_t^\ell}) p_t(\xi_{t-1}^{I_t^\ell}, \xi_t^\ell)}$$

and the expectation $\phi_{\chi,t}[h]$ is approximated by

$$\phi_{\chi,t}^N[h] := \frac{1}{\Omega_t^N} \sum_{\ell=1}^N \omega_t^\ell h(\xi_t^\ell), \quad \Omega_t^N := \sum_{\ell=1}^N \omega_t^\ell.$$

The backward pass is based on a family of nonnegative measurable functions $\{\gamma_t\}_{t \geq 0}$ defined on X chosen by the user (required assumptions are given in Chapter 4). These backward functions are used to define an auxiliary distribution called

backward information filter which may be computed recursively backward in time. The backward information filter is defined, for any bounded and measurable function h on X , by:

$$\psi_{\gamma, t|T}[h] := \frac{\int \gamma_t(x_t) \left[\prod_{u=t+1}^T g_{u-1}(x_{u-1}, x_u) \right] g_T(x_T) h(x_t) \nu(dx_{0:T})}{\int \gamma_t(x_t) \left[\prod_{u=t+1}^T g_{u-1}(x_{u-1}, x_u) \right] g_T(x_T) \nu(dx_{0:T})}.$$

The backward information filter cannot be explicitly computed but may be numerically approximated using Sequential Monte Carlo methods. At $t = T$, sample independently N particles $\{\check{\xi}_{T|T}^i\}_{i=1}^N$ from the instrumental density $\check{\rho}_T$. Each particle $\check{\xi}_{T|T}^i$ is associated with the following importance weight

$$\check{\omega}_{T|T}^i := g_T(\check{\xi}_{T|T}^i) \gamma_T(\check{\xi}_{T|T}^i) / \check{\rho}_T(\check{\xi}_{T|T}^i).$$

$\psi_{\gamma, T|T}[h]$ is then approximated by

$$\psi_{\gamma, T|T}^N[h] := \frac{1}{\check{\Omega}_{T|T}} \sum_{i=1}^N \check{\omega}_{T|T}^i h(\check{\xi}_{T|T}^i), \quad \check{\Omega}_{T|T} = \sum_{i=1}^N \check{\omega}_{T|T}^i.$$

A particle approximation of the backward information filter at time t can be derived by choosing an adjustment weight $\vartheta_{t|T}$ and an instrumental density $r_{t|T}$ as for the forward pass. Pairs $\{(\check{I}_t^i, \check{\xi}_{t|T}^i)\}_{i=1}^N$ of indices and particles at time t are sampled from the instrumental distribution:

$$\pi_{t|T}(i, x_t) \propto \frac{\check{\omega}_{t+1|T}^i \vartheta_{t|T}(\check{\xi}_{t+1|T}^i)}{\gamma_{t+1}(\check{\xi}_{t+1|T}^i)} r_{t|T}(\check{\xi}_{t+1|T}^i, x_t).$$

Each new particle $\check{\xi}_{t|T}^i$ at time t is sampled in two steps:

- choose a particle index \check{I}_t^i at time $t + 1$ in $\{1, \dots, N\}$ with probabilities proportional to $\check{\omega}_{t+1|T}^i \vartheta_{t|T}(\check{\xi}_{t+1|T}^i) / \gamma_{t+1}(\check{\xi}_{t+1|T}^i)$;
- sample $\check{\xi}_{t|T}^i$ using this chosen particle according to $\check{\xi}_{t|T}^i \sim r_{t|T}(\check{\xi}_{t+1|T}^i, \cdot)$.

The particles $\{\check{\xi}_{t|T}^i\}_{i=1}^N$ are then associated with the importance weights:

$$\check{\omega}_{t|T}^i := \frac{\gamma_t(\check{\xi}_{t|T}^i) g_t(\check{\xi}_{t|T}^i) q(\check{\xi}_{t|T}^i, \check{\xi}_{t+1|T}^i)}{\vartheta_{t|T}(\check{\xi}_{t+1|T}^i) r_{t|T}(\check{\xi}_{t+1|T}^i, \check{\xi}_{t|T}^i)}.$$

CHAPTER 2. TWO-FILTER ALGORITHMS

Combining the definition of the forward filter at time $t - 1$ and backward filter at time $t + 1$, the marginal smoothing distribution at time t may be expressed as:

$$\phi_{\chi,t|T}[h] \propto \int \phi_{\chi,t-1}(dx) \psi_{\gamma,t+1|T}(dx') \left[q(x, x_t) g_t(x_t) \frac{q(x_t, x')}{\gamma_{t+1}(x')} \right] h(x_t) \nu(dx_t). \quad (2.1)$$

Using this formula for the marginal smoothing distribution $\phi_{\chi,t|T}$, several algorithms may be used to combine the particles produced in the forward filtering pass and in the backward filtering pass. Chapter 4 provides exponential deviation inequalities and central limit theorems for all the algorithms proposed in the Sequential Monte Carlo literature based on the two filter approach. These algorithms are briefly recalled below.

TwoFilt_{fw} approach: Plugging the particle approximations of the forward and backward filters in (2.1) yields the following mixture approximation of the smoothing distribution $\phi_{\chi,t|T}[h]$:

$$\hat{\phi}_{\chi,t|T}^{tar}[h] \propto \sum_{i=1}^N \sum_{j=1}^N \frac{\omega_{t-1}^i \check{\omega}_{t+1|T}^j}{\gamma_{t+1}(\check{\xi}_{t+1|T}^j)} q(\xi_{t-1}^i, x_t) g_t(x_t) q(x_t, \check{\xi}_{t+1|T}^j).$$

Based on the auxiliary procedure introduced in [34], pairs $\{(I^\ell, J^\ell, \check{\xi}_{t|T}^\ell)\}_{\ell=1}^T$ of indices and particles are sampled at time t from the instrumental density:

$$\pi_{t|T}(i, j, x_t) \propto \frac{\omega_{t-1}^i \tilde{\vartheta}_{t|T}(\xi_{t-1}^i, \check{\xi}_{t+1|T}^j) \check{\xi}_{t+1|T}^j}{\gamma_{t+1}(\check{\xi}_{t+1|T}^j)} \tilde{r}_{t|T}(\xi_{t-1}^i, \check{\xi}_{t+1|T}^j; x_t),$$

where, as above, $\tilde{\vartheta}_{t|T}$ is the adjustment multiplier weight and $\tilde{r}_{t|T}$ is the instrumental density. Each tuple $\{(I^\ell, J^\ell, \check{\xi}_{t|T}^\ell)\}_{\ell=1}^T$ is associated with the importance weight:

$$\tilde{\omega}_{t|T}^\ell := \frac{q(\xi_{t-1}^{I^\ell}, \check{\xi}_{t|T}^\ell) g_t(\check{\xi}_{t|T}^\ell) q(\check{\xi}_{t|T}^\ell, \check{\xi}_{t+1|T}^{J^\ell})}{\tilde{\vartheta}_{t|T}(\xi_{t-1}^{I^\ell}, \check{\xi}_{t+1|T}^{J^\ell}) \tilde{r}_{t|T}(\xi_{t-1}^{I^\ell}, \check{\xi}_{t+1|T}^{J^\ell}, \check{\xi}_{t|T}^\ell)}.$$

Then, the auxiliary indices $\{(I^\ell, J^\ell)\}_{\ell=1}^N$ are discarded and $\{\tilde{\omega}_{t|T}^\ell, \check{\xi}_{t|T}^\ell\}_{\ell=1}^N$ approximate the target smoothing distribution.

TwoFilt_{bdm} approach: Instead of considering the marginal of the target distribution over pairs of indices, [10] considers two partial auxiliary distributions having densities defined on $\{1, \dots, N\} \times X$,

$$\hat{\phi}_{t|T}^{aux,f}(i, x_t) \propto \omega_{t-1}^i q(\xi_{t-1}^i, x_t) g_t(x_t) \sum_{j=1}^N \frac{\check{\omega}_{t+1|T}^j}{\gamma_{t+1}(\check{\xi}_{t+1|T}^j)} q(x_t, \check{\xi}_{t+1|T}^j),$$

$$\hat{\phi}_{t|T}^{aux,b}(j, x_t) \propto \frac{\check{\omega}_{t+1|T}^j}{\gamma_{t+1}(\check{\xi}_{t+1|T}^j)} q(x_t, \check{\xi}_{t+1|T}^j) g_t(x_t) \sum_{i=1}^N \omega_{t-1}^i q(\xi_{t-1}^i, x_t).$$

Then, we may sample from $\hat{\phi}_{\chi, t|T}^{tar}$ by simulating either $\{(I^\ell, \check{\xi}_{t|T}^\ell)\}_{\ell=1}^N$ or $\{(J^\ell, \check{\xi}_{t|T}^\ell)\}_{\ell=1}^N$ from the instrumental distributions:

$$\pi_{t|T}^f(i, x_t) \propto \omega_{t-1}^i \vartheta(\xi_{t-1}^i) p_t(\xi_{t-1}^i, x_t),$$

$$\pi_{t|T}^b(j, x_t) \propto \frac{\check{\omega}_{t+1|T}^j \vartheta(\check{\xi}_{t+1|T}^j) r_{t|T}(\check{\xi}_{t+1|T}^j, x_t)}{\gamma_{t+1}(\check{\xi}_{t+1|T}^j)},$$

where (ϑ_t, p_t) and $(\vartheta_{t+1|T}, r_{t|T})$ are the adjustment multiplier weight functions and the instrumental kernels used in the forward and backward passes, respectively. In this case, the particles $\{\xi_{t|T}^i\}_{i=1}^N$ and $\{\check{\xi}_{t|T}^j\}_{i=1}^N$ are associated with the forward $\{\tilde{\omega}_{t|T}^{i,f}\}_{i=1}^N$ and backward $\{\tilde{\omega}_{t|T}^{i,b}\}_{i=1}^N$ importance weights, respectively, given by:

$$\tilde{\omega}_{t|T}^{i,f} \propto \omega_t^i \sum_{j=1}^N \check{\omega}_{t+1|T}^j q(\xi_t^i, \check{\xi}_{t+1|T}^j) / \gamma_{t+1}(\check{\xi}_{t+1|T}^j),$$

$$\tilde{\omega}_{t|T}^{i,b} \propto \check{\omega}_{t|T}^j \sum_{i=1}^N \omega_{t-1}^i q(\xi_{t-1}^i, \check{\xi}_{t|T}^j) / \gamma_t(\check{\xi}_{t|T}^j).$$

Both TwoFilt_{fw} and TwoFilt_{bdm} approaches have a computational complexity of order $\mathcal{O}(N^2)$. However, [34] proposed a specific way to implement TwoFilt_{fw} to obtain an algorithm with complexity of order $\mathcal{O}(N)$. We also show in Chapter 4 that there is another way of interpreting TwoFilt_{bdm} to provide an approximation of the smoothing distributions with complexity of order $\mathcal{O}(N)$.

2.2 Other Sequential Monte Carlo algorithms to approximate smoothing distributions

Poor man's smoother

The Poor man's smoother is the first algorithm introduced to approximate joint smoothing distributions using Sequential Monte Carlo methods. As explained in [59] or [20], the forward filter (forward pass of the two-filter algorithm) can be used to obtain not only an approximation of $\phi_{\chi,t}$ for $t = 0$ to $t = T$ but also an approximation of the smoothing distribution $\phi_{\chi,0:T|T}$. The estimator is based on the genealogy of all the particles obtained at time T . At each time step $t \in \{0 \leq t \leq T - 1\}$, the N trajectories stored by the Poor man's smoother are given, for $1 \leq \ell \leq N$, by:

$$\xi_{0:t+1}^\ell := \left(\xi_{0:t}^{I_{t+1}^\ell}, \xi_{t+1}^\ell \right).$$

Using this recursion, we obtain a family of N weighted trajectories which allows to approach the expectation $\phi_{\chi,0:T|T}[h]$ of any measurable function h on X^{T+1} by

$$\phi_{\chi,0:T|T}^{N,\text{pms}}[h] := \frac{1}{\Omega_T^N} \sum_{\ell=1}^N \omega_T^\ell h(\xi_{0:T}^\ell).$$

However, as explained for instance in [59], [60] and [34], [81], the above formula is not an efficient approximation of the smoothing distribution because of the path degeneracy issue. At each time step $t \geq 1$, a new trajectory is obtained by combining the newly sampled particle ξ_t^ℓ with the ancestral trajectory $\xi_{0:t-1}^{I_t^\ell}$ chosen among N existing trajectories. Therefore, if a ancestral trajectory is not selected by any new particle, it is discarded and will not be taken into account to approximate the smoothing distribution. As t increases, the number of resampling steps increases accordingly and more and more ancestral trajectories are discarded. At time T , it is common that all trajectories share a common ancestral path up to a given time in the past which leads to poor estimators: no matter the number of particles N , for a small value of t , $\phi_{\chi,t|T}^{N,\text{pms}}$ is approximated by a unique particle duplicated N times. As an illustration, it is shown in [81] that when the Poor man's smoother is used to

approximate smoothed expectation of additive functionals, the asymptotic variance of the estimator grows at least with T^2 .

FFBS algorithm

An alternative to overcome the path degeneracy issue of the Poor man's smoother is the Forward Filtering Backward Smoothing (FFBS) algorithm introduced in [28]. This algorithm requires first that a forward pass has produced a set of particles and importance weights $\{(\xi_t^\ell, \omega_t^\ell)\}$, for $\ell \in \{1, \dots, N\}$ and $t \in \{0, \dots, T\}$ to approximate $\phi_{\chi, t}$ for $t = 0$ to $t = T$. Instead of discarding all particles which are not chosen during the multiple resampling steps, The FFBS algorithm stores all particles and weights without taking into consideration the genealogy of the particles. Then, a backward pass is performed, keeping all the particles fixed but modifying all importance weights computed during the forward pass. For any measurable function h on X , the FFBS approximation of $\phi_{\chi, t|T}[h]$ is written:

$$\phi_{\chi, t|T}^{N, \text{FFBS}}[h] = \sum_{i=1}^N \omega_{t|T}^i h(\xi_t^i), \quad \sum_{i=1}^N \omega_{t|T}^i = 1.$$

The new normalized weights $\omega_{t|T}^i$ for $1 \leq i \leq N$ and $0 \leq t \leq T$ are obtained by using the backward Markov kernel associated with the hidden Markov model.

- At time T the new weights are given, for all $i \in \{1, \dots, N\}$, by $\omega_{T|T}^i = \omega_T^i / \Omega_T^N$.
- For $t = T - 1$ to $t = 0$ note by [57] that

$$\phi_{\chi, t|T}[h] = \int \phi_{\chi, t+1|T}(\mathrm{d}x_{t+1}) \frac{q(x_t, x_{t+1}) \phi_{\chi, t}(\mathrm{d}x_t)}{\int q(x, x_{t+1}) \phi_{\chi, t}(\mathrm{d}x)} h(x_t).$$

Then, plugging the approximation $\phi_{\chi, t+1|T}^{N, \text{FFBS}}$ and $\phi_{\chi, t}^N$ in this expression yields the approximation

$$\phi_{\chi, t|T}^{N, \text{FFBS}}[h] = \sum_{i=1}^N \omega_t^i \left(\sum_{j=1}^N \omega_{t+1|T}^j \frac{q(\xi_t^i, \xi_{t+1}^j)}{\sum_{\ell=1}^N \omega_t^\ell q(\xi_t^\ell, \xi_{t+1}^i)} \right) h(\xi_t^i).$$

Then the new importance weights at time t are given, for all $1 \leq i \leq N$, by:

$$\omega_{t|T}^i := \omega_t^i \left(\frac{\sum_{j=1}^N \omega_{t+1|T}^j \frac{q(\xi_t^i, \xi_{t+1}^j)}{\sum_{\ell=1}^N \omega_t^\ell q(\xi_t^\ell, \xi_{t+1}^i)} \right).$$

This algorithm has a computational complexity which grows quadratically with the number of particles due to the weights updates in the backward pass. This $\mathcal{O}(N^2)$ complexity is only true when approximating marginal smoothing distribution. The FFBS algorithm may also be used to approximate joint smoothing distributions of the form $\phi_{\chi, s:t|T}$ with $s < t$ but with a complexity of order $\mathcal{O}(N^{t-s+1})$ which makes it prohibitive. An important feature of the FFBS algorithm shown in [73, 13, 22] is that it can be implemented using only a forward pass when approximating smoothed expectations of additive functionals.

FFBSi algorithm

The Forward Filtering Backward Simulation (FFBSi) algorithm is another alternative proposed by [40] to avoid the degeneracy issue when approximating smoothing distributions with Sequential Monte Carlo methods. As for the FFBS algorithm, a forward pass is performed to produce $\{(\xi_t^\ell, \omega_t^\ell)\}$, for $\ell \in \{1, \dots, N\}$ and $t \in \{0, \dots, T\}$ to approximate $\phi_{\chi, t}$ for $t = 0$ to $t = T$ and all these particles and weights are stored without taking into consideration the genealogy of the particles. Instead of computing new weights in the backward pass, the FFBSi samples backward new trajectories from time T to time 0 among all the N^{T+1} trajectories which can be chosen in $\{\xi_t^\ell\}$, $1 \leq \ell \leq N$ and $0 \leq t \leq T$. Each trajectory is sampled according to the following steps.

- At time $t = T$, sample $J_T \in \{1, \dots, N\}$ with probabilities proportional to $\omega_T^\ell / \Omega_T^N$, for $\ell \in \{1, \dots, N\}$.
- For $t = T - 1$ to $t = 0$, J_t is sampled in $\{1, \dots, N\}$ with probabilities

$$\Lambda_t^N(J_{t+1}, i) := \frac{\omega_t^i q(\xi_t^i, \xi_{t+1}^{J_{t+1}})}{\sum_{\ell=1}^N \omega_t^\ell q(\xi_t^\ell, \xi_{t+1}^{J_{t+1}})}, \quad 1 \leq i \leq N.$$

These sampling steps are repeated N times to produce $\{J_0^\ell, \dots, J_T^\ell\}$, $1 \leq \ell \leq N$. Then, the smoothed expectation of any measurable function h on X^{T+1} is approximated by:

$$\phi_{\chi, 0:T|T}^{N, \text{FFBSi}}[h] := \frac{1}{N} \sum_{\ell=1}^N h(\xi_0^\ell, \dots, \xi_T^\ell).$$

Contrary to the FFBS algorithm, the FFBSi algorithm samples trajectories and may be used to approximate joint smoothing distributions without increasing its complexity. Nevertheless, the computation of the transition probabilities Λ_t , $0 \leq t \leq T-1$ makes it a $\mathcal{O}(N^2)$ algorithm. In the case where the transition probability q is upper bounded, [24] introduced an acceptance rejection mechanism to implement the FFBSi algorithm with a $\mathcal{O}(N)$ complexity, see also [30, 31] for an application to the approximation of conditional expectations of additive functionals of the hidden states given the observations.

PaRis algorithm

More recently, [78] introduced the particle-based, rapid incremental smoother (PaRis) algorithm which combines both the forward only version of the FFBS algorithm with the sampling mechanism of the FFBSi algorithm. The algorithm does not produce an approximation of the smoothing distributions but of the smoothed expectation of a fixed additive functional. It may be used to approximate $\phi_{\chi, 0:T|T}[S_T]$ where, for all $x_{0:T} \in X^{T+1}$,

$$S_T(x_{0:T}) = \sum_{t=1}^T h_t(x_{t-1}, x_t), \quad (2.2)$$

where $\{h_t\}_{t=1}^T$ are bounded measurable functions on X^2 . The algorithm is based on sufficient statistics τ_t^i , for $1 \leq i \leq N$ and $0 \leq t \leq T$ starting with τ_0^i for all $1 \leq i \leq N$. Let $\tilde{N} \geq 1$. Then, at each time step $1 \leq t \leq n$ these statistics are update according to the following steps.

- Run a particle filter to produce $\{(\xi_t^\ell, \omega_t^\ell)\}$ for $1 \leq \ell \leq N$.
- For all $1 \leq i \leq N$, sample independently J_t^ℓ in $\{1, \dots, N\}$ for $1 \leq \ell \leq \tilde{N}$ with probabil-

ities $\Lambda_{t-1}^N(i, \cdot)$. Then, set

$$\tau_t^i := \frac{1}{\tilde{N}} \sum_{\ell=1}^{\tilde{N}} \left\{ \tau_{t-1}^{J_t^\ell} + h_t \left(\xi_{t-1}^{J_t^\ell}, \xi_t^i \right) \right\}.$$

Then, $\phi_{\chi, 0:T|T}[S_T]$ is approximated by

$$\phi_{\chi, 0:T|T}^{N, \text{PaRis}}[S_T] = \frac{1}{\Omega_T^N} \sum_{i=1}^N \omega_T^i \tau_T^i.$$

As for the FFBSi algorithm, when q is upper bounded the PaRis algorithm may be implemented with $\mathcal{O}(N)$ complexity using the procedure of [24]. As proved in [78], the algorithm is asymptotically consistent (as N goes to infinity) for any fixed precision parameter \tilde{N} . However, there is a significant qualitative difference between the cases $\tilde{N} = 1$ and $\tilde{N} \geq 2$.

2.3 Contribution

This section focuses on the convergence properties of the two-filter algorithms in the general state-space models. These results are proved in the article displayed in Chapter 4. We provide exponential deviation inequalities, central limit theorems and time-uniform bounds under *strong mixing* assumptions.

2.3.1 Exponential Deviation Inequalities

Exponential deviation inequalities have been established in the past few years for the FFBS, FFBSi and PaRis algorithms. In [24] the authors provide exponential deviation inequalities for $\phi_{0:T|T}^{N, \text{FFBS}}$ and $\phi_{0:T|T}^{N, \text{FFBSi}}$. Under mild assumptions on the model, it is proved in [24, Theorem 5] and [24, Corollary 6] that there exist B_T and C_T such that for all $N \geq 1$, all $\varepsilon > 0$ and all bounded measurable function h defined on X^{T+1} ,

$$\begin{aligned} \mathbb{P} \left(\left| \phi_{0:T|T}[h] - \phi_{0:T|T}^{N, \text{FFBS}}[h] \right| \geq \varepsilon \right) &\leq B_T \exp \left\{ -\frac{C_T N \varepsilon^2}{\text{osc}(h)^2} \right\}, \\ \mathbb{P} \left(\left| \phi_{0:T|T}[h] - \phi_{0:T|T}^{N, \text{FFBSi}}[h] \right| \geq \varepsilon \right) &\leq B_T \exp \left\{ -\frac{C_T N \varepsilon^2}{\text{osc}(h)^2} \right\}, \end{aligned}$$

where for any real valued function h defined on a space Z , $\text{osc}(h) := \sup_{x, x' \in Z} |h(x) - h(x')|$. Similar results have been obtained simultaneously in [22]. Under additional strong mixing conditions, in particular assuming that q is uniformly upper and lower bounded, it is proved in [24, Theorem 6] that these results are uniform in time (with constants which do not depend on T) for the marginal smoothing distributions $\phi_{t|T}^{N, \text{FFBS}}[h]$ and $\phi_{t|T}^{N, \text{FFBSi}}[h]$ for any function h defined on X .

Similar deviation inequalities as in [24] are given in [78, Theorem 1 and Corollary 2] for $\phi_{x, 0:T|T}^{N, \text{PaRis}}[S_T]$, where S_T is an additive functional defined as in (2.2). Using the strong mixing condition, [31] obtained L^q mean error bounds for $\phi_{0:T|T}^{N, \text{FFBS}}[S_T]$ and $\phi_{0:T|T}^{N, \text{FFBSi}}[S_T]$. For all $q \geq 2$, it is proved in [31, Theorem 1] that there exists $C > 0$ such that for all $T < \infty$ and all sequences of bounded measurable functions $\{h_t\}_{t=1}^T$,

$$\begin{aligned} \left\| \phi_{0:T|T}[S_T] - \phi_{0:T|T}^{N, \text{FFBS}}[S_T] \right\|_q &\leq C \sqrt{\frac{T}{N}} \left(1 + \sqrt{\frac{T}{N}} \right), \\ \left\| \phi_{0:T|T}[S_T] - \phi_{0:T|T}^{N, \text{FFBSi}}[S_T] \right\|_q &\leq C \sqrt{\frac{T}{N}} \left(1 + \sqrt{\frac{T}{N}} \right). \end{aligned}$$

By [31, Theorem 2], we also know that there exists $C > 0$ such that for all $T < \infty$ and all sequences of bounded measurable functions $\{h_t\}_{t=1}^T$,

$$\begin{aligned} \mathbb{P} \left(\left| \phi_{0:T|T}[S_T] - \phi_{0:T|T}^{N, \text{FFBS}}[S_T] \right| \geq \varepsilon \right) &\leq B \exp \left\{ -CN\varepsilon^2/T \right\}, \\ \mathbb{P} \left(\left| \phi_{0:T|T}[S_T] - \phi_{0:T|T}^{N, \text{FFBSi}}[S_T] \right| \geq \varepsilon \right) &\leq B \exp \left\{ -CN\varepsilon^2/T \right\}. \end{aligned}$$

All these results are crucial to assess the statistical efficiency of Sequential Monte Carlo methods to approximate smoothing distributions which is a key step to ensure consistency and asymptotic normality of Monte Carlo maximum likelihood inference procedures. However, there are very few results in the literature to prove convergence properties of two-filter algorithms. The following deviation inequalities are proved in Chapter 4.

Deviation inequality for the product distribution: establishing controls for the two filter algorithms requires first to obtain deviation inequalities for the product distribution of the forward and backward filters. Therefore, we prove that for all

CHAPTER 2. TWO-FILTER ALGORITHMS

$0 \leq s < t \leq T$, there exist constants $B_{s,t|T}, C_{s,t|T} > 0$, such that, for all $N, \epsilon > 0$ and all measurable functions h on X^2 ,

$$\mathbb{P} \left[\left| \sum_{i,j=1}^N \frac{\omega_s^i \check{\omega}_{t|T}^j}{\check{\Omega}_s \check{\Omega}_{t|T}} h(\xi_{s,s}^i, \xi_{t|T}^j) - \phi_{\chi,s} \otimes \psi_{\gamma,t|T}[h] \right| \geq \epsilon \right] \leq B_{s,t|T} e^{-C_{s,t|T} N \epsilon^2 / \text{osc}(h)^2}.$$

Deviation inequality for the TwoFilt_{fwt} and TwoFilt_{bdm} algorithms: the previous result is then used to provide exponential deviation inequalities for the TwoFilt_{fwt} and TwoFilt_{bdm} estimators: it is proved in Chapter 4 that for all $1 \leq t \leq T - 1$, there exist constants $B_{t|T}, C_{t|T} > 0$, such that, for a measurable function h on X ,

$$\begin{aligned} \mathbb{P} \left[\left| \frac{\sum_{i=1}^N \tilde{\omega}_{t|T}^i h(\tilde{\xi}_{t|T}^i)}{\sum_{i=1}^N \tilde{\omega}_{t|T}^i} - \phi_{\chi,t|T}[h] \right| \geq \epsilon \right] &\leq B_{t|T} \exp \left\{ -C_{t|T} N \epsilon^2 / \text{osc}(h)^2 \right\}, \\ \mathbb{P} \left[\left| \frac{\sum_{i=1}^N \tilde{\omega}_{t|T}^{i,f} h(\xi_t^i)}{\sum_{i=1}^N \tilde{\omega}_{t|T}^{i,f}} - \phi_{\chi,t|T}[h] \right| \geq \epsilon \right] &\leq B_{t|T} \exp \left\{ -C_{t|T} N \epsilon^2 / \text{osc}(h)^2 \right\}, \\ \mathbb{P} \left[\left| \frac{\sum_{i=1}^N \tilde{\omega}_{t|T}^{i,b} h(\check{\xi}_t^i)}{\sum_{i=1}^N \tilde{\omega}_{t|T}^{i,f}} - \phi_{\chi,t|T}[h] \right| \geq \epsilon \right] &\leq B_{t|T} \exp \left\{ -C_{t|T} N \epsilon^2 / \text{osc}(h)^2 \right\}. \end{aligned}$$

However, these bounds are hardly interesting in practice since there is no explicit dependency on T . To overcome this issue we use additional *strong mixing* to obtain time uniform deviation inequalities:

H1. Assumptions for the forward filter: there exist $0 < \sigma_- < \sigma_+ < \infty$ and $c_- > 0$ such that for all $x, x' \in X$, $q(x, x') \in [\sigma_-, \sigma_+]$, and for all $t \geq 0$,

$$\int \chi(dx_0) g_0(x_0) \geq c_- \quad \text{and} \quad \inf_{x \in X} \int Q(x, dx') g_t(x') \geq c_-.$$

H2. Assumptions for the backward filter: there exist $0 < \gamma_- < \gamma_+ < \infty$ and $\check{c} > 0$ such that for all $x \in X$ and all $t \geq 0$, $\gamma_t(x) \in [\gamma_-, \gamma_+]$, and

$$\int \gamma_T(x_T) g_T(x_T) dx_T \geq \check{c}_- \quad \text{and} \quad \inf_{x \in X} \int \gamma_t(x_t) g_t(x_t) q(x_t, x) \gamma_{t+1}^{-1}(x) dx_t \geq \check{c}_-.$$

Under these strong mixing assumptions, similarly to [24], [31], [78] the exponential deviation inequalities for the TwoFilt_{fwt} and TwoFilt_{bdm} algorithms are uniform in time (with constants that do not depend on T).

2.3.2 Asymptotic normality

[24] established a central limit theorem for the FFBS and FFBSi algorithms. They proved that both $\sqrt{N}(\phi_{0:T|T}[h] - \phi_{0:T|T}^{N,FFBS}[h])$ and $\sqrt{N}(\phi_{0:T|T}[h] - \phi_{0:T|T}^{N,FFBSi}[h])$ are asymptotically normal for all measurable functions h on X^{T+1} with an explicit expression of the asymptotic variance as N goes to infinity under mild assumptions on the hidden Markov model. Under strong mixing assumptions, [24, Theorem 12] provides a time uniform bound on the asymptotic variance of $\sqrt{N}(\phi_{t|T}[h] - \phi_{t|T}^{N,FFBS}[h])$ and $\sqrt{N}(\phi_{t|T}[h] - \phi_{t|T}^{N,FFBSi}[h])$ for all measurable functions h on X .

[78, Corollary 5] proved the asymptotic normality of $\sqrt{N}(\phi_{0:T|T}[S_T] - \phi_{0:T|T}^{N,PaRis}[S_T])$ as N goes to infinity for additive functionals defined as in (2.2). The asymptotic variance is given by the sum of the asymptotic variance of the FFBSi estimator and an additional term due to all the additional sampling steps.

Chapter 4 provides new central limit theorems for the two-filter algorithms. These results are derived by using the limit theorems for weighted samples given in [25]. Note first that the asymptotic normality of the forward filter has been established for instance in [20]. For all $0 \leq t \leq T$ and all measurable functions h on X ,

$$\sqrt{N} \left[\sum_{i=1}^N \frac{\omega_t^i}{\Omega_t} h(\xi_t^i) - \phi_{\chi,t}[h] \right] \xrightarrow[N \rightarrow +\infty]{\mathcal{D}} \mathcal{N}(0, \Gamma_{\chi,s}[h - \phi_{\chi,t}[h]]),$$

where $\Gamma_{\chi,s}[h]$ is calculated by a recursive formula which is detailed in [20]. Following the same steps, it is easily proved that for any measurable function h on X ,

$$\sqrt{N} \left[\sum_{j=1}^N \frac{\check{\omega}_{t|T}^j}{\check{\Omega}_{t|T}} h(\check{\xi}_{t|T}^j) - \psi_{\gamma,t|T}[h] \right] \xrightarrow[N \rightarrow +\infty]{\mathcal{D}} \mathcal{N}(0, \check{\Gamma}_{t|T}[h - \psi_{\gamma,t|T}[h]]),$$

where $\check{\Gamma}_{t|T}[h]$ is calculated by a recursive formula in the backward time direction. Combining these two results yields, for all $0 \leq s < t \leq T$ and any measurable function h on X^2 ,

$$\sqrt{N} \left[\sum_{i,j=1}^N \frac{\omega_s^i \check{\omega}_{t|T}^j}{\Omega_s \check{\Omega}_{t|T}} h(\xi_s^i, \check{\xi}_{t|T}^j) - \phi_{\chi,s} \otimes \psi_{\gamma,t|T}[h] \right] \xrightarrow[N \rightarrow +\infty]{\mathcal{D}} \mathcal{N}(0, \tilde{\Gamma}_{s,t|T}[h - \phi_{\chi,s} \otimes \psi_{\gamma,t|T}[h]]).$$

where $\tilde{\Gamma}_{s,t|T}[h]$ is defined by:

$$\tilde{\Gamma}_{s,t|T}[h] := \Gamma_{\chi,s} \left[\int \psi_{\gamma,t|T}(dx_t) h(\cdot, x_t) \right] + \check{\Gamma}_{\gamma,t|T} \left[\int \phi_{\chi,s}(dx_s) h(x_s, \cdot) \right].$$

This result is crucial since they allow to obtain asymptotic normality of the TwoFilt_{fwt} and TwoFilt_{bdm} algorithms in Chapter 4. On the other hand, its specific form is such that all asymptotic variances of two-filter algorithms rely only on the asymptotic variances of the forward filter and the backward information filter. These variances are much more easy to interpret than asymptotic variances of the FFBS, FFBSi and PaRis algorithms. Note that under strong mixing assumptions we also obtain a time uniform bound on the asymptotic variance of all two filter algorithms.

Chapter 3

Regime switching applied to commodity term structure

Thousands of commodities are traded daily around the world, some may be traded in spot markets for immediate delivery in exchange of another commodity or payment of the given spot price and others in derivative markets through such instruments as forwards, futures, swaps, options... The possibility of trading in spot market or derivative markets depends on the characteristics of a commodity. For example, precious metals and electricity can be traded for immediate delivery but crude oil or natural gas are traded through forward contracts where buyers and sellers agree on forward delivery and payment at a specified future time.

- A spot contract is an agreement to trade a certain amount of a particular commodity on the spot date. For commodities, the spot date can be an immediate transaction for electricity or two business days after today for precious metals for instance.
- A forward contract is an agreement between a buyer and a seller, who both commit to a transaction at a future date at a price set by negotiation today (*forward price*). There is no money transfer between the buyer and the seller initially or during the lifetime of the forward contract, only on the maturity date.
- A future contract is an agreement between two parties: a short position – the party

CHAPTER 3. COMMODITY TERM STRUCTURE MODELING

who agrees to deliver a commodity – and a long position – the party who agrees to receive a commodity at a specified future date at a price fixed today (*future price*). The future exchange states an “initial margin” (the minimum amount of money to be deposited into an account by both the short and long positions) and the account is refunded every trading day with the initial margin plus (or minus) any gains or losses over the span of the future contract. If the amount of an account is below the “maintenance margin” (the lowest amount an account can reach), a “margin call” (an additional amount) is requested to bring the account back to the initial margin. This goes on until the position (long or short) is closed out. The difference in the payment schedules distinguishes forward and future prices in general although [6] proved that if the interest rates are deterministic, the two contracts are essentially equivalent and forward prices are equal to futures prices.

Term Structure in commodity markets refers to the relationship between the future prices of different maturities on a given date, i.e the future contract valorisation for several maturities of one asset. In the case where future prices of shorter maturities are greater than longer maturity ones, the future curve is said to be in *backwardation*. In the other case (that is, when future prices of shorter maturities are less than longer maturity ones), the future curve is said to be in *contango*. A market can switch from backwardation to contango effects. Figure 3.1 displays CME WTI crude oil future prices as a function of the maturity on two different dates in 2007 (left) and in 2011 (right). Each curve is a series of 21 future prices associated with 21 maturities, from 1 to 21 months roughly. These two figures show two different behaviors: in 2007, the term structure is in backwardation (the curve is decreasing) and in 2011 the term structure is in contango (the curve is increasing). The term structure behaviors of WTI crude oil market changed between 2007 and 2011. It is usually hard to analyze the behavior of future contracts on commodities without understanding delivery rules or the features of each specific commodity such as storage cost or technological requirements to extract raw materials. For instance, the WTI crude oil future contract (NYMEX) expires roughly the third business day prior to the 25th day of the month but the delivery procedure does not happen immediately but with a perimeter of 1 month, i.e a non-specified day in the delivery

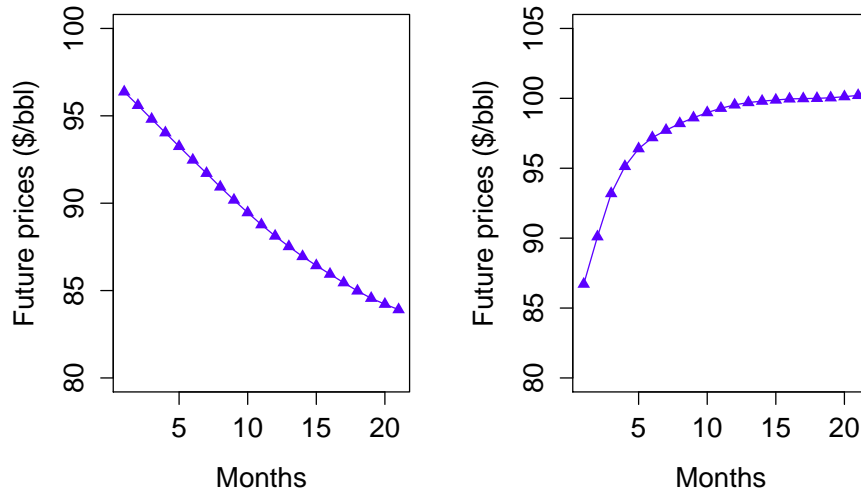


Figure 3.1: Futures term structure of CME WTI Crude oil observed in 2007 (*left*) and in 2011 (*right*).

month¹. People consider these complexities as drawbacks for investment but in exchange this provides more opportunities to the investors mastering the underlying specificities as well as contract rules of the markets.

Term structure of commodity future prices has been studied by many researchers and practitioners since it was first formulated by [54]. The convenience yield was defined by [51] as the benefit or cost required for holding a physical commodity. This factor has been further developed in the theory of storage of [92] and [7], where it is shown that the marginal convenience yield on inventory falls at a decreasing rate as inventories increase. It is commonly assumed that holding a commodity can only generate some additional fees while physical possession offers more flexibility to a firm, by reducing risk of shortage for instance. However, while inventories accumulate the cost associated to their storage increases and the benefit of holding the commodity declines. The role of the convenience yield in future prices dynamics has also been demonstrated in [38] through the notion known as the *cost of carry*

¹For more details on expiration date rules see www.cmegroup.com

which includes at least a cost of storage of oil and which is supposed to be positive (and constant). This cost of carry is not sufficient to understand the term structure of future prices which confirms the usefulness of a more complex factor, the convenience yield.

Inventory level also explains the seasonal behavior of spot and future prices of agricultural assets: future prices are below spot prices when inventories are low and marginal convenience yield on inventory is high. Agricultural products being essential for human consumption and industrial production, when inventories are low demand becomes greater than supply which implies a spot price greater than the future price and a high marginal convenience yield (that is, benefit of holding the commodity). The relationship between the marginal convenience yield and the inventories also exists for other commodities as explained in [32] for metals prices for which inventories are not affected by seasonality as agricultural products but by general business cycles. In particular, when inventories fall, future prices are below spot prices and spot prices are more variable than future prices around business cycle peaks.

3.1 Commodity Modeling

Stochastic properties of commodity markets are widely studied in terms of future prices to take into account the time needed for asset delivery and storage costs in this specific framework. Thus, commodity term structure modeling should be designed to match the structure and variability of future prices in order to (a) price derivative products based on spot or future prices and (b) fit the structure of future prices to given historical to forecast other future prices. Many models have been proposed to perform derivative products pricing. [19] introduced a single factor model for the forward curve which can be used to obtain a stochastic model for the spot price and explicit derivation of the prices of contingent claims on commodities. The model is used to evaluate derivative products like standard options, caps, floors, collars and swaptions by analytical pricing formulas, or American, Asian option by using trinomial trees method. This model has been extended in [18] to estimate the

volatilities and correlations of forward prices to value several energy derivatives. In [55], the future price of a given maturity T depends on two Brownian motions in the following way:

$$dF(t, T) = F(t, T) \left[e^{-\kappa(T-t)} \sigma_1 dW_t^1 + \sigma_2 dW_t^2 \right].$$

This choice is motivated by the observed volatility in the EEX (European Energy Exchange): futures maturing later have a lower volatility (exponential decrease) than futures maturing soon. As $T - t$ becomes very large, the volatility assigned to the contract by the first factor is close to zero. As this is not the case in practice, the second factor keeps the volatility away from zero. The model is calibrated using European call options on electricity futures traded in EEX. [90] developed a tractable model for pricing commodity derivatives in the presence of unspanned stochastic volatility and applied it to crude oil markets.

In this thesis, we are interested in modeling commodity term structure and in fitting models to historical data. [88] proposed a two-factor model decomposing the logarithm of the spot price into two factors: $\ln P_t = X_t + \xi_t$ where ξ_t represents a long-term equilibrium price following $d\xi_t = \mu_\xi dt + \sigma_\xi dW_t^\xi$ and X_t is an Ornstein-Uhlenbeck process representing a short-term deviation $dX_t = -\kappa X_t dt + \sigma_X dW_t^X$. Although the notion of convenience yield does not exist in the model dynamics, this model is equivalent to the stochastic convenience yield modeling proposed by [39] which is detailed below. [69] studied and extended several models based on one or two factors by taking into account the seasonal effects with application to the Nordic Power Exchange electricity prices. The authors described the spot price P_t (or logarithm of spot price $\ln P_t$) as a sum of two components: a deterministic function and an Ornstein-Uhlenbeck process with zero mean reversion X_t such that: $P_t = f(t) + X_t$ (or $\ln P_t = f(t) + X_t$). They also allow the spot price to depend on more than one factor as $P_t = f(t) + X_t + \xi_t$ (or $\ln P_t = f(t) + X_t + \xi_t$) where X_t and ξ_t are the same components as in [88]. The function $f(t)$ is an important factor that determines the shape of future curves that reflects the seasonal pattern. [89] applied the two-factor model of [88] to agricultural future prices by using a suitable deterministic component as seasonal factor. [1] proposed another extension by

CHAPTER 3. COMMODITY TERM STRUCTURE MODELING

including jumps in short-term variation to explain the behavior of oil prices. [70] constructed a linear and Gaussian state-space model where the state variable is a vector of m components (ξ^1, \dots, ξ^n) such that the logarithm of spot price is described as the sum of these components. Each ξ^i is assumed to be an extended Ornstein-Uhlenbeck process depending on n Brownian motions:

$$d\xi_t^i = (\alpha_t^{ij} - \kappa_t^i \xi_t^i) dt + \sum_{j=1}^n \eta_t^{ij} dW_t^j,$$

with κ_t^i , α_t^{ij} and η_t^{ij} deterministic functions of time. In the case where the parameters are constant functions of time, the model leads to an analytic formula to compute future prices as functions of ξ_t^i . In practice, the authors considered the model in two cases:

- one factor with one stochastic mean-reverting process and a deterministic function describing the seasonal behaviors ;
- two factors with one stochastic mean-reverting process and another given by a Brownian motion, and a deterministic function for seasonality.

They applied these two models to Henry Hub natural gas from Bloomberg data, the result showed that the two models can capture the seasonal pattern present in the data, the two-factor model performing better than the one-factor model. In all these models, the observations are given by the future prices observed in Gaussian noise. Time discretization of such models leads to a common Linear and Gaussian state-space model setting in which calibration, pricing or computation of expectations may be performed using Kalman filtering techniques for instance.

[83] proposed a three-factor model (spot price, convenience yield and volatility) in which both the convenience yield and the spot price volatility depend on a seasonality factor described as a sum of trigonometric functions. The logarithm of future price can be formulated as an affine function of these factors (spot price, convenience yield and seasonality) whose coefficients are numerically estimated by solving ordinary differential equations assuming that some contract prices are observed without errors. Other complex models to describe the seasonality have been

introduced such as [72] which proposed a model where the spot price logarithm is a sum of four stochastic processes: a long-term component, a short-term component and two seasonality components. Nevertheless, decomposing the spot price of a commodity into two components, a de-seasonalized spot price and a deterministic seasonal function, seems sufficiently consistent based on existing works. Therefore, this thesis focuses on modeling the de-seasonalized part of the spot price.

The work presented in this thesis is based on the model of [39] where the spot price S_t and convenience yield δ_t are described, under the risk-neutral probability \mathbb{Q} , as:

$$\begin{aligned} dS_t &= (r - \delta_t)S_t dt + \sigma S_t dW_t^1, \\ d\delta_t &= \kappa(\alpha - \delta_t)dt + \eta dW_t^2, \quad d\langle W_t^1, W_t^2 \rangle = \rho dt, \end{aligned}$$

where the parameter $(\mu, \sigma, \kappa, \alpha, \eta, \rho)$ are constant and $((W_t^1, W_t^2), t \geq 0)$ are standard Brownian motions under \mathbb{Q} . In this model, the logarithm of the future price $F_t(S_t, \delta_t, T - t)$ at time t with time to maturity $T - t$ is a linear function of $(\ln S_t, \delta_t)$. In the case where the logarithm of future prices are discretely observed with additional Gaussian noise, parameter estimation may be performed using Kalman filtering techniques. Some extensions of [39] have been developed: [87] and [71] valued options on commodity futures by using a three-factor model: spot price, stochastic interest rate and stochastic convenience yield. Numerical results highlighted that the volatility level of both the spot price and the convenience yield are much more significant than interest rate volatility level. The stochastic interest rate does not have a significant impact on commodity derivative pricing and it may be assumed to be deterministic without loss of generality. [48] extended this three-factor model by introducing jumps in the commodity spot prices. Term Structure behaviors particularly depend on the mean-reversion α , when α is high (resp. low) the market is likely to be on backwardation (resp. contango). However, with constant parameters, the model is not sufficiently flexible to capture simultaneously both effects.

[43] proposed a regime switching autoregressive model for short-term interest rate time series depending on a two-state Markov process characterizing the two regimes the market may run through. The standard deviation of the autoregres-

sive model may switch between two values according to a_t . The author provided a new numerical method estimate this model with an application to interest rate term structure. Regime switching GARCH models have been then developed in [44], [12], [42]. Common features of these models are: (i) the conditional volatility is allowed to switch between a finite number of regimes driven by a first-order Markov process, (ii) the transition probability is assumed to be constant or a time-varying function of exogenous variables. While [12] combined the regime switching model of [43] and the ARCH effect with Gaussian-type errors, [44] were interested in regime switchings incorporated in ARCH models where the innovations can be either Gaussian or Student-t distributions. These models are estimated by maximizing the log-likelihood function.

[36] showed that regime switchings are present in commodity modeling using crude oil data. The authors modeled the daily log-returns of crude oil futures by using regime switching ARMA processes with Student-t distributed innovations. Different cases are studied, the standard deviation of the innovations: (i) are regime dependent, (ii) follow an ARCH dynamics with regime dependent parameters, and (iii) follow a GARCH dynamics where the parameters are regime dependent. The regime in this framework is driven by a latent finite state-space Markov chain. Their numerical experiments showed that regime switching models outperformed GARCH models with constant parameters. [74] proposed a model where the logarithm of electricity price linearly depends on the log-price of the previous date and an explanatory variable (additional variables can be included) with all parameters driven by a finite-state process. Regime switchings applied to commodity markets have been even more developed after the recent financial crisis. [91] proposed a stochastic volatility (SV) model with the log-volatility following an Orstein-Uhlenbeck process where the mean-reverting level is driven by a two-state first-order Markov process. The author applied this model to the return of weekly CME WTI price time series to illustrate that this regime switching model captures most major events affecting the oil market and significantly enhances the forecasting power of the SV model. [17] modeled the evolution of future curves by using a two-regime approach for the two-factor model of [18]. The model is applied to NBP (National Balancing Point)

natural gas future prices data. The impact of oil prices changes on stock market return in the case of Gulf Cooperation Council (GCC) countries is analyzed in [75] using a regime switching model. Practical significance and important implications for both GCC economic policy and financial stability have been concluded (relationship between stock and oil markets; or interest rate and oil markets).

Regime switching models applied to commodity markets usually focus on one specific fixed maturity, but not on commodity term structure. Modeling jointly future time series of several maturities is complex since it requires to take into account statistical dependency among future prices of various maturities. [2] proposed Gibson-Schwartz-based dynamics with parameters driven by a two-state Markov process $\{a_t\}_{t \geq 0}$ under the risk-neutral probability \mathbb{Q} as follows:

$$\begin{aligned} dS_t &= (r - \delta_t)S_t dt + \sigma_{a_t} S_t dW_t^1, \\ d\delta_t &= \kappa(\alpha_{a_t} - \delta_t)dt + \eta_{a_t} dW_t^2, \quad d\langle W_t^1, W_t^2 \rangle = \rho_{a_t} dt. \end{aligned}$$

Statistical estimation is performed using market observations and cannot be done directly under the risk-neutral probability. It is assumed that each factor (spot, convenience yield) has its own market price of risk, $\lambda_{a_t}^S$ for S_t and $\lambda_{a_t}^\delta$ for δ_t . The historical dynamics of the Gibson-Schwartz model is then described as follows:

$$\begin{aligned} dS_t &= (\mu_{a_t} - \delta_t)S_t dt + \sigma_{a_t} S_t d\widetilde{W}_t^1, \\ d\delta_t &= \kappa(\widetilde{\alpha}_{a_t} - \delta_t)dt + \eta_{a_t} d\widetilde{W}_t^2, \quad d\langle \widetilde{W}_t^1, \widetilde{W}_t^2 \rangle = \rho_{a_t} dt, \end{aligned}$$

where $((\widetilde{W}_t^1, \widetilde{W}_t^2), t \geq 0)$ are standard Brownian motions under the historical probability \mathbb{P} with correlation ρ_{a_t} , $\mu_{a_t} = r - \lambda_{a_t}^S$ and $\widetilde{\alpha}_{a_t} = \alpha_{a_t} - \lambda_{a_t}^\delta / \kappa$. In this model, the convenience yield is meant to revert to its equilibrium value $\widetilde{\alpha}_{a_t}$ at speed κ . Based on the technique of [83], the author assumed that future prices of two maturities are observed without error to filter out the state variables associated with these prices and to estimate model parameters. The model we used to analyze commodity term structure is closely related to the one introduced in [2]: under the historical probability \mathbb{P} our model is given by:

$$\begin{aligned} dS_t &= (\mu - \delta_t)S_t dt + \sigma_{a_t} S_t d\widetilde{W}_t^1, \\ d\delta_t &= \kappa(\widetilde{\alpha}_{a_t} - \delta_t)dt + \eta_{a_t} d\widetilde{W}_t^2, \quad d\langle \widetilde{W}_t^1, \widetilde{W}_t^2 \rangle = \rho_{a_t} dt, \end{aligned}$$

CHAPTER 3. COMMODITY TERM STRUCTURE MODELING

where $((\widetilde{W}_t^1, \widetilde{W}_t^2), t \geq 0)$ are standard Brownian motions under the historical probability \mathbb{P} . Contrary to the approach of [2], all future prices for all maturities are observed with Gaussian errors and time discretization of this model with $Z_k = (\ln S_k, \delta_k)$ and noisy future prices vectors $(Y_k)_{k \geq 0}$ leads to the following model (see section 5.3 for a detailed expression of each parameter):

$$Z_k = d_{a_{k-1}} + T_{a_{k-1}} Z_{k-1} + H_{a_{k-1}} \varepsilon_k, \quad (3.1)$$

$$Y_k = c_{a_k} + B_{a_k} Z_k + G_{a_k} \eta_k, \quad (3.2)$$

where

- $(\varepsilon_k)_{k \geq 0}$ is a sequence of independent and identically distributed (i.i.d.) standard Gaussian variables.
- $(a_k)_{k \geq 0}$ is a homogeneous Markov chain taking values in $\{1, \dots, J\}$. We denote by Q the transition matrix of this Markov chain and by π its initial distribution under the risk-neutral probability \mathbb{Q} .
- $(H_j)_{1 \leq j \leq J}$ are positive-definite matrices.
- Z_1 is a Gaussian random vector of size m with mean μ_1 and variance Σ_1 and is independent of $(\varepsilon_k)_{k \geq 0}$.
- $(\eta_k)_{k \geq 0}$ is a i.i.d. sequence of standard Gaussian variables, independent of $(\varepsilon_k)_{k \geq 0}$ and Z_1 .
- $(G_j)_{1 \leq j \leq J}$ are $p \times p$ positive-definite matrices and $(B_j)_{1 \leq j \leq J}$ $m \times m$ matrices.

Define

$$\bar{G}_j := G_j G_j', \bar{H}_j := H_j H_j'$$

The unknown parameters are collectively referred to as θ , which is given by

$$\theta := \left\{ Q, \mu_1, \Sigma_1, (d_j, c_j, T_j, B_j, \bar{H}_j, \bar{G}_j)_{1 \leq j \leq J} \right\}.$$

Our objective is to estimate θ using the sequence of noisy observations of future prices Y_k , $1 \leq k \leq n$. Then, once the model is calibrated, it can be used to predict future prices or to detect regime switchings.

3.2 Contribution

In this section, we introduce our main contributions to address modeling, inference and forecasting issues in commodity markets. We expand the regime switching based two-filter algorithm proposed in [10]. In our extension, we propose a more flexible match between the forward and backward filters to approximate the posterior distribution of two consecutive states $(a_{k-1}, Z_{k-1}, a_k, Z_k)$ given the observations. The performance of this method has been numerically compared to the algorithm of [10], see more details in 5.3.

We applied our two-filter algorithm to approximate the E-step of an EM algorithm in order to estimate the parameters of the regime switching model (3.1) and (3.2) and for CME Crude Oil data. Then, the M-step of the EM algorithm is approximated using the Covariance Matrix Adaptation Evolution Strategy optimization method introduced in [47] (see also [46] and [45]), see in section 5.4. These two contributions are described in an article submitted to EURASIP Journal on Advances in Signal Processing and inserted in Chapter 5. Note that recently, [66] proposed a forward backward approach to approximate these smoothing distributions without introducing artificial distributions. They also took advantage of the tractable substructure of the model to introduce a new Rao-Blackwellized particle smoother. Although it does not rely on artificial distributions, their method does not require any approximation of the model.

3.2.1 Smoothing algorithm

Parameter estimation of regime switching models based only on the sequence of observations (Y_1, Y_2, \dots, Y_n) is a challenging statistical problem. Our algorithm relies on the Expectation Maximization (EM) algorithm which iteratively produces parameter estimates that increase the likelihood of the observations. As explained in Chapter 2, the computation of the intermediate quantity $\theta \mapsto Q(\theta, \theta_p)$ of the EM algorithm for a current parameter θ_p requires the conditional distributions of the states $(a_{i-1}, a_i, z_{i-1}, z_i)$ in (3.1) and (3.2) given the observations (Y_1, \dots, Y_n) for all $1 \leq i \leq n$,

written $\phi_i^n(a_{i-1:i}, z_{i-1:i})$.

Let $m(a_{i-1}, z_{i-1}; z_i)$ be the density of the conditional distribution of Z_i given (a_{i-1}, Z_{i-1}) and $g(a_i, z_i; y_i)$ be the density of the conditional distribution of Y_i given (a_i, Z_i) when the parameter value is θ_p (θ_p is dropped from notations for better clarity). Conditional on the sequence $(a_i)_{1 \leq i \leq n}$, (3.1) and (3.2) define a standard linear and Gaussian model. Using this specific structure, [10] proposed to use Rao-Blackwellisation combined with a two-filter algorithm to provide a Sequential Monte Carlo procedure to approximate $\phi_i^n(a_{i-1:i}, z_{i-1:i})$. [10] first introduced a family of artificial probability densities $(\gamma_i)_{1 \leq i \leq n}$ such that for all $1 \leq i \leq n$,

$$\prod_{k=i+1}^n Q(a_{k-1}, a_k) p(y_{i:n} | a_{i:n}, z_i) \propto \gamma_i^{-1}(a_i, z_i) \tilde{p}(a_{i:n}, z_i | y_{i:n}),$$

where $\tilde{p}(a_{i:n}, z_i | y_{i:n})$ is an artificial probability density on $\{1, \dots, J\}^{n-i+1} \times \mathbb{R}^m$ although $p(y_{i:n} | a_{i:n}, z_i)$ is not. Then, $\phi_i^n(a_{i-1:i}, z_{i-1:i})$ may be written as:

$$\begin{aligned} \phi_i^n(a_{i-1:i}, z_{i-1:i}) &\propto \gamma_i^{-1}(a_i, z_i) Q(a_{i-1}, a_i) m(a_{i-1}, z_{i-1}; z_i) \\ &\quad \times p(a_{i-1}, z_{i-1} | y_{1:i-1}) \tilde{p}(a_i, z_i | y_{i:n}). \end{aligned} \quad (3.3)$$

Then the algorithm of [10] goes as follows:

- (i) A forward filter approximates $p(a_{i-1} | y_{1:i-1})$ for all $i \in \{2, \dots, n\}$ with weighted particles $\{(a_{i-1}^k, \omega_{i-1}^k)\}_{k=1}^N$ using the mixture Kalman techniques introduced in [16]. Conditional on $\{(a_{i-1}^k, \omega_{i-1}^k)\}_{k=1}^N$ and $y_{1:i-1}$, Z_{i-1} has a Gaussian distribution with mean μ_i^k and covariance matrix P_i^k which may be updated on-the-fly. Then, the conditional density $p(a_{i-1}, z_{i-1} | y_{1:i-1})$ at time $i-1$ is approximated by:

$$p^N(a_{i-1}, z_{i-1} | y_{1:i-1}) := \sum_{k=1}^N \omega_{i-1}^k \delta_{a_{i-1}^k}(a_{i-1}) \varphi_{\mu_i^k, P_i^k}(z_{i-1}),$$

where $\varphi_{\mu_i^k, P_i^k}$ is the probability density of a Gaussian distribution with mean μ_i^k and covariance matrix P_i^k .

- (ii) A backward filter produces weighted samples $\{(\tilde{a}_{i:n}^\ell, \tilde{\omega}_i^\ell)\}_{k=1}^N$ for all $i \in \{1, \dots, n\}$ to approximate $\tilde{p}(a_i, z_i | y_{i:n})$ as a mixture of distributions using Kalman filtering techniques.

In [10], the approximation of the smoothing distribution combines directly both particle filters by plugging the approximations in (3.3). If these probability have their regions of high probability masses significantly disjoint, the resulting approximation cannot be expected to perform well, the approximation of the smoothing distribution is such that most particles have negligible weights.

To overcome this difficulty, we introduced the following marginalization of the bivariate smoothing distributions:

$$\phi_i^n(a_{i-1:i}, z_{i-1:i}) = \sum_{a_{i-2}} \sum_{a_{i+1}} \int_{z_{i-1}} \int_{z_{i+1}} \psi_i^n(a_{i-2:i+1}, z_{i-2:i+1}) dz_{i-2} dz_{i+1}. \quad (3.4)$$

where $\psi_i^n(a_{i-2:i+1}, z_{i-2:i+1})$ is the conditional distribution of the states and hidden regimes from time indices $i-2$ to $i+1$ given all the observations. Writing

$$r_i(a_{i-1}, z_{i-1}, y_{1:i-2}) = \sum_{a_{i-2}} \int_{z_{i-2}} p(a_{i-2}, z_{i-2} | y_{1:i-2}) m(a_{i-2}, z_{i-2}; z_{i-1}) Q(a_{i-2}, a_{i-1}) dz_{i-2}$$

and

$$t_i(a_i, z_i, y_{i:n}) = \sum_{a_{i+1}} \int_{z_{i+1}} m(a_i, z_i; z_{i+1}) Q(a_i, a_{i+1}) g(a_i, z_i; y_i) p(y_{i+1:n} | a_{i+1}, z_{i+1}) dz_{i+1}$$

and using (3.4) leads to the following expression for $\phi_i^n(a_{i-1:i}, z_{i-1:i})$:

$$\begin{aligned} \phi_i^n(a_{i-1:i}, z_{i-1:i}) &\propto r_i(a_{i-1}, z_{i-1}, y_{1:i-2}) [m(a_{i-1}, z_{i-1}; z_i) Q(a_{i-1}, a_i) g(a_{i-1}, z_{i-1}; y_{i-1})] \\ &\quad \times t_i(a_i, z_i, y_{i:n}), \end{aligned} \quad (3.5)$$

We propose to approximate the marginal smoothing distributions by merging forward and backward trajectories corresponding to hidden states up to time $i-2$ for the forward filter and time $i+1$ for the backward filter. Then, the computation of r_{i-1} and t_i requires to integrate out all possible paths between $i-2$ and $i-1$ and between i and $i+1$ time steps, respectively. This integration is available explicitly in our linear and Gaussian case and is shown to produce a more robust approximation than the algorithm of [10]. Following [10] our algorithm is decomposed into a forward pass to approximate r_i and a backward pass to approximate t_i using two-filter techniques. Both path are based on sequential Monte Carlo methods: the forward pass provides

a sequence of states $a_{1:i-2}^k$ associated with importance weights $\omega_{i-2}^k, 1 \leq k \leq N$ to estimate $p(a_{i-2}|y_{1:i-2})$. This provides an approximation of r_i which replaces step (i) of [10]. Another particle filter is used in the backward pass using artificial distributions to produce a sequence of backward trajectories $\tilde{a}_{i+1:n}^k$ associated with importance weights $\tilde{\omega}_{i-2}^k, 1 \leq k \leq N$ to estimate $p(y_{i+1:n}|a_{i+1}, z_{i+1})$ and then t_i which replaces step (ii) of [10]. These approximations are then plugged in (3.5) to obtain a particle based approximation of $\phi_i^n(a_{i-1:i}, z_{i-1:i})$. Two-filter methods performance depends heavily on the choice of artificial distributions. We proposed an artificial distribution defined as a mixture of Gaussian distributions based on the particles produced during the forward pass.

3.2.2 Statistical inference applied to crude oil data

The algorithm presented in the previous section has been applied to estimate the parameters of the regime switching hidden Markov models described by (3.1) and (3.1) using an EM algorithm with Crude oil data. We used the Wednesday future prices of 1, 4, 6 and 13 months from 11 January 1995 to 29 May 2013 in CME WTI crude oil market. These prices are considered as observations in our algorithm and used to estimate the posterior distribution of the state variables (spot price, convenience yield and regime label). Illustrations of the convergence of the estimation procedure are given in Chapter 5.

Chapter 4

On the two-filter approximations of marginal smoothing distributions in general state space models

Abstract

A prevalent problem in general state space models is the approximation of the smoothing distribution of a state conditional on the observations from the past, the present, and the future. The aim of this paper is to provide a rigorous analysis of such approximations of smoothed distributions provided by the two-filter algorithms. We extend the results available for the approximation of smoothing distributions to these two-filter approaches which combine a forward filter approximating the filtering distributions with a backward information filter approximating a quantity proportional to the posterior distribution of the state given future observations.

4.1 Introduction

State-space models play a key role in a large variety of disciplines such as engineering, econometrics, computational biology or signal processing, see [27, 26] and references therein. This paper provides a nonasymptotic analysis of a Sequential

Monte Carlo Method (SMC) which aims at performing optimal smoothing in nonlinear and non Gaussian state space models. Given two measurable spaces (X, \mathcal{X}) and (Y, \mathcal{Y}) , consider a bivariate stochastic process $\{(X_t, Y_t)\}_{t \geq 0}$ taking values in the product space $(X \times Y, \mathcal{X} \otimes \mathcal{Y})$, where the hidden state sequence $\{X_t\}_{t \geq 0}$ is observed only through the observation process $\{Y_t\}_{t \geq 0}$. Statistical inference in general state space models usually involves the computation of conditional distributions of some unobserved states given a set of observations. These posterior distributions are crucial to compute smoothed expectations of additive functionals which appear naturally for maximum likelihood parameter inference in hidden Markov models (computation of the Fisher score or of the intermediate quantity of the Expectation Maximization algorithm), see [14, Chapter 10 and 11], [53, 81, 62, 63].

Nevertheless, exact computation of the filtering and smoothing distributions is possible only for linear and Gaussian state spaces or when the state space X is finite. This paper focuses on particular instances of Sequential Monte Carlo methods which approximate sequences of distributions in a general state space X with random samples, named particles, associated with nonnegative importance weights. Those particle filters and smoothers rely on the combination of sequential importance sampling steps to propagate particles in the state space and importance resampling steps to duplicate or discard particles according to their importance weights. The first implementation of these SMC methods, introduced in [41, 59], propagates the particles using the Markov kernel of the hidden process $\{X_t\}_{t \geq 0}$ and uses a multinomial resampling step based on the importance weights to select particles at each time step. An interesting feature of this *Poor man's smoother* is that it provides an approximation of the joint smoothing distribution by storing the ancestral line of each particle with a complexity growing only linearly with the number N of particles, see for instance [20]. However, this smoothing algorithm has a major shortcoming since the successive resampling steps induce an important depletion of the particle trajectories. This degeneracy of the particle sequences leads to trajectories sharing a common ancestor path; see [81, 50] for a discussion.

Approximations of the smoothing distributions may also be obtained using the forward filtering backward smoothing decomposition in general state space models.

The Forward Filtering Backward Smoothing algorithm (FFBS) and the Forward Filtering Backward Simulation algorithm (FFBSi) developed respectively in [59, 49, 28] and [40] avoid the path degeneracy issue of the *Poor man's smoother* at the cost of a computational complexity growing with N^2 . Both algorithms rely on a forward pass which produces a set of particles and weights approximating the sequence of filtering distributions up to time T . Then, the backward pass of the FFBS algorithm modifies all the weights computed in the forward pass according to the so-called backward decomposition of the smoothing distribution keeping all the particles fixed. On the other hand, the FFBSi algorithm samples independently particle trajectories among all the possible paths produced by the forward pass. It is shown in [73, 13, 22] that the FFBS algorithm can be implemented using only a forward pass when approximating smoothed expectations of additive functionals but with a complexity still growing quadratically with N . Under the mild assumption that the transition density of the hidden chain $\{X_t\}_{t \geq 0}$ is uniformly bounded above, [24] proposed an accept-reject mechanism to implement the FFBSi algorithm with a complexity growing only linearly with N . Concentration inequalities, controls of the L_q -norm of the deviation between smoothed functionals and their approximations and Central Limit Theorems (CLT) for the FFBS and the FFBSi algorithms have been established in [22, 24, 31].

Recently, [78] proposed a new SMC algorithm, the particle-based rapid incremental smoother (PaRis), to approximate online, using only a forward pass, smoothed expectations of additive functionals. The crucial feature of this algorithm is that its complexity grows only linearly with N as it samples on-the-fly particles distributed according to the backward dynamics of the hidden chain conditionally on the observations Y_0, \dots, Y_T . The authors show concentration inequalities and CLT for the estimators provided by the PaRis algorithm.

In this paper, we extend the theoretical results available for the SMC approximations of smoothing distributions to the estimators given by the two-filter algorithms. These methods were first introduced in the particle filter literature by [59] and developed further by [10] and [34]. The two-filter approach combines the output of two independent filters, one that evolves forward in time and approximates the filtering

distributions and another that evolves backward in time approximating a quantity proportional to the posterior distribution of a state given future observations. In [34], the authors introduced a proposal mechanism leading to algorithms whose complexity grows linearly with the number of particles. An algorithm similar to the algorithm of [10] may also be implemented with an $\mathcal{O}(N)$ computational complexity following the same idea. We analyze all these algorithms which approximate the marginal smoothing distributions (smoothing distributions of one state given all the observations) and provide concentration inequalities as well as CLT.

This paper is organized as follows. Section 4.2 introduces the different particle approximations of the marginal smoothing distributions given by the two-filter algorithms. Sections 4.3 and 4.4 provide exponential deviation inequalities and CLT for the particle approximations under mild assumptions on the hidden Markov chain. Under additional *strong mixing assumptions*, it is shown that the results of Section 4.3 are uniform in time and that the asymptotic variance in Section 4.4 may be uniformly bounded in time. All proofs are postponed to Section 4.5.

Notations and conventions

Let X and Y be two general state-spaces endowed with countably generated σ -fields \mathcal{X} and \mathcal{Y} . $\mathbb{F}_b(X, \mathcal{X})$ is the set of all real valued bounded measurable functions on (X, \mathcal{X}) . Q is a Markov transition kernel defined on $X \times \mathcal{X}$ and $\{g_i\}_{i \geq 0}$ a family of positive functions defined on X . For any $x \in X$, $Q(x, \cdot)$ has a density $q(x, \cdot)$ with respect to a measure λ on (X, \mathcal{X}) . The oscillation of a real valued function defined on a space Z is given by: $\text{osc}(h) := \sup_{z, z' \in Z} |h(z) - h(z')|$.

4.2 The two-filter algorithms

For any measurable function h on X^{t-s+1} , probability distribution χ on (X, \mathcal{X}) , $T \geq 0$ and $0 \leq s \leq t \leq T$, define the joint smoothing distribution by:

$$\phi_{\chi, s:t|T}[h] := \frac{\int \chi(dx_0) g_0(x_0) \prod_{u=1}^T Q(x_{u-1}, dx_u) g_u(x_u) h(x_{s:t})}{\int \chi(dx_0) g_0(x_0) \prod_{u=1}^T Q(x_{u-1}, dx_u) g_u(x_u)}, \quad (4.1)$$

where $a_{u:v}$ is a short-hand notation for $\{a_s\}_{s=u}^v$. In the following we use the notations $\phi_{\chi,s|T} := \phi_{\chi,s:s|T}$ and $\phi_{\chi,t} := \phi_{\chi,t:t|t}$. The aim of this paper is to provide a rigorous analysis of the performance of SMC algorithms approximating the sequence $\phi_{\chi,s|T}$ for $0 \leq s \leq T$. The algorithms analyzed in this paper are based on the *two-filter formula* introduced in [10, 34], which we now detail.

4.2.1 Forward filter

Let $\{\xi_0^\ell\}_{\ell=1}^N$ be i.i.d. and distributed according to the instrumental distribution ρ_0 and define the importance weights

$$\omega_0^\ell := \frac{d\chi}{d\rho_0}(\xi_0^\ell) g_0(\xi_0^\ell).$$

For any $h \in \mathbb{F}_b(X, \mathcal{X})$,

$$\phi_{\chi,0}^N[h] := \Omega_0^{-1} \sum_{\ell=1}^N \omega_0^\ell h(\xi_0^\ell), \quad \text{where} \quad \Omega_0 := \sum_{\ell=1}^N \omega_0^\ell,$$

is a consistent estimator of $\phi_{\chi,0}[h]$, see for instance [20]. Then, based on $\{(\xi_{s-1}^\ell, \omega_{s-1}^\ell)\}_{\ell=1}^N$ a new set of particles and importance weights is obtained using the auxiliary sampler introduced in [80]. Pairs $\{(I_s^\ell, \xi_s^\ell)\}_{\ell=1}^N$ of indices and particles are simulated independently from the instrumental distribution with density on $\{1, \dots, N\} \times X$:

$$\pi_{s|s}(\ell, x) \propto \omega_{s-1}^\ell \vartheta_s(\xi_{s-1}^\ell) p_s(\xi_{s-1}^\ell, x), \quad (4.2)$$

where ϑ_s is the adjustment multiplier weight function and p_s is a Markovian transition density. For any $\ell \in \{1, \dots, N\}$, ξ_s^ℓ is associated with the importance weight defined by:

$$\omega_s^\ell := \frac{q(\xi_{s-1}^{I_s^\ell}, \xi_s^\ell) g_s(\xi_s^\ell)}{\vartheta_s(\xi_{s-1}^{I_s^\ell}) p_s(\xi_{s-1}^{I_s^\ell}, \xi_s^\ell)} \quad (4.3)$$

to produce the following approximation of $\phi_{\chi,s}[h]$:

$$\phi_{\chi,s}^N[h] := \Omega_s^{-1} \sum_{\ell=1}^N \omega_s^\ell h(\xi_s^\ell), \quad \text{where} \quad \Omega_s := \sum_{\ell=1}^N \omega_s^\ell.$$

4.2.2 Backward filter

Let $\{\gamma_t\}_{t \geq 0}$ be a family of positive measurable functions such that, for all $t \in \{0, \dots, T\}$,

$$\int \gamma_t(x_t) dx_t \left[\prod_{u=t+1}^T g_{u-1}(x_{u-1}) Q(x_{u-1}, dx_u) \right] g_T(x_T) < \infty. \quad (4.4)$$

Following [10], for any $0 \leq t \leq T$ we introduce the backward filtering distribution $\psi_{\gamma, t|T}$ on \mathcal{X} (referred to as the *backward information filter* in [59] and [10]) defined, for any $h \in \mathbb{F}_b(\mathcal{X}, \mathcal{X})$, by:

$$\psi_{\gamma, t|T}[h] := \frac{\int \gamma_t(x_t) dx_t \left[\prod_{u=t+1}^T g_{u-1}(x_{u-1}) Q(x_{u-1}, dx_u) \right] g_T(x_T) h(x_t)}{\int \gamma_t(x_t) dx_t \left[\prod_{u=t+1}^T g_{u-1}(x_{u-1}) Q(x_{u-1}, dx_u) \right] g_T(x_T)}.$$

If the distribution of X_t has probability density function γ_t , then $\psi_{\gamma, t|T}$ is the conditional distribution of X_t given $Y_{t:T}$. Contrary to [10] or [34], $\int \gamma_t(x_t) dx_t$ may be infinite. The only requirement about the nonnegative functions $\{\gamma_t\}_{t \geq 0}$ is the condition (4.4) and the fact that γ_t should be available in closed form. Here γ_t is a possibly improper prior introduced to make $\psi_{\gamma, t|T}$ a proper posterior distribution, which is of key importance when producing particle approximations of such quantities. For $0 \leq t \leq T-1$, the backward information filter is computed by the recursion

$$\psi_{\gamma, t|T}[h] \propto \int \psi_{\gamma, t+1|T}(dx_{t+1}) \left[\gamma_t(x_t) g_t(x_t) \frac{q(x_t, x_{t+1})}{\gamma_{t+1}(x_{t+1})} \right] h(x_t) dx_t, \quad (4.5)$$

in the backward time direction. (4.5) is analogous to the forward filter recursion and particle approximations of the backward information filter can be obtained similarly. Using the definition of the forward filtering distribution at time $s-1$ and the backward information filter at time $s+1$, the marginal smoothing distribution may be expressed as

$$\phi_{\chi, s|T}[h] \propto \int \phi_{\chi, s-1}(dx_{s-1}) \psi_{\gamma, s+1|T}(dx_{s+1}) q(x_{s-1}, x_s) g_s(x_s) \frac{q(x_s, x_{s+1})}{\gamma_{s+1}(x_{s+1})} h(x_s) dx_s. \quad (4.6)$$

We now describe the Sequential Monte Carlo methods used to approximate the recursion (4.5) in [10], [34]. Let $\check{\rho}_T$ be an instrumental probability density on \mathcal{X} and $\{\check{\xi}_{T|T}^i\}_{i=1}^N$ be i.i.d. random variables such that $\check{\xi}_{T|T}^i \sim \check{\rho}_T$ and define

$$\check{\omega}_{T|T}^i := \frac{g_T(\check{\xi}_{T|T}^i) \gamma_T(\check{\xi}_{T|T}^i)}{\check{\rho}_T(\check{\xi}_{T|T}^i)}.$$

Let now $\{(\check{\xi}_{t+1|T}^i, \check{\omega}_{t+1|T}^i)\}_{i=1}^N$ be a weighted sample targeting the backward information filter distribution $\psi_{\gamma, t+1|T}[h]$ at time $t+1$:

$$\psi_{\gamma, t+1|T}^N[h] := \check{\Omega}_{t+1|T}^{-1} \sum_{i=1}^N \check{\omega}_{t+1|T}^i h(\check{\xi}_{t+1|T}^i), \quad \text{where} \quad \check{\Omega}_{t+1|T} := \sum_{i=1}^N \check{\omega}_{t+1|T}^i.$$

Plugging this approximation into (4.5) yields the target probability density

$$\hat{\psi}_{\gamma, t|T}^{\text{tar}}(x_t) \propto \sum_{i=1}^N \check{\omega}_{t+1|T}^i \left[\gamma_t(x_t) g_t(x_t) \frac{q(x_t, \check{\xi}_{t+1|T}^i)}{\gamma_{t+1}(\check{\xi}_{t+1|T}^i)} \right],$$

which is the marginal probability density function of x_t of the joint density

$$\hat{\psi}_{\gamma, t|T}^{\text{aux}}(i, x_t) \propto \frac{\check{\omega}_{t+1|T}^i}{\gamma_{t+1}(\check{\xi}_{t+1|T}^i)} \gamma_t(x_t) g_t(x_t) q(x_t, \check{\xi}_{t+1|T}^i).$$

A particle approximation of the backward information filter at time t can be derived by choosing an adjustment weight function $\vartheta_{t|T}$ and an instrumental density kernel $r_{t|T}$, and simulating $\{(\check{I}_t^i, \check{\xi}_{t|T}^i)\}_{i=1}^N$ from the instrumental probability density on $\{1, \dots, N\} \times X$ given by

$$\pi_{t|T}(i, x_t) \propto \frac{\check{\omega}_{t+1|T}^i \vartheta_{t|T}(\check{\xi}_{t+1|T}^i)}{\gamma_{t+1}(\check{\xi}_{t+1|T}^i)} r_{t|T}(\check{\xi}_{t+1|T}^i, x_t). \quad (4.7)$$

Subsequently, the particles are associated with the importance weights

$$\check{\omega}_{t|T}^i := \frac{\gamma_t(\check{\xi}_{t|T}^i) g_t(\check{\xi}_{t|T}^i) q(\check{\xi}_{t|T}^i, \check{\xi}_{t+1|T}^i)}{\vartheta_{t|T}(\check{\xi}_{t+1|T}^i) r_{t|T}(\check{\xi}_{t+1|T}^i, \check{\xi}_{t|T}^i)}. \quad (4.8)$$

Ideally, a fully adapted version of the auxiliary backward information filter is obtained by using the adjustment weights $\vartheta_{t|T}^*(x) = \int \gamma_t(x_t) g_t(x_t) q(x_t, x) dx_t$ and the proposal kernel density

$$r_{t|T}^*(x, x_t) = \gamma_t(x_t) g_t(x_t) \frac{q(x_t, x)}{\vartheta_{t|T}^*(x)},$$

yielding uniform importance weights. Such a solution is most likely to be cumbersome from a computational perspective.

4.2.3 Two-filter approximations of the marginal smoothing distributions

Plugging the particle approximations of the forward and backward filter distributions into (4.6) provides the following mixture approximation of the smoothing distribution:

$$\hat{\phi}_{\chi, s|T}^{\text{tar}}(x_s) \propto \sum_{i=1}^N \sum_{j=1}^N \frac{\omega_{s-1}^i \check{\omega}_{s+1|T}^j}{\gamma_{s+1}(\check{\xi}_{s+1|T}^j)} q(\xi_{s-1}^i, x_s) g_s(x_s) q(x_s, \check{\xi}_{s+1|T}^j). \quad (4.9)$$

Following the $\text{TwoFilt}_{f_{wt}}$ algorithm of Fearnhead, Wyncoll and Tawn [34], the probability density (4.9) might be seen as the marginal density of x_s obtained from the joint density on the product space $\{1, \dots, N\}^2 \times X$ given by

$$\hat{\phi}_{\chi, s|T}^{\text{aux}}(i, j, x_s) \propto \frac{\omega_{s-1}^i \check{\omega}_{s+1|T}^j}{\gamma_{s+1}(\check{\xi}_{s+1|T}^j)} q(\xi_{s-1}^i, x_s) g_s(x_s) q(x_s, \check{\xi}_{s+1|T}^j). \quad (4.10)$$

The $\text{TwoFilt}_{f_{wt}}$ algorithm draws a set $\{(I_s^\ell, \check{I}_s^\ell, \check{\xi}_{s|T}^\ell)\}_{\ell=1}^N$ of indices and particle positions from the instrumental density

$$\pi_{s|T}(i, j, x_s) \propto \frac{\omega_{s-1}^i \tilde{\vartheta}_{s|T}(\xi_{s-1}^i, \check{\xi}_{s+1|T}^j) \check{\omega}_{s+1|T}^j}{\gamma_{s+1}(\check{\xi}_{s+1|T}^j)} \tilde{r}_{s|T}(\xi_{s-1}^i, \check{\xi}_{s+1|T}^j; x_s), \quad (4.11)$$

where, as above, $\tilde{\vartheta}_{s|T}(x, x')$ is an adjustment multiplier weight function (which now depends on the forward and backward particles) and $\tilde{r}_{s|T}$ is an instrumental kernel. We then associate with each draw $(I_s^\ell, \check{I}_s^\ell, \check{\xi}_{s|T}^\ell)$ the importance weight

$$\tilde{\omega}_{s|T}^\ell := \frac{q(\xi_{s-1}^{I_s^\ell}, \check{\xi}_{s|T}^{\check{I}_s^\ell}) g_s(\check{\xi}_{s|T}^{\check{I}_s^\ell}) q(\check{\xi}_{s|T}^{\check{I}_s^\ell}, \check{\xi}_{s+1|T}^{\check{I}_s^\ell})}{\tilde{\vartheta}_{s|T}(\xi_{s-1}^{I_s^\ell}, \check{\xi}_{s+1|T}^{\check{I}_s^\ell}) \tilde{r}_{s|T}(\xi_{s-1}^{I_s^\ell}, \check{\xi}_{s+1|T}^{\check{I}_s^\ell}; \check{\xi}_{s|T}^{\check{I}_s^\ell})}, \quad \tilde{\Omega}_{s|T} := \sum_{\ell=1}^N \tilde{\omega}_{s|T}^\ell. \quad (4.12)$$

Then, the auxiliary indices $\{(I_s^\ell, \check{I}_s^\ell)\}_{\ell=1}^N$ are discarded and $\{(\tilde{\omega}_{s|T}^\ell, \check{\xi}_{s|T}^\ell)\}_{\ell=1}^N$ approximate the target smoothing density $\hat{\phi}_{\chi, s|T}^{\text{tar}}$. Mimicking the arguments in [49] and further developed in [61], the auxiliary particle filter is fully adapted if the adjustment weight function is $\vartheta_{s|T}^*(x, x') = \int q(x, x_s) g_s(x_s) q(x_s, x') dx_s$ and the instrumental kernel is

$$r_{s|T}^*(x, x'; x_s) = q(x, x_s) g_s(x_s) q(x_s, x') / \vartheta_{s|T}^*(x, x').$$

Except in simple scenarios, simulating from the fully adapted auxiliary filter is computationally intractable.

Instead of considering the target distribution (4.9) as the marginal of the auxiliary distribution (4.10) over pairs of indices, the TwoFilt_{bdm} algorithm of [10] uses the following *partial* auxiliary distributions having densities,

$$\begin{aligned}\hat{\phi}_{s|T}^{\text{aux},f}(i, x_s) &\propto \omega_{s-1}^i q(\xi_{s-1}^i, x_s) g_s(x_s) \sum_{j=1}^N \frac{\check{\omega}_{s+1|T}^j}{\gamma_{s+1}(\check{\xi}_{s+1|T}^j)} q(x_s, \check{\xi}_{s+1|T}^j), \\ \hat{\phi}_{s|T}^{\text{aux},b}(j, x_s) &\propto \frac{\check{\omega}_{s+1|T}^j}{\gamma_{s+1}(\check{\xi}_{s+1|T}^j)} q(x_s, \check{\xi}_{s+1|T}^j) g_s(x_s) \sum_{i=1}^N \omega_{s-1}^i q(\xi_{s-1}^i, x_s).\end{aligned}$$

Since $\hat{\phi}_{\chi, s|T}^{\text{tar}}$ is the marginal probability density of the partial auxiliary distributions $\hat{\phi}_{s|T}^{\text{aux},f}$ and $\hat{\phi}_{s|T}^{\text{aux},b}$ with respect to the forward and the backward particle indices, respectively, we may sample from $\hat{\phi}_{\chi, s|T}^{\text{tar}}$ by simulating instead $\{(I_s^\ell, \xi_s^\ell)\}_{\ell=1}^N$ or $\{(\check{I}_s^\ell, \check{\xi}_s^\ell)\}_{\ell=1}^N$ from the instrumental probability density functions

$$\begin{aligned}\pi_{s|T}^f(i, x_s) &\propto \omega_{s-1}^i \vartheta_s(\xi_{s-1}^i) p_s(\xi_{s-1}^i, x_s), \\ \pi_{s|T}^b(j, x_s) &\propto \vartheta_{s|T}(\check{\xi}_{s+1|T}^j) \check{\omega}_{s+1|T}^j r_{s|T}(\check{\xi}_{s+1|T}^j, x_s) / \gamma_{s+1}(\check{\xi}_{s+1|T}^j),\end{aligned}$$

where (ϑ_s, p_s) and $(\vartheta_{s+1|T}, r_{s|T})$ are the adjustment multiplier weight functions and the instrumental kernels used in the forward and backward passes. In this case the algorithm uses the particles obtained when approximating the forward filter and backward information filter to provide two different weighted samples $\{(\tilde{\omega}_{s|T}^{i,f}, \xi_s^i)\}_{i=1}^N$ and $\{(\tilde{\omega}_{s|T}^{i,b}, \check{\xi}_s^i)\}_{i=1}^N$ targeting the marginal smoothing distribution, where the *forward* $\{\tilde{\omega}_{s|T}^{i,f}\}_{i=1}^N$ and *backward* $\{\tilde{\omega}_{s|T}^{i,b}\}_{i=1}^N$ importance weights are given by

$$\tilde{\omega}_{s|T}^{i,f} := \omega_s^i \sum_{j=1}^N \check{\omega}_{s+1|T}^j q(\xi_s^i, \check{\xi}_{s+1|T}^j) / \gamma_{s+1}(\check{\xi}_{s+1|T}^j), \quad \tilde{\Omega}_{s|T}^f := \sum_{j=1}^N \check{\omega}_{s|T}^{j,f}, \quad (4.13)$$

$$\tilde{\omega}_{s|T}^{j,b} := \check{\omega}_{s|T}^j \sum_{i=1}^N \omega_{s-1}^i q(\xi_{s-1}^i, \check{\xi}_s^j) / \gamma_s(\check{\xi}_s^j), \quad \tilde{\Omega}_{s|T}^b := \sum_{j=1}^N \check{\omega}_{s|T}^{j,b}. \quad (4.14)$$

An important drawback of these algorithms is that the computation of the forward and backward importance weights grows quadratically with the number N of particles.

4.2.4 $\mathcal{O}(N)$ approximations of the marginal smoothing distributions

In [34], the authors introduced a proposal mechanism in (4.11) such that the indices (I_s, \check{I}_s) of the forward and backward particles chosen at time $s-1$ and $s+1$ are sampled independently. Such choices lead to algorithms whose complexity grows linearly with the number of particles. The $\mathcal{O}(N)$ algorithm displayed in [34] suggests to use an adjustment multiplier weight function in (4.11) such that I_s and \check{I}_s are chosen according to the same distributions as the indices sampled in the forward filter and in the backward information filter. It is done in [34] by choosing $\tilde{\vartheta}_{s|T}(x, x') = \vartheta_s(x)\vartheta_{s|T}(x')$ so that (4.11) becomes

$$\pi_{s|T}(i, j, x_s) \propto \omega_{s-1}^i \vartheta_s(\xi_{s-1}^i) \frac{\vartheta_{s|T}(\check{\xi}_{s+1|T}^j) \tilde{\omega}_{s+1|T}^j}{\gamma_{s+1}(\check{\xi}_{s+1|T}^j)} \tilde{r}_{s|T}(\xi_{s-1}^i, \check{\xi}_{s+1|T}^j; x_s). \quad (4.15)$$

In this case, the importance weight (4.12) associated with each draw $(I_s^\ell, \check{I}_s^\ell, \check{\xi}_{s|T}^\ell)$ is given by

$$\tilde{\omega}_{s|T}^\ell := \frac{q(\xi_{s-1}^{I_s^\ell}, \check{\xi}_{s|T}^\ell) g_s(\check{\xi}_{s|T}^\ell) q(\check{\xi}_{s|T}^\ell, \check{\xi}_{s+1|T}^{\check{I}_s^\ell})}{\vartheta_s(\xi_{s-1}^{I_s^\ell}) \vartheta_{s|T}(\check{\xi}_{s+1|T}^{\check{I}_s^\ell}) \tilde{r}_{s|T}(\xi_{s-1}^{I_s^\ell}, \check{\xi}_{s+1|T}^{\check{I}_s^\ell}; \check{\xi}_{s|T}^\ell)}. \quad (4.16)$$

Instead of sampling new particles at time s , an algorithm similar to the `TwoFiltbdm` algorithm of [10] which uses the forward particles $\{\xi_{s-1}^{\ell=1}^N\}$ or backward particles $\{\check{\xi}_{s|T}^{\ell=1}^N\}$ may also be implemented with an $\mathcal{O}(N)$ computational complexity.

(a) Choosing $\tilde{r}_{s|T}(\xi_{s-1}^{I_s^\ell}, \check{\xi}_{s+1|T}^{\check{I}_s^\ell}; x_s) = r_{s|T}(\check{\xi}_{s+1|T}^{\check{I}_s^\ell}, x_s)$ in (4.15), the smoothing distribution approximation is obtained by reweighting the particles obtained in the backward pass. The backward particles $\{\check{\xi}_{s|T}^{\ell=1}^N\}$ are associated with the importance weights:

$$\begin{aligned} \tilde{\omega}_{s|T}^\ell &:= \frac{\gamma_s(\check{\xi}_{s|T}^\ell) g_s(\check{\xi}_{s|T}^\ell) q(\check{\xi}_{s|T}^\ell, \check{\xi}_{s+1|T}^{\check{I}_s^\ell})}{\vartheta_{s|T}(\check{\xi}_{s+1|T}^{\check{I}_s^\ell}) r_{s|T}(\check{\xi}_{s+1|T}^{\check{I}_s^\ell}, \check{\xi}_{s|T}^\ell)} \frac{q(\xi_{s-1}^{I_s^\ell}, \check{\xi}_{s|T}^\ell)}{\gamma_s(\check{\xi}_{s|T}^\ell) \vartheta_s(\xi_{s-1}^{I_s^\ell})}, \\ &= \tilde{\omega}_{s|T}^\ell \frac{q(\xi_{s-1}^{I_s^\ell}, \check{\xi}_{s|T}^\ell)}{\gamma_s(\check{\xi}_{s|T}^\ell) \vartheta_s(\xi_{s-1}^{I_s^\ell})}. \end{aligned} \quad (4.17)$$

- (b) Choosing $\tilde{r}_{s|T}(\xi_{s-1}^{I_s^\ell}, \check{\xi}_{s+1|T}^{\check{I}_s^\ell}; x_s) = p_s(\xi_{s-1}^{I_s^\ell}, x_s)$ in (4.15), the smoothing distribution approximation is obtained by reweighting the particles obtained in the forward filtering pass. The forward particles $\{\xi_s^\ell\}_{\ell=1}^N$ are associated with the importance weights:

$$\tilde{\omega}_{s|T}^\ell := \frac{q(\xi_{s-1}^{I_s^\ell}, \xi_s^\ell) g_s(\xi_s^\ell)}{\vartheta_s(\xi_{s-1}^{I_s^\ell}) p_s(\xi_{s-1}^{I_s^\ell}, \xi_s^\ell)} \frac{q(\xi_s^\ell, \check{\xi}_{s+1|T}^{\check{I}_s^\ell})}{\vartheta_{s|T}(\check{\xi}_{s+1|T}^{\check{I}_s^\ell})} = \omega_s^\ell \frac{q(\xi_s^\ell, \check{\xi}_{s+1|T}^{\check{I}_s^\ell})}{\vartheta_{s|T}(\check{\xi}_{s+1|T}^{\check{I}_s^\ell})}. \quad (4.18)$$

4.3 Exponential deviation inequality for the two-filter algorithms

In this section, we establish exponential deviation inequalities for the two-filter algorithms introduced in Section 4.2. Before stating the results, some additional notations are required. Define, for all $(x, x', x'') \in X^3$,

$$q^{[2]}(x, x'; x'') = q(x, x'')q(x'', x')$$

and for any functions $f : X^2 \rightarrow \mathbb{R}$ and $g : X \rightarrow \mathbb{R}$,

$$f \odot g(x, x') := f(x, x')g(x').$$

Consider the following assumptions:

A1. $|q|_\infty < \infty$ and for all $0 \leq t \leq T$, g_t is positive and $|g_t|_\infty < \infty$.

A2. For all $0 \leq t \leq T$, $|\vartheta_t|_\infty < \infty$, $|p_t|_\infty < \infty$ and $|\omega_t|_\infty < \infty$ where

$$\omega_0(x) := \frac{d\chi}{d\rho_0}(x)g_0(x) \quad \text{and for all } t \geq 1 \quad \omega_t(x, x') := \frac{q(x, x')g_t(x')}{\vartheta_t(x)p_t(x, x')}.$$

A3. - For all $0 \leq t \leq T - 1$, $|\vartheta_{t|T}/\gamma_{t+1}|_\infty < \infty$ and $|r_{t|T}|_\infty < \infty$. For all $0 \leq t \leq T$ $|\check{\omega}_{t|T}|_\infty < \infty$, where

$$\check{\omega}_{T|T}(x) := \frac{g_T(x)\gamma_T(x)}{\check{\rho}_T(x)} \quad \text{and for all } 0 \leq t < T, \quad \check{\omega}_{t|T}(x, x') := \frac{\gamma_t(x)g_t(x)q(x, x')}{\vartheta_{t|T}(x')r_{t|T}(x', x)}.$$

- For all $1 \leq t \leq T-1$, $|\tilde{\vartheta}_{t|T} \odot \gamma_{t+1}^{-1}|_\infty < \infty$, $|q \odot \gamma_{t+1}^{-1}|_\infty < \infty$, $|\tilde{\omega}_{t|T}|_\infty < \infty$ and $|\tilde{r}_{t|T}|_\infty < \infty$ where

$$\tilde{\omega}_{t|T}(x, x'; x'') := \frac{q^{[2]}(x, x'; x'')g_s(x'')}{\tilde{\vartheta}_{t|T}(x, x'')\tilde{r}_{t|T}(x, x'; x'')}.$$

We first show that the weighted sample $\{(\omega_s^i \tilde{\omega}_{t|T}^j), (\xi_s^i, \check{\xi}_{t|T}^j)\}_{i,j=1}^N$ targets the product distribution $\phi_{\chi,s} \otimes \psi_{\gamma,t|T}$.

Theorem 1. Assume that A1, A2 and A3 hold for some $T < \infty$. Then, for all $0 \leq s < t \leq T$, there exist $0 < B_{s,t|T}, C_{s,t|T} < \infty$ such that for all $N \geq 1$, $\epsilon > 0$ and all $h \in \mathbb{F}_b(\mathcal{X} \times \mathcal{X}, \mathcal{X} \otimes \mathcal{X})$,

$$\mathbb{P}\left(\left|\sum_{i,j=1}^N \frac{\omega_s^i \tilde{\omega}_{t|T}^j}{\tilde{\Omega}_{t|T}} h(\xi_s^i, \check{\xi}_{t|T}^j) - \phi_{\chi,s} \otimes \psi_{\gamma,t|T}[h]\right| > \epsilon\right) \leq B_{s,t|T} e^{-C_{s,t|T} N \epsilon^2 / \text{osc}^2(h)}.$$

Proof. The proof is postponed to Section 4.5.1. \square

We now study the weighted sample $\{(\tilde{\omega}_{s|T}^i, \tilde{\xi}_{s|T}^\ell)\}_{\ell=1}^N$ produced by the TwoFilt_{fwt} algorithm of Fearnhead, Wyncoll and Tawn [34] defined in (4.11) and (4.12) and targeting the marginal smoothing distribution $\phi_{\chi,s|T}$.

Theorem 2 (deviation inequality for TwoFilt_{fwt} of [34]). Assume that A1, A2 and A3 hold for some $T < \infty$. Then, for all $s < T$, there exist $0 < B_{s|T}, C_{s|T} < \infty$ such that for all $N \geq 1$, $\epsilon > 0$ and all $h \in \mathbb{F}_b(\mathcal{X}, \mathcal{X})$,

$$\mathbb{P}\left(\left|\sum_{i=1}^N \frac{\tilde{\omega}_{s|T}^i}{\tilde{\Omega}_{s|T}} h(\tilde{\xi}_{s|T}^i) - \phi_{\chi,s|T}[h]\right| > \epsilon\right) \leq B_{s|T} e^{-C_{s|T} N \epsilon^2 / \text{osc}^2(h)}.$$

Proof. The proof is postponed to Section 4.5.2. \square

Using Theorem 1 and Lemma 7, we may derive an exponential inequality for the weighted samples $\{(\xi_s^i, \tilde{\omega}_{s|T}^{i,f})\}_{i=1}^N$ and $\{(\check{\xi}_{s|T}^i, \tilde{\omega}_{s|T}^{i,b})\}_{i=1}^N$ produced by the TwoFilt_{bdtm} algorithm of [10], where $\tilde{\omega}_{s|T}^{i,f}$ and $\tilde{\omega}_{s|T}^{i,b}$ are defined in (4.13) and (4.14). Therefore, both the forward and the backward particle approximations of the smoothing distribution converge to the marginal smoothing distribution, and these two approximations satisfy an exponential inequality.

Theorem 3 (deviation inequality for the TwoFilt_{bdm} algorithm of [10]). Assume that A1, A2 and A3 hold for some $T < \infty$. Then, for all $1 \leq s \leq T - 1$, there exist $0 < B_{s|T}, C_{s|T} < \infty$ such that for all $N \geq 1$, $\epsilon > 0$ and all $h \in \mathbb{F}_b(X, \mathcal{X})$,

$$\mathbb{P} \left(\left| \sum_{i=1}^N \frac{\tilde{\omega}_{s|T}^{i,f}}{\bar{\Omega}_{s|T}^f} h(\xi_s^i) - \phi_{\chi, s|T}[h] \right| > \epsilon \right) \leq B_{s|T} e^{-C_{s|T} N \epsilon^2 / \text{osc}^2(h)}, \quad (4.19)$$

$$\mathbb{P} \left(\left| \sum_{i=1}^N \frac{\tilde{\omega}_{s|T}^{i,b}}{\bar{\Omega}_{s|T}^b} h(\check{\xi}_{s|T}^i) - \phi_{\chi, s|T}[h] \right| > \epsilon \right) \leq B_{s|T} e^{-C_{s|T} N \epsilon^2 / \text{osc}^2(h)}. \quad (4.20)$$

Proof. The proof is postponed to Section 4.5.3. \square

Remark 1. Following [78, 24, 31], time uniform exponential inequalities for the two-filter approximations of the marginal smoothing distributions may be obtained using *strong mixing* assumptions which are standard in the SMC literature:

H1. There exist $0 < \sigma_- < \sigma_+ < \infty$ and $c_- > 0$ such that for all $x, x' \in X$, $\sigma_- \leq q(x, x') \leq \sigma_+$ and for all $t \geq 0$,

$$\int \chi(dx_0) g_0(x_0) \geq c_- \quad \text{and} \quad \inf_{x \in X} \int Q(x, dx') g_t(x') \geq c_-.$$

H2. There exist $0 < \gamma_- < \gamma_+ < \infty$ and $\check{c}_- > 0$ such that for all $x \in X$ and all $t \geq 0$, $\gamma_- \leq \gamma_t(x) \leq \gamma_+$ and for all $t \geq 0$,

$$\int \gamma_T(x_T) g_T(x_T) dx_T \geq \check{c}_- \quad \text{and} \quad \inf_{x \in X} \int \gamma_t(x_t) g_t(x_t) q(x_t, x) \gamma_{t+1}^{-1}(x) dx_t \geq \check{c}_-.$$

(i) If A1 and A2 hold uniformly in T and if H1 holds, then, it is proved in [24] that Proposition 8 holds with constants that are uniform in time : there exist $0 < B, C < \infty$ such that for all $s \geq 0$, $N > 0$, $\epsilon > 0$ and all $h \in \mathbb{F}_b(X, \mathcal{X})$,

$$\mathbb{P} \left(\left| \Omega_s^{-1} \sum_{i=1}^N \omega_s^i h(\xi_s^i) - \phi_{\chi, s}[h] \right| \geq \epsilon \right) \leq B e^{-C N \epsilon^2 / \text{osc}(h)^2}.$$

(ii) It can be shown following the exact same steps that if A1 and A3 hold uniformly in T and if H1 and H2 hold then Proposition 9 holds with constants that are

uniform in time: there exist $0 < B, C < \infty$ such that for all $t \geq 0$, $N \geq 1$, $\epsilon > 0$, and all $h \in \mathbb{F}_b(\mathbb{X}, \mathcal{X})$,

$$\mathbb{P} \left[\left| \check{\Omega}_{t|T}^{-1} \sum_{i=1}^N \check{\omega}_{t|T}^i h(\check{\xi}_{t|T}^i) - \psi_{\gamma, t|T}[h] \right| \geq \epsilon \right] \leq B e^{-CN\epsilon^2 / \text{osc}(h)^2}.$$

(iii) Therefore, if A1, A2 and A3 hold uniformly in T and if H1 and H2 hold, then Theorem 1 holds with constants that are uniform in time. As a direct consequence, Theorems 2 and 3 hold also with constants that are uniform in time.

4.4 Asymptotic normality of the two-filter algorithms

We now establish CLT for the two-filter algorithms. Note first that under assumptions A1, A2 and A3, for all $0 \leq s, t \leq T$ a CLT may be derived for the weighted samples $\{(\xi_s^\ell, \omega_s^\ell)\}_{\ell=1}^N$ and $\{(\check{\xi}_{t|T}^i, \check{\omega}_{t|T}^i)\}_{i=1}^N$ which target respectively the filtering distribution $\phi_{\chi, s}$ and the backward information filter $\psi_{\gamma, t|T}$. By Propositions 10 and 11, there exist $\Gamma_{\chi, s}$ and $\check{\Gamma}_{\gamma, t|T}$ such that for any $h \in \mathbb{F}_b(\mathbb{X}, \mathcal{X})$,

$$N^{1/2} \sum_{i=1}^N \frac{\omega_s^i}{\Omega_s} \left(h(\xi_s^i) - \phi_{\chi, s}[h] \right) \xrightarrow{D}_{N \rightarrow \infty} \mathcal{N} \left(0, \Gamma_{\chi, s} [h - \phi_{\chi, s}[h]] \right), \quad (4.21)$$

$$N^{1/2} \sum_{j=1}^N \frac{\check{\omega}_{t|T}^j}{\check{\Omega}_{t|T}} \left(h(\check{\xi}_{t|T}^j) - \psi_{\gamma, t|T}[h] \right) \xrightarrow{D}_{N \rightarrow \infty} \mathcal{N} \left(0, \check{\Gamma}_{\gamma, t|T} [h - \psi_{\gamma, t|T}[h]] \right). \quad (4.22)$$

Theorem 4 establishes a CLT for the weighted sample $\{\omega_s^i \check{\omega}_{t|T}^j, (\xi_s^i, \check{\xi}_{t|T}^j)\}_{i, j=1}^N$ which targets the product distribution $\phi_{\chi, s} \otimes \psi_{\gamma, t|T}$. As an important consequence, the asymptotic variance of the weighted sample $\{\omega_s^i \check{\omega}_{t|T}^j, (\xi_s^i, \check{\xi}_{t|T}^j)\}_{i, j=1}^N$ is the sum of two contributions, the first one involves $\Gamma_{\chi, s}$ and the second one $\check{\Gamma}_{\gamma, t|T}$. Intuitively, this may be explained by the fact that the estimator $\phi_{\chi, s}^N \otimes \psi_{\gamma, t|T}^N[h]$ is obtained by mixing

two independent weighted samples which suggests the following decomposition:

$$\begin{aligned} \sqrt{N} \sum_{i,j=1}^N \frac{\omega_s^i \check{\omega}_{t|T}^j}{\Omega_s \check{\Omega}_{t|T}} \tilde{h}_{s,t}(\xi_s^i, \check{\xi}_{t|T}^j) &= \sqrt{N} \sum_{j=1}^N \frac{\check{\omega}_{t|T}^j}{\check{\Omega}_{t|T}} \phi_{\chi,s}[\tilde{h}_{s,t}(\cdot, \check{\xi}_{t|T}^j)] \\ &\quad + \sqrt{N} \sum_{i=1}^N \frac{\omega_s^i}{\Omega_s} \psi_{\gamma,t|T}[\tilde{h}_{s,t}(\xi_s^i, \cdot)] + \mathcal{E}_{s,T|t}^N(\tilde{h}_{s,t}), \end{aligned}$$

where $\tilde{h}_{s,t} = h - \phi_{\chi,s} \otimes \psi_{\gamma,t|T}[h]$ and

$$\mathcal{E}_{s,T|t}^N(h) := \sqrt{N} \sum_{i,j=1}^N \frac{\omega_s^i \check{\omega}_{t|T}^j}{\Omega_s \check{\Omega}_{t|T}} \left\{ h(\xi_s^i, \check{\xi}_{t|T}^j) - \phi_{\chi,s}[h(\cdot, \check{\xi}_{t|T}^j)] - \psi_{\gamma,t|T}[h(\xi_s^i, \cdot)] \right\}.$$

A CLT for the two independent first terms is obtained by (4.21) and (4.22). It remains then to prove that $\mathcal{E}_{s,T|t}^N(h)$ converges in probability to 0. However, this cannot be obtained directly from the exponential deviation inequality derived in Theorem 1 and requires sharper controls of the smoothing error (for instance nonasymptotic L^p -mean error bounds). Theorem 4 provides a direct proof following the asymptotic theory of weighted system of particles developed in [25].

Theorem 4. Assume that A1, A2 and A3 hold for some $T < \infty$. Then, for all $0 \leq s < t \leq T$ and all $h \in \mathbb{F}_b(X \times X, \mathcal{X} \times \mathcal{X})$,

$$\begin{aligned} \sqrt{N} \left(\sum_{i,j=1}^N \frac{\omega_s^i \check{\omega}_{t|T}^j}{\Omega_s \check{\Omega}_{t|T}} h(\xi_s^i, \check{\xi}_{t|T}^j) - \phi_{\chi,s} \otimes \psi_{\gamma,t|T}[h] \right) \\ \xrightarrow{\mathcal{D}}_{N \rightarrow \infty} \mathcal{N} \left(0, \tilde{\Gamma}_{s,t|T} [h - \phi_{\chi,s} \otimes \psi_{\gamma,t|T}[h]] \right), \end{aligned}$$

where $\tilde{\Gamma}_{s,t|T}[h]$ is defined by:

$$\tilde{\Gamma}_{s,t|T}[h] := \Gamma_{\chi,s} \left[\int \psi_{\gamma,t|T}(dx_t) h(\cdot, x_t) \right] + \check{\Gamma}_{\gamma,t|T} \left[\int \phi_{\chi,s}(dx_s) h(x_s, \cdot) \right], \quad (4.23)$$

with $\Gamma_{\chi,s}$ and $\check{\Gamma}_{\gamma,t|T}$ are given in Proposition 10 and Proposition 11.

Proof. The proof is postponed to Section 4.5.4. □

Define

$$\begin{aligned}\sigma_s &:= \phi_{\chi,s-1} \otimes \psi_{\gamma,s+1|T} \left[\int q^{[2]}(\cdot, x) g_s(x) dx \odot \gamma_{s+1}^{-1} \right], \\ \Sigma_s[h] &:= \tilde{\Gamma}_{s-1,s+1|T} \left[\int q^{[2]}(\cdot; x) g_s(x) h(x) dx \odot \gamma_{s+1}^{-1} \right].\end{aligned}$$

Theorem 5 provides a CLT for the TwoFilt_{fwt} algorithm of [34]

Theorem 5 (CLT for the TwoFilt_{fwt} algorithm of [34]). Assume that A1, A2 and A3 hold for some $T < \infty$. Then, for all $1 \leq s \leq T - 1$ and all $h \in \mathbb{F}_b(\mathcal{X}, \mathcal{X})$,

$$\sqrt{N} \left(\sum_{i=1}^N \frac{\tilde{\omega}_{s|T}^i}{\tilde{\Omega}_{s|T}} h(\tilde{\xi}_{s|T}^i) - \phi_{\chi,s|T}[h] \right) \xrightarrow{D}_{N \rightarrow \infty} \mathcal{N} \left(0, \Upsilon_{\chi,s|T} [h - \phi_{\chi,s|T}[h]] \right).$$

where

$$\begin{aligned}\Upsilon_{\chi,s|T} [h] &= \sigma_s^{-2} \left\{ \Sigma_s[h] + \phi_{\chi,s-1} \otimes \psi_{\gamma,s+1|T} \left[\tilde{\mathfrak{D}}_{s|T} \odot \gamma_{s+1}^{-1} \right] \right. \\ &\quad \left. \times \phi_{\chi,s-1} \otimes \psi_{\gamma,s+1|T} \left[\int \tilde{\omega}_{s|T}(\cdot; x) q^{[2]}(\cdot, x) g_s(x) h^2(x) dx \odot \gamma_{s+1}^{-1} \right] \right\}. \quad (4.24)\end{aligned}$$

Proof. The proof is postponed to Section 4.5.5. \square

The decompositions (4.27) and (4.28) together with Theorem 4 allow to prove a CLT form the forward and the backward approximations of the marginal smoothing distribution. Theorem 6 is a direct consequence of Proposition 11, Theorem 4 and Slutsky Lemma.

Theorem 6 (CLT for the TwoFilt_{bdm} algorithm of [10]). Assume that A1, A2 and A3 hold for some $T < \infty$. Then, for all $1 \leq s \leq T - 1$ and all $h \in \mathbb{F}_b(\mathcal{X}, \mathcal{X})$,

$$\sqrt{N} \left(\sum_{i=1}^N \frac{\tilde{\omega}_{s|T}^{i,f}}{\tilde{\Omega}_{s|T}^f} h(\xi_s^i) - \phi_{\chi,s|T}[h] \right) \xrightarrow{D}_{N \rightarrow \infty} \mathcal{N} \left(0, \Delta_{\chi,s|T}^f [h - \phi_{\chi,s|T}[h]] \right),$$

where

$$\begin{aligned}\Delta_{\chi,s|T}^f [h] &:= \tilde{\Gamma}_{s,s+1|T} [H_s^f] / \left(\phi_{\chi,s} \otimes \psi_{\gamma,s+1|T} [q \odot \gamma_{s+1}^{-1}] \right)^2, \\ H_s^f(x, x') &:= h(x) q(x, x') \gamma_{s+1}^{-1}(x').\end{aligned}$$

Similarly,

$$\sqrt{N} \left(\sum_{i=1}^N \frac{\tilde{\omega}_{s|T}^{i,b}}{\tilde{\Omega}_{s|T}^b} h(\xi_s^i) - \phi_{\chi,s|T}[h] \right) \xrightarrow{\mathcal{D}}_{N \rightarrow \infty} \mathcal{N} \left(0, \Delta_{\chi,s|T}^b [h - \phi_{\chi,s|T}[h]] \right),$$

where

$$\begin{aligned} \Delta_{\chi,s|T}^b [h] &:= \tilde{\Gamma}_{s-1,s|T} [H_s^b] / \{ \phi_{\chi,s-1} \otimes \psi_{\gamma,s|T} [q \odot \gamma_s^{-1}] \}^2, \\ H_s^b(x, x') &:= q(x, x') \gamma_s^{-1}(x') h(x'). \end{aligned}$$

Note that σ_s and $\Sigma_s[h]$ may be written as:

$$\sigma_s = \phi_{\chi,s} \otimes \psi_{\gamma,s+1|T} [q \odot \gamma_{s+1}^{-1}] \times \phi_{\chi,s-1} \left[\int q(\cdot, x) g_s(x) dx \right]$$

and by Theorem 4,

$$\begin{aligned} \Sigma_s[h] &= \Gamma_{\chi,s-1} \left[\int q(\cdot, x) g_s(x) h_{s+1}^1(x) dx \right] \\ &\quad + \phi_{\chi,s-1}^2 \left[\int q(\cdot, x) g_s(x) dx \right] \check{\Gamma}_{\gamma,s+1|T} [h_{s+1}^2], \end{aligned}$$

with $h_{s+1}^1(x) := h(x) \psi_{\gamma,s+1|T} [q(x, \cdot) \gamma_{s+1}^{-1}]$ and $h_{s+1}^2(x) := \gamma_{s+1}^{-1}(x) \phi_{\chi,s} [h(\cdot) q(\cdot, x)]$. In the case where $\tilde{r}_{s|T}(x_s, x_{s+1}; x_s) = p_s(x_{s-1}, x_s)$ in (4.15) and $\tilde{\vartheta}_{s|T}(x, x') = \vartheta_s(x) \vartheta_{s|T}(x')$, the smoothing distribution approximation given by the `TwoFiltfw` algorithm is obtained by reweighting the particles obtained in the forward filtering pass and $\Upsilon_{\chi,s|T}[h]$ may be compared to $\Delta_{\chi,s|T}^f[h]$ as both approximations of $\phi_{\chi,s|T}[h]$ are based on the same particles (associated with different importance weights). In this case, the two last terms in (4.24) are easily interpreted in the case $\vartheta_{s|T} = \gamma_{s+1}$:

$$\phi_{\chi,s-1} \otimes \psi_{\gamma,s+1|T} [\tilde{\vartheta}_{s|T} \odot \gamma_{s+1}^{-1}] = \phi_{\chi,s-1} [\vartheta_s] \psi_{\gamma,s+1|T} [\vartheta_{s|T} \gamma_{s+1}^{-1}] = \phi_{\chi,s-1} [\vartheta_s]$$

and by Jensen's inequality,

$$\begin{aligned} &\phi_{\chi,s-1} \otimes \psi_{\gamma,s+1|T} \left[\int \tilde{\omega}_{s|T}(\cdot; x) q^{[2]}(\cdot, x) g_s(x) h^2(x) dx \odot \gamma_{s+1}^{-1} \right] \\ &= \int \phi_{\chi,s-1}(dx_{s-1}) \omega_s(x_{s-1}, x) g_s(x) q(x_{s-1}, x) \psi_{\gamma,s+1|T} [q^2(x, \cdot) \gamma_{s+1}^{-2}] h^2(x) dx, \\ &\geq \int \phi_{\chi,s-1}(dx_{s-1}) \omega_s(x_{s-1}, x) g_s(x) q(x_{s-1}, x) (h_{s+1}^1(x))^2 dx. \end{aligned}$$

Therefore, by Proposition 11 and Theorem 6

$$\Upsilon_{\mathcal{X},s|T}[h] \geq \frac{\Gamma_{\mathcal{X},s}[h_{s+1}^1] + \check{\Gamma}_{\gamma,s+1|T}[h_{s+1}^2]}{\left(\phi_{\mathcal{X},s} \otimes \psi_{\gamma,s+1|T}[q \odot \gamma_{s+1}^{-1}]\right)^2} = \Delta_{\mathcal{X},s|T}^f[h],$$

where the last inequality comes from Theorem 4. The same inequality holds for $\Delta_{\mathcal{X},s|T}^b[h]$ when $\tilde{r}_{s|T}(x_{s-1}, x_{s+1}; x_s) = r_{s|T}(x_{s+1}, x_s)$ in (4.15).

Remark 2. Under the strong mixing assumptions H1 and H2, time uniform bounds for the asymptotic variances of the two-filter approximations of the marginal smoothing distributions may be obtained.

- (i) If A1 and A2 hold uniformly in T and if H1 holds, then it is proved in [24] that there exists $C > 0$ such that for all $s \geq 0$ and all $h \in \mathbb{F}_b(\mathcal{X}, \mathcal{X})$, the asymptotic variance $\Gamma_{\mathcal{X},s}[h]$ defined in Proposition 10 satisfies:

$$\Gamma_{\mathcal{X},s}[h] \leq C|h|_\infty^2.$$

- (ii) Following the same steps, if A1 and A3 hold uniformly in T and if H1 and H2 hold, there exists $C > 0$ such that for all $0 \leq t \leq T$ and all $h \in \mathbb{F}_b(\mathcal{X}, \mathcal{X})$, the asymptotic variance $\check{\Gamma}_{t|T}[h]$ defined in Proposition 11 satisfies:

$$\check{\Gamma}_{\gamma,t|T}[h] \leq C|h|_\infty^2.$$

- (iii) As a consequence, if A1, A2 and A3 hold uniformly in T and if H1 and H2 hold, the asymptotic variances $\tilde{\Gamma}_{s,t|T}[h]$, $\Delta_{\mathcal{X},s|T}^f[h]$, $\Delta_{\mathcal{X},s|T}^b[h]$ and $\Upsilon_{\mathcal{X},s|T}[h]$ defined in Theorem 4, Theorem 5 and Theorem 6 are all uniformly bounded.

4.5 Proofs

4.5.1 Proof of Theorem 1

Define $\mathcal{G}_{t|T}^N := \sigma(\check{\xi}_{t|T}^j, \check{\omega}_{t|T}^j, 1 \leq j \leq N)$ and

$$f_{t|T}(x) := \check{\Omega}_{t|T}^{-1} \sum_{j=1}^N \check{\omega}_{t|T}^j h(x, \check{\xi}_{t|T}^j)$$

whose oscillation is bounded by $\text{osc}(h)$. By the exponential inequality for the auxiliary particle filter (Proposition 8), there exist constants B_s and C_s such that

$$\begin{aligned} & \mathbb{P} \left(\left| \sum_{i,j=1}^N \frac{\omega_s^i \check{\omega}_{t|T}^j}{\check{\Omega}_{t|T}} h(\xi_s^i, \check{\xi}_{t|T}^j) - \sum_{j=1}^N \frac{\check{\omega}_{t|T}^j}{\check{\Omega}_{t|T}} \int \phi_{\chi,s}(\mathrm{d}x_s) h(x_s, \check{\xi}_{t|T}^j) \right| > \epsilon \right) \\ &= \mathbb{E} \left[\mathbb{P} \left(\left| \sum_{i=1}^N \frac{\omega_s^i}{\check{\Omega}_s} f_{t|T}(\xi_s^i) - \phi_{\chi,s}(f_{t|T}) \right| > \epsilon \mid \mathcal{G}_{t|T}^N \right) \right] \leq B_s e^{-C_s N \epsilon^2 / \text{osc}^2(h)}. \end{aligned} \quad (4.25)$$

Since the oscillation of the function $x \mapsto \int \phi_{\chi,s}(\mathrm{d}x_s) h(x_s, x)$ is bounded by $\text{osc}(h)$, by Proposition 9 there exist constants $B_{t|T}$ and $C_{t|T}$ such that

$$\begin{aligned} & \mathbb{P} \left(\left| \sum_{j=1}^N \frac{\check{\omega}_{t|T}^j}{\check{\Omega}_{t|T}} \int \phi_{\chi,s}(\mathrm{d}x_s) h(x_s, \check{\xi}_{t|T}^j) - \phi_{\chi,s} \otimes \psi_{\gamma,t|T}[h] \right| > \epsilon \right) \\ & \leq B_{t|T} e^{-C_{t|T} N \epsilon^2 / \text{osc}^2(h)}, \end{aligned} \quad (4.26)$$

which concludes the proof.

4.5.2 Proof of Theorem 2

Define $\tilde{h}_{s|T} := h - \phi_{\chi,s|T}[h]$. Lemma 7 is used with

$$\begin{aligned} a_N &:= N^{-1} \sum_{i=1}^N \tilde{\omega}_{s|T}^i \tilde{h}_{s|T}(\tilde{\xi}_{s|T}^i), \quad b_N := N^{-1} \tilde{\Omega}_{s|T}, \\ b &:= \frac{\phi_{\chi,s} \otimes \psi_{\gamma,s+1|T} \left[\int q^{[2]}(\cdot; x_s) g_s(x_s) \mathrm{d}x_s \odot \gamma_{s+1}^{-1} \right]}{\phi_{\chi,s} \otimes \psi_{\gamma,s+1|T} \left[\tilde{\mathfrak{D}}_{s|T} \odot \gamma_{s+1}^{-1} \right]}. \end{aligned}$$

Lemma 7-(i) is satisfied using $\beta := b$ and $|a_N|/|b_N| \leq \text{osc}(h)$. To prove Lemma 7-(ii) for a_N , note that Hoeffding inequality implies that, for any $\epsilon > 0$,

$$\mathbb{P} \left(\left| a_N - \mathbb{E} \left[\tilde{\omega}_{s|T}^1 \tilde{h}_{s|T}(\tilde{\xi}_{s|T}^1) \mid \mathcal{G}_{s,T}^N \right] \right| \geq \epsilon \mid \mathcal{G}_{s,T}^N \right) \leq 2 \exp \left\{ - \frac{N \epsilon^2}{8 \left| \tilde{\omega}_{s|T} \right|_{\infty}^2 \text{osc}^2(h)} \right\},$$

where $\mathcal{G}_{s,T}^N := \mathcal{G}_{s-1}^{N,+} \vee \mathcal{G}_{s+1,T}^{N,-}$ and

$$\begin{aligned} \mathcal{G}_s^{N,+} &:= \sigma \left\{ \{(\omega_u^i, \xi_u^i)\}_{i=1}^N, u = 1, \dots, s-1 \right\}, \\ \mathcal{G}_{s,T}^{N,-} &:= \sigma \left\{ \{(\check{\omega}_{u|T}^i, \check{\xi}_{u|T}^i)\}_{i=1}^N, u = s+1, \dots, T \right\}. \end{aligned}$$

On the other hand, for all $\ell \in \{1, \dots, N\}$,

$$\begin{aligned} & \mathbb{E} \left[\tilde{\omega}_{s|T}^\ell \tilde{h}_{s|T}(\tilde{\xi}_{s|T}^\ell) \middle| \mathcal{G}_{s,T}^N \right] \\ &= \frac{\sum_{i,j=1}^N \omega_{s-1}^i \check{\omega}_{s+1|T}^j \gamma_{s+1}^{-1}(\check{\xi}_{s+1|T}^j) \int q^{[2]}(\xi_{s-1}^i, \check{\xi}_{s+1|T}^j; x_s) g_s(x_s) \tilde{h}_{s|T}(x_s) dx_s}{\sum_{i,j=1}^N \omega_{s-1}^i \check{\omega}_{s+1|T}^j \gamma_{s+1}^{-1}(\check{\xi}_{s+1|T}^j) \tilde{\vartheta}_{s|T}(\xi_{s-1}^i, \check{\xi}_{s+1|T}^j)}. \end{aligned}$$

The proof of Lemma 7-(ii) is then completed by applying Lemma 7 to a'_N , b'_N and b' defined by:

$$\begin{aligned} a'_N &:= \sum_{i,j=1}^N \frac{\omega_{s-1}^i \check{\omega}_{s+1|T}^j}{\Omega_{s-1} \check{\Omega}_{s+1|T}} \gamma_{s+1}^{-1}(\check{\xi}_{s+1|T}^j) \int q^{[2]}(\xi_{s-1}^i, \check{\xi}_{s+1|T}^j; x_s) g_s(x_s) \tilde{h}_{s|T}(x_s) dx_s, \\ b'_N &:= \sum_{i,j=1}^N \frac{\omega_{s-1}^i \check{\omega}_{s+1|T}^j}{\Omega_{s-1} \check{\Omega}_{s+1|T}} \gamma_{s+1}^{-1}(\check{\xi}_{s+1|T}^j) \tilde{\vartheta}_{s|T}(\xi_{s-1}^i, \check{\xi}_{s+1|T}^j), \\ b' &:= \phi_{\chi,s} \otimes \psi_{\gamma,s+1|T}[\tilde{\vartheta}_{s|T} \odot \gamma_{s+1}^{-1}]. \end{aligned}$$

Note first that Lemma 7-(i) is satisfied using $\beta' := b'$ and $|a'_N/b'_N| \leq |\tilde{\omega}_{s|T}|_\infty \text{osc}(h)$. In addition, by (4.6),

$$\phi_{\chi,s} \otimes \psi_{\gamma,s+1|T}[\bar{h}_{s|T}] \propto \phi_{\chi,s|T}[\tilde{h}_{s|T}] = 0,$$

where

$$\bar{h}_{s|T}(x, x') := \int q^{[2]}(\cdot; x_s) g_s(x_s) \tilde{h}_{s|T}(x_s) dx_s \odot \gamma_{s+1}^{-1}(x, x').$$

Theorem 1 ensures that Lemma 7-(ii) is satisfied for a'_N as

$$\text{osc}(\bar{h}_{s|T}) \leq 2 |\tilde{\vartheta}_{s|T} \odot \gamma_{s+1}^{-1}|_\infty |\tilde{\omega}_{s|T}|_\infty \text{osc}(h).$$

Similarly, Theorem 1 yields:

$$\mathbb{P}(|b'_N - b'| \geq \epsilon) \leq B_s e^{-C_s N \epsilon^2 / \text{osc}^2(\tilde{\vartheta}_{s|T} \odot \gamma_{s+1}^{-1})},$$

which proves Lemma 7-(iii) for b'_N and concludes the proof of Lemma 7-(ii) for a_N . The proof of Lemma 7-(iii) for b_N is along the same lines.

4.5.3 Proof of Theorem 3

Define

$$\bar{h}_s(x, x') := \gamma_{s+1}^{-1}(x')h(x)q(x, x') \quad \text{and} \quad \underline{h}_s(x, x') := \gamma_s^{-1}(x')q(x, x')h(x').$$

It follows from the definition of the forward and backward smoothing weights (4.13) and (4.14) that,

$$\sum_{i=1}^N \frac{\tilde{\omega}_{s|T}^{i,f}}{\tilde{\Omega}_{s|T}^f} h(\xi_s^i) = \frac{\Omega_s^{-1} \check{\Omega}_{s+1|T}^{-1} \sum_{i,j=1}^N \omega_s^i \check{\omega}_{s+1|T}^j \bar{h}_s(\xi_s^i, \check{\xi}_{s+1|T}^j)}{\Omega_s^{-1} \check{\Omega}_{s+1|T}^{-1} \sum_{i,j=1}^N \omega_s^i \check{\omega}_{s+1|T}^j \bar{\mathbb{I}}_s(\xi_s^i, \check{\xi}_{s+1|T}^j)}, \quad (4.27)$$

$$\sum_{i=1}^N \frac{\tilde{\omega}_{s|T}^{i,b}}{\tilde{\Omega}_{s|T}^b} h(\xi_{s|T}^i) = \frac{\Omega_{s-1}^{-1} \check{\Omega}_{s|T}^{-1} \sum_{i,j=1}^N \omega_{s-1}^i \check{\omega}_{s|T}^j \underline{h}_s(\xi_{s-1}^i, \check{\xi}_{s|T}^j)}{\Omega_{s-1}^{-1} \check{\Omega}_{s|T}^{-1} \sum_{i,j=1}^N \omega_{s-1}^i \check{\omega}_{s|T}^j \underline{\mathbb{I}}_s(\xi_{s-1}^i, \check{\xi}_{s|T}^j)}. \quad (4.28)$$

On the other hand, from the definition of the filtering distribution and of the backward information filter

$$\begin{aligned} \phi_{\chi,s|T}[h] &= \phi_{\chi,s} \otimes \psi_{s+1|T}[\bar{h}_s] / \phi_{\chi,s} \otimes \psi_{s+1|T}[\bar{\mathbb{I}}_s], \\ \phi_{\chi,s|T}[h] &= \phi_{\chi,s-1} \otimes \psi_{s|T}[\underline{h}_s] / \phi_{\chi,s-1} \otimes \psi_{s|T}[\underline{\mathbb{I}}_s]. \end{aligned}$$

Then, (4.19) is established by writing:

$$\sum_{i=1}^N \frac{\tilde{\omega}_{s|T}^{i,f}}{\tilde{\Omega}_{s|T}^f} h(\xi_s^i) - \phi_{\chi,s|T}[h] = a_N^{i,f} / b_N^{i,f},$$

where

$$\begin{aligned} a_N^{i,f} &:= \sum_{i,j=1}^N \frac{\omega_s^i \check{\omega}_{s+1|T}^j}{\Omega_s \check{\Omega}_{s+1|T}} \bar{\mathbb{I}}_s(\xi_s^i, \check{\xi}_{s+1|T}^j) \left\{ \frac{\bar{h}_s(\xi_s^i, \check{\xi}_{s+1|T}^j)}{\bar{\mathbb{I}}_s(\xi_s^i, \check{\xi}_{s+1|T}^j)} - \frac{\phi_{\chi,s} \otimes \psi_{s+1|T}[\bar{h}_s]}{\phi_{\chi,s} \otimes \psi_{s+1|T}[\bar{\mathbb{I}}_s]} \right\}, \\ b_N^{i,f} &:= \sum_{i,j=1}^N \frac{\omega_s^i \check{\omega}_{s+1|T}^j}{\Omega_s \check{\Omega}_{s+1|T}} \bar{\mathbb{I}}_s(\xi_s^i, \check{\xi}_{s+1|T}^j), \quad b := \phi_{\chi,s} \otimes \psi_{s+1|T}[\bar{\mathbb{I}}_s]. \end{aligned}$$

Lemma 7 may then be applied with $\beta := b$. Note that

$$\frac{\bar{h}_s(\xi_s^i, \check{\xi}_{s+1|T}^j)}{\bar{\mathbb{I}}_s(\xi_s^i, \check{\xi}_{s+1|T}^j)} - \frac{\phi_{\chi,s} \otimes \psi_{s+1|T}[\bar{h}_s]}{\phi_{\chi,s} \otimes \psi_{s+1|T}[\bar{\mathbb{I}}_s]} = h(\xi_s^i) - \phi_{\chi,s|T}[h],$$

which ensures that $\left| a_N^{i,f}/b_N^{i,f} \right| \leq \text{osc}(h)$ and that Lemma 7-(i) is satisfied. By

$$\begin{aligned} \text{osc}(\bar{\mathbb{I}}_s) &= \text{osc}(q \odot \gamma_{s+1}^{-1}), \\ \text{osc}(\bar{\mathbb{I}}_s \odot \{h(\xi_s^i) - \phi_{\chi,s|T}[h]\}) &\leq 2 \left| q \odot \gamma_{s+1}^{-1} \right|_{\infty} \text{osc}(h), \end{aligned}$$

Theorem 1 shows that Lemma 7-(ii) and (iii) are satisfied. The proof of (4.20) follows the exact same lines.

4.5.4 Proof of Theorem 4

For all $1 \leq t \leq T$, the result is shown by induction on s where $s \in \{0, \dots, t-1\}$. Write $\tilde{h}_{0,t} := h - \phi_{\chi,0} \otimes \psi_{\gamma,t|T}[h]$ and set, for $i \in \{1, \dots, N\}$,

$$U_{N,i} := N^{-1/2} \omega_0^i \sum_{j=1}^N \frac{\check{\omega}_{t|T}^j}{\check{\Omega}_{t|T}} \tilde{h}_{0,t}(\xi_0^i, \check{\xi}_{t|T}^j).$$

Then,

$$\sqrt{N} \left(\sum_{i,j=1}^N \frac{\omega_0^i}{\Omega_0} \frac{\check{\omega}_{t|T}^j}{\check{\Omega}_{t|T}} h(\xi_0^i, \check{\xi}_{t|T}^j) - \phi_{\chi,0} \otimes \psi_{\gamma,t|T}[h] \right) = (\Omega_0/N)^{-1} \sum_{i=1}^N U_{N,i}.$$

Define $\mathcal{G}_{N,i} := \sigma(\{\xi_0^\ell\}_{\ell \leq i}, \{\check{\xi}_{u|T}^j\}_{t \leq u \leq T}, j = 1, \dots, N)$. Then,

$$\sum_{i=1}^N \mathbb{E}[U_{N,i} | \mathcal{G}_{N,i-1}] = N^{1/2} \sum_{j=1}^N \frac{\check{\omega}_{t|T}^j}{\check{\Omega}_{t|T}} \rho_0 \left[\omega_0 \tilde{h}_{0,t}(\cdot, \check{\xi}_{t|T}^j) \right].$$

As $\int \psi_{\gamma,t|T}(dx_t) \rho_0(dx_0) \omega_0(x_0) \tilde{h}_{0,t}(x_0, x_t) = 0$, by the CLT for the backward information filter (Proposition 11),

$$\sum_{i=1}^N \mathbb{E}[U_{N,i} | \mathcal{G}_{N,i-1}] \xrightarrow{D}_{N \rightarrow \infty} \mathcal{N}(0, \check{\Gamma}_{t|T}[H_{0,t}]),$$

where $H_{0,t}(x_t) := \int \rho_0(dx_0) \omega_0(x_0) \tilde{h}_{0,t}(x_0, x_t)$. We now prove that

$$\mathbb{E} \left[\exp \left(iu \sum_{i=1}^N \{U_{N,i} - \mathbb{E}[U_{N,i} | \mathcal{G}_{N,i-1}]\} \right) \middle| \mathcal{G}_{N,0} \right] \xrightarrow{P}_{N \rightarrow \infty} \exp \left(-\frac{u^2 \sigma_{0,t|T}^2[h]}{2} \right),$$

where

$$\sigma_{0,t|T}^2[h] := \int \rho_0(dx) \omega_0^2(x) \psi_{\gamma,t|T}^2[\tilde{h}_{0,t}(x, \cdot)].$$

This is done by applying [25, Theorem A.3] which requires to show that

$$\sum_{i=1}^N \left(\mathbb{E}[U_{N,i}^2 | \mathcal{G}_{N,i-1}] - \mathbb{E}[U_{N,i} | \mathcal{G}_{N,i-1}]^2 \right) \xrightarrow{P}_{N \rightarrow \infty} \sigma_{0,t|T}^2[h], \quad (4.29)$$

$$\sum_{i=1}^N \mathbb{E}[U_{N,i}^2 \mathbb{1}\{|U_{N,i}| > \varepsilon\} | \mathcal{G}_{N,i-1}] \xrightarrow{P}_{N \rightarrow \infty} 0. \quad (4.30)$$

By Proposition 9,

$$\sum_{i=1}^N \mathbb{E}[U_{N,i} | \mathcal{G}_{N,i-1}]^2 = \left(\sum_{j=1}^N \frac{\check{\omega}_{t|T}^j}{\check{\Omega}_{t|T}} H_{0,t}(\check{\xi}_{t|T}^j) \right)^2 \xrightarrow{P}_{N \rightarrow \infty} \psi_{\gamma,t|T}^2[H_{0,t}] = 0.$$

On the other hand,

$$\begin{aligned} & \mathbb{E} \left[\left| \sum_{i=1}^N \mathbb{E}[U_{N,i}^2 | \mathcal{G}_{N,i-1}] - \sigma_{0,t|T}^2[h] \right| \right] \\ &= \int \rho_0(dx) \omega_0^2(x) \mathbb{E} \left[\left| \left(\sum_{j=1}^N \frac{\check{\omega}_{t|T}^j}{\check{\Omega}_{t|T}} \tilde{h}_{0,t}(x, \check{\xi}_{t|T}^j) \right)^2 - \psi_{\gamma,t|T}^2[\tilde{h}_{0,t}(x, \cdot)] \right| \right] \\ &\leq 2 \operatorname{osc}(h) \int \rho_0(dx) \omega_0^2(x) \mathbb{E}[A_N(x)], \end{aligned}$$

where

$$A_N(x) := \left| \sum_{j=1}^N \frac{\check{\omega}_{t|T}^j}{\check{\Omega}_{t|T}} \tilde{h}_{0,t}(x, \check{\xi}_{t|T}^j) - \psi_{\gamma,t|T}[\tilde{h}_{0,t}(x, \cdot)] \right|.$$

By Proposition 9, there exist $B_{t|T}$ and $C_{t|T}$ such that for all $x \in X$,

$$\begin{aligned} \mathbb{E}[A_N(x)] &= \int_0^\infty \mathbb{P}(A_N(x) \geq \varepsilon) d\varepsilon \\ &\leq B_{t|T} \int_0^\infty e^{-C_{t|T} N \varepsilon^2 / \operatorname{osc}(h)^2} d\varepsilon \leq D_{t|T} \operatorname{osc}(h) N^{-1/2}, \quad (4.31) \end{aligned}$$

which shows that

$$\sum_{i=1}^N \mathbb{E}[U_{N,i}^2 | \mathcal{G}_{N,i-1}] \xrightarrow{P}_{N \rightarrow \infty} \sigma_{0,t|T}^2[h]$$

and concludes the proof of (4.29). For all $N \geq 1$,

$$\{|U_{N,i}| \geq \varepsilon\} \subseteq \{\omega_0^i \geq \varepsilon N^{1/2} \text{osc}(h)^{-1}\},$$

which implies that

$$\sum_{i=1}^N \mathbb{E} \left[U_{N,i}^2 \mathbb{1}_{\{|U_{N,i}| \geq \varepsilon\}} \middle| \mathcal{G}_{N,i-1} \right] \leq \text{osc}(h)^2 \int \rho_0(dx) \omega_0^2(x) \mathbb{1}_{\{\omega_0(x) \geq N^{1/2} \text{osc}(h)^{-1}\}}$$

and (4.30) follows by letting $N \rightarrow \infty$. Note that

$$N^{-1} \Omega_0 \xrightarrow{P}_{N \rightarrow \infty} \int \chi(dx_0) g_0(x_0),$$

which shows (4.23) since

$$\begin{aligned} \tilde{\Gamma}_{0,t|T}[\tilde{h}_{0,t}] &= \left(\int \chi(dx_0) g_0(x_0) \right)^{-2} \\ &\quad \times \left(\check{\Gamma}_{\gamma,t|T}[H_{0,t}] + \int \rho_0(dx) \omega_0^2(x) \psi_{\gamma,t|T}^2[\tilde{h}_{0,t}(x, \cdot)] \right), \\ &= \check{\Gamma}_{\gamma,t|T} \left[\int \phi_{\chi,0}(dx_0) \tilde{h}_{0,t}(x_0, \cdot) \right] + \Gamma_{\chi,0} \left[\int \psi_{\gamma,t|T}(dx_t) \tilde{h}_{0,t}(\cdot, x_t) \right]. \end{aligned}$$

Assume now that the result holds for some $s-1$. Write $\tilde{h}_{s,t} := h - \phi_{\chi,s} \otimes \psi_{\gamma,t|T}[h]$ and set, for $i \in \{1, \dots, N\}$,

$$U_{N,i} := N^{-1/2} \omega_s^i \sum_{j=1}^N \frac{\check{\omega}_{t|T}^j}{\check{\Omega}_{t|T}} \tilde{h}_{s,t}(\xi_s^i, \check{\xi}_{t|T}^j).$$

Then,

$$\sqrt{N} \left(\sum_{i,j=1}^N \frac{\omega_s^i \check{\omega}_{t|T}^j}{\Omega_s \check{\Omega}_{t|T}} h(\xi_s^i, \check{\xi}_{t|T}^j) - \phi_{\chi,0} \otimes \psi_{\gamma,t|T}[h] \right) = (\Omega_s/N)^{-1} \sum_{i=1}^N U_{N,i}.$$

Define, for $1 \leq i \leq N$,

$$\mathcal{G}_{N,i} := \sigma \left(\left\{ \xi_s^j \right\}_{j=1}^i, \left\{ \xi_u^\ell \right\}_{\ell=1}^N, \left\{ \check{\xi}_{v|T}^j \right\}_{j=1}^N, 1 \leq u < s, t \leq v \leq T \right).$$

Then,

$$\sum_{i=1}^N \mathbb{E} [U_{N,i} | \mathcal{G}_{N,i-1}] = (\phi_{\chi,s-1}^N[\vartheta_s])^{-1} N^{1/2} \sum_{i,j=1}^N \frac{\omega_{s-1}^i}{\Omega_{s-1}} \frac{\check{\omega}_{t|T}^j}{\check{\Omega}_{t|T}} H_s(\xi_{s-1}^i, \check{\xi}_{t|T}^j),$$

where

$$H_{s,t}(x_{s-1}, x_t) := \int q(x_{s-1}, x) g_s(x) \tilde{h}_{s,t}(x, x_t) dx. \quad (4.32)$$

Since $\phi_{\chi,s-1} \otimes \psi_{\gamma,t|T}[H_{s,t}] = 0$, by the induction assumption,

$$\sum_{i=1}^N \mathbb{E} [U_{N,i} | \mathcal{G}_{N,i-1}] \xrightarrow{D}_{N \rightarrow \infty} \mathcal{N}(0, \tilde{\Gamma}_{s-1,t|T}[H_{s,t}] / \phi_{\chi,s-1}^2[\vartheta_s]).$$

We will now prove that

$$\mathbb{E} \left[\exp \left(iu \sum_{i=1}^N \{U_{N,i} - \mathbb{E}[U_{N,i} | \mathcal{G}_{N,i-1}]\} \right) \middle| \mathcal{G}_{N,0} \right] \xrightarrow{P}_{N \rightarrow \infty} \exp \left(-\frac{u^2 \sigma_{s,t|T}^2[h]}{2} \right),$$

where

$$\begin{aligned} \sigma_{s,t|T}^2[h] &:= \phi_{\chi,s-1}[\vartheta_s]^{-1} \phi_{\chi,s-1}[f_{s-1,t}], \\ f_{s-1,t}(x_{s-1}) &:= \int q(x_{s-1}, x_s) \omega_s(x_{s-1}, x_s) \psi_{\gamma,t|T}^2[\tilde{h}_{s,t}(x_s, \cdot)] g_s(x_s) dx_s. \end{aligned}$$

This is done using again [25, Theorem A.3] and proving that (4.29) and (4.30) hold with $\sigma_{0,t|T}^2[h]$ replaced by $\sigma_{s,t|T}^2[h]$. Note that

$$\sum_{i=1}^N \mathbb{E} [U_{N,i} | \mathcal{G}_{N,i-1}]^2 = \left(\sum_{i,j=1}^N \frac{\omega_{s-1}^i}{\Omega_{s-1}} \frac{\check{\omega}_{t|T}^j}{\check{\Omega}_{t|T}} H_{s,t}(\xi_{s-1}^i, \check{\xi}_{t|T}^j) \right)^2 / (\phi_{\chi,s-1}^N[\vartheta_s])^2,$$

CHAPTER 4. CONVERGENCE RESULTS FOR TWO-FILTER ALGORITHMS

which converges in probability to 0 by Theorem 1 and the fact that $\phi_{\chi,s-1} \otimes \psi_{\gamma,t|T}[H_{s,t}] = 0$. In addition,

$$\begin{aligned} \sum_{i=1}^N \mathbb{E} \left[U_{N,i}^2 \mid \mathcal{G}_{N,i-1} \right] &= \left(\phi_{\chi,s-1}^N[\vartheta_s] \right)^{-1} \sum_{i=1}^N \frac{\omega_{s-1}^i}{\Omega_{s-1}} \int \frac{q^2(\xi_{s-1}^i, x_s) g_s(x_s)}{\vartheta_s(\xi_{s-1}^i) p_s(\xi_{s-1}^i, x_s)} \\ &\quad \times \left(\psi_{\gamma,t|T}^N[\tilde{h}_{s,t}(x_s, \cdot)] \right)^2 g_s(x_s) dx_s, \\ &= \left(\phi_{\chi,s-1}^N[\vartheta_s] \right)^{-1} \sum_{i=1}^N \frac{\omega_{s-1}^i}{\Omega_{s-1}} \int \omega_s(\xi_{s-1}^i, x_s) q(\xi_{s-1}^i, x_s) \\ &\quad \times \left(\psi_{\gamma,t|T}^N[\tilde{h}_{s,t}(x_s, \cdot)] \right)^2 g_s(x_s) dx_s, \\ &= \left(\phi_{\chi,s-1}^N[\vartheta_s] \right)^{-1} \phi_{\chi,s-1}^N[f_{s-1,t}^N], \end{aligned}$$

where

$$f_{s-1,t}^N(x_{s-1}) := \int q(x_{s-1}, x_s) \omega_s(x_{s-1}, x_s) \left(\psi_{\gamma,t|T}^N[\tilde{h}_{s,t}(x_s, \cdot)] \right)^2 g_s(x_s) dx_s.$$

First note that $\phi_{\chi,s-1}^N[\vartheta_s] \xrightarrow{P}_{N \rightarrow \infty} \phi_{\chi,s-1}[\vartheta_s]$ and write

$$\left| \phi_{\chi,s-1}^N[f_{s-1,t}^N] - \phi_{\chi,s-1}[f_{s-1,t}] \right| \leq A_{s,t}^N + B_{s,t}^N,$$

where $A_{s,t}^N := |\phi_{\chi,s-1}^N[f_{s-1,t}^N] - \phi_{\chi,s-1}^N[f_{s-1,t}]|$ and $B_{s,t}^N := |\phi_{\chi,s-1}^N[f_{s-1,t}] - \phi_{\chi,s-1}[f_{s-1,t}]|$. As $(\omega_{s-1}^i, \xi_{s-1}^i)_{i=1}^N$ and $(\check{\omega}_{t|T}^j, \check{\xi}_{t|T}^j)_{j=1}^N$ are independent,

$$\mathbb{E} \left[A_{s,t}^N \right] \leq |\omega_s|_\infty |g_s|_\infty \mathbb{E} \left[\sum_{i=1}^N \frac{\omega_{s-1}^i}{\Omega_{s-1}} \int q(\xi_{s-1}^i, x_s) \mathbb{E} \left[\left| \Delta \psi_{\gamma,t|T}^N[\tilde{h}_{s,t}](x_s) \right| \right] dx_s \right],$$

where $\Delta \psi_{\gamma,t|T}^N[\tilde{h}_{s,t}](x_s) := (\psi_{\gamma,t|T}^N[\tilde{h}_{s,t}(x_s, \cdot)])^2 - \psi_{\gamma,t|T}^2[\tilde{h}_{s,t}(x_s, \cdot)]$. Following the same steps as in (4.31), there exists $D_{t|T}$ such that

$$\mathbb{E} \left[\left| \Delta \psi_{\gamma,t|T}^N[\tilde{h}_{s,t}](x_s) \right| \right] \leq 2D_{t|T} \text{osc}(h)^2 / \sqrt{N},$$

which yields

$$\begin{aligned} \mathbb{E} \left[A_{s,t}^N \right] &\leq 2 \text{osc}(h)^2 |\omega_s|_\infty |g_s|_\infty D_{t|T} \mathbb{E} \left[\sum_{i=1}^N \frac{\omega_{s-1}^i}{\Omega_{s-1}} \int q(\xi_{s-1}^i, x_s) dx_s \right] / \sqrt{N}, \\ &\leq 2 \text{osc}(h)^2 |\omega_s|_\infty |g_s|_\infty D_{t|T} / \sqrt{N} \end{aligned}$$

and $\mathbb{E}[A_{s,t}^N] \xrightarrow{N \rightarrow \infty} 0$. On the other hand, as $\text{osc}(f_{s-1,t}) \leq \text{osc}(h)^2 |\omega_s|_\infty |g_s|_\infty$, $B_{s,t}^N \xrightarrow{N \rightarrow \infty} 0$ by Proposition 8. Finally, the tightness condition (4.30) holds since $|U_{N,i}| \leq N^{-1/2} |\omega_s|_\infty \text{osc}(h)$. Note that,

$$N^{-1} \Omega_s \xrightarrow{N \rightarrow \infty} \phi_{\chi,s-1} \left[\int q(\cdot, x_s) g_s(x_s) dx_s \right] / \phi_{\chi,s-1}[\vartheta_s].$$

Therefore (4.23) holds with

$$\begin{aligned} \tilde{\Gamma}_{s,t|T}[\tilde{h}_{s,t}] &= \frac{\phi_{\chi,s-1}^2[\vartheta_s]}{\phi_{\chi,s-1}^2 \left[\int q(\cdot, x_s) g_s(x_s) dx_s \right]} \left\{ \frac{\tilde{\Gamma}_{s-1,t|T}[H_{s,t}]}{\phi_{\chi,s-1}^2[\vartheta_s]} + \frac{\phi_{\chi,s-1}[f_{s-1,t}]}{\phi_{\chi,s-1}[\vartheta_s]} \right\}, \\ &= \frac{\tilde{\Gamma}_{s-1,t|T}[H_{s,t}]}{\phi_{\chi,s-1}^2 \left[\int q(\cdot, x_s) g_s(x_s) dx_s \right]} + \frac{\phi_{\chi,s-1}[f_{s-1,t}] \phi_{\chi,s-1}[\vartheta_s]}{\phi_{\chi,s-1}^2 \left[\int q(\cdot, x_s) g_s(x_s) dx_s \right]}, \end{aligned}$$

where, by induction assumption,

$$\begin{aligned} \tilde{\Gamma}_{s-1,t|T}[H_{s,t}] &= \Gamma_{\chi,s-1} \left[\int \psi_{\gamma,t|T}(dx_t) H_{s,t}(\cdot, x_t) \right] \\ &\quad + \check{\Gamma}_{\gamma,t|T} \left[\int \phi_{\chi,s-1}(dx_{s-1}) H_{s,t}(x_{s-1}, \cdot) \right]. \end{aligned}$$

The proof is completed upon noting that

$$\begin{aligned} \frac{\int \phi_{\chi,s-1}(dx_{s-1}) H_{s,t}(x_{s-1}, \cdot)}{\phi_{\chi,s-1} \left[\int q(\cdot, x_s) g_s(x_s) dx_s \right]} &= \frac{\int \phi_{\chi,s-1}(dx_{s-1}) q(x_{s-1}, x_s) g_s(x_s) \tilde{h}_{s,t}(x_s, \cdot) dx_s}{\phi_{\chi,s-1} \left[\int q(\cdot, x_s) g_s(x_s) dx_s \right]}, \\ &= \int \phi_{\chi,s}(dx_s) \tilde{h}_{s,t}(x_s, \cdot) \end{aligned}$$

and, by Proposition 10,

$$\begin{aligned} \Gamma_{\chi,s-1} \left[\frac{\int \psi_{\gamma,t|T}(dx_t) H_{s,t}(\cdot, x_t)}{\phi_{\chi,s-1} \left[\int q(\cdot, x_s) g_s(x_s) dx_s \right]} \right] &+ \frac{\phi_{\chi,s-1}[f_{s-1,t}] \phi_{\chi,s-1}[\vartheta_s]}{\phi_{\chi,s-1}^2 \left[\int q(\cdot, x_s) g_s(x_s) dx_s \right]} \\ &= \Gamma_{\chi,s} \left[\int \psi_{\gamma,t|T}(dx_t) \tilde{h}_{s,t}(\cdot, x_t) \right]. \end{aligned}$$

4.5.5 Proof of Theorem 5

Write $\tilde{h}_{s,T} = h - \phi_{\chi,s|T}(h)$. Note that

$$\sqrt{N} \sum_{i=1}^N \frac{\tilde{\omega}_{s|T}^i}{\tilde{\Omega}_{s|T}} \tilde{h}_{s,T}(\tilde{\xi}_{s|T}^i) = (\tilde{\Omega}_{s|T}/N)^{-1} \sum_{i=1}^N U_{N,i},$$

CHAPTER 4. CONVERGENCE RESULTS FOR TWO-FILTER ALGORITHMS

where $U_{N,\ell} := N^{-1/2} \tilde{\omega}_{s|T}^\ell \tilde{h}_{s,T}(\tilde{\xi}_{s|T}^\ell)$. Set, for $i \in \{1, \dots, N\}$,

$$\mathcal{G}_{N,i} := \sigma \left\{ \{(\tilde{\omega}_{s|T}^\ell, \tilde{\xi}_{s|T}^\ell)\}_{\ell=1}^i, \{(\omega_u^\ell, \xi_u^\ell)\}_{\ell=1}^N, u = 0, \dots, s-1, \right. \\ \left. \{(\tilde{\omega}_{v|T}^\ell, \tilde{\xi}_{v|T}^\ell)\}_{\ell=1}^N, v = s+1, \dots, T \right\}.$$

By the proof of Theorem 2,

$$N^{-1} \tilde{\Omega}_{s|T} \xrightarrow{P}_{N \rightarrow \infty} \frac{\phi_{\chi,s-1} \otimes \psi_{s+1|T} \left[\int q^{[2]}(\cdot, x) g_s(x) dx \odot \gamma_{s+1}^{-1} \right]}{\phi_{\chi,s-1} \otimes \psi_{s+1|T} \left[\tilde{\mathfrak{D}}_{s|T} \odot \gamma_{s+1}^{-1} \right]}.$$

The proof therefore amounts to establish a CLT for $\sum_{\ell=1}^N U_{N,\ell}$ and then to use Slutsky Lemma. The limit distribution of $\sum_{\ell=1}^N U_{N,\ell}$ is again obtained using the invariance principle for triangular array of dependent random variables derived in [25].

As

$$\sum_{i=1}^N \mathbb{E} \left[U_{N,i} \mid \mathcal{G}_{N,i-1} \right] \\ = \sqrt{N} \frac{\sum_{i,j=1}^N \omega_{s-1}^i \tilde{\omega}_{s+1|T}^j \int q^{[2]}(\cdot; x_s) g_s(x_s) \tilde{h}_{s,T}(x_s) dx_s \odot \gamma_{s+1}^{-1}(\xi_{s-1}^i, \tilde{\xi}_{s+1|T}^j)}{\sum_{i,j=1}^N \omega_{s-1}^i \tilde{\omega}_{s+1|T}^j \gamma_{s+1}^{-1}(\tilde{\xi}_{s+1|T}^j) \tilde{\mathfrak{D}}_{s|T}(\xi_{s-1}^i, \tilde{\xi}_{s+1|T}^j)},$$

it follows from Theorems 1 and 4 that

$$\sum_{i=1}^N \mathbb{E} \left[U_{N,i} \mid \mathcal{G}_{N,i-1} \right] \xrightarrow{P}_{N \rightarrow \infty} \mathcal{N} \left(0, \frac{\Sigma_s[\tilde{h}_{s,T}]}{\left(\phi_{\chi,s-1} \otimes \psi_{s+1|T} \left[\tilde{\mathfrak{D}}_{s|T} \odot \gamma_{s+1}^{-1} \right] \right)^2} \right).$$

Using that

$$\phi_{\chi,s|T}[\tilde{h}_{s,T}] = \frac{\phi_{\chi,s-1} \otimes \psi_{s+1|T} \left[\int q^{[2]}(\cdot; x) g_s(x) \tilde{h}_{s,T}(x) dx \right] \odot \gamma_{s+1}^{-1}}{\phi_{\chi,s-1} \otimes \psi_{s+1|T} \left[\int q^{[2]}(\cdot; x) g_s(x) dx \right] \odot \gamma_{s+1}^{-1}} = 0,$$

Theorem 1 yields

$$\begin{aligned} \sum_{i=1}^N \mathbb{E} \left[U_{N,i} | \mathcal{G}_{N,i-1} \right]^2 &= \\ &\left(\frac{\sum_{i,j=1}^N \omega_{s-1}^i \check{\omega}_{s+1|T}^j \gamma_{s+1}^{-1}(\check{\xi}_{s+1|T}^j) \int q^{[2]}(\xi_{s-1}^i, \check{\xi}_{s+1|T}^j; x) g_s(x) \tilde{h}_{s,T}(x)}{\sum_{i,j=1}^N \omega_{s-1}^i \check{\omega}_{s+1|T}^j \gamma_{s+1}^{-1}(\check{\xi}_{s+1|T}^j) \tilde{\mathfrak{D}}_{s|T}(\xi_{s-1}^i, \check{\xi}_{s+1|T}^j)} \right)^2 \\ &\xrightarrow{P}_{N \rightarrow \infty} \left(\frac{\phi_{\chi,s-1} \otimes \psi_{s+1|T} \left[\int q^{[2]}(\cdot; x) g_s(x) \tilde{h}_{s,T}(x) dx \odot \gamma_{s+1}^{-1} \right]}{\phi_{\chi,s-1} \otimes \psi_{s+1|T} [\tilde{\mathfrak{D}}_{s|T} \odot \gamma_{s+1}^{-1}]} \right)^2 = 0. \end{aligned}$$

Similarly, using again Theorem 1,

$$\begin{aligned} \sum_{i=1}^N \mathbb{E} \left[U_{N,i}^2 | \mathcal{G}_{N,i-1} \right] &\xrightarrow{P}_{N \rightarrow \infty} \\ &\frac{\phi_{\chi,s-1} \otimes \psi_{s+1|T} \left[\int \tilde{\omega}_{s|T}(\cdot; x) q^{[2]}(\cdot; x) g_s(x) \tilde{h}_{s,T}^2(x) dx \odot \gamma_{s+1}^{-1} \right]}{\phi_{\chi,s-1} \otimes \psi_{s+1|T} [\tilde{\mathfrak{D}}_{s|T} \odot \gamma_{s+1}^{-1}]} . \end{aligned}$$

Since under A2, $|U_{N,i}| \leq N^{-1/2} |\tilde{\omega}_{s|T}|_{\infty} \text{osc}(h)$, for any $\epsilon > 0$,

$$\sum_{i=1}^N \mathbb{E} \left[U_{N,i}^2 \mathbb{1}_{\{|U_{N,i}| \geq \epsilon\}} | \mathcal{G}_{N,i-1} \right] \xrightarrow{P}_{N \rightarrow \infty} 0,$$

which concludes the proof.

Chapter 5

A survey of Rao-Blackwellized Sequential Monte Carlo smoothers for Conditionally Linear and Gaussian models

Abstract

This paper focuses on Sequential Monte Carlo based approximations of smoothing distributions in conditionally linear and Gaussian state spaces. To reduce Monte Carlo variance of smoothers, it is typical in these models to use Rao-Blackwellization: particle approximation is used to sample the sequence of hidden regimes while the Gaussian states are explicitly integrated conditional on the sequence of regimes and observations, using variants of the Kalman filter / smoother. The first successful attempt to use Rao-Blackwellization for smoothing has been presented in [10] which extends the Bryson-Frazier smoother for Gaussian linear state space models using the generalized two-filter formula together with Kalman filters / smoothers. More recently, in a series of article, [67, 85, 65] proposed to use a forward backward decomposition of the smoothing distribution, mimicking the Rauch-Tung-Striebel smoother for the regime combined with backward Kalman updates. In this paper, we provide a survey of these algorithms in a common unifying framework. We then investigate

the benefit of introducing an additional rejuvenation step which allows to sample at each time instant new regimes conditional on the forward and backward particles. To support our claims, a Monte Carlo study is presented. These procedures are applied to commodity markets which are described using a two factor model based on the spot price and a convenience yield for crude oil data.

5.1 Introduction

State space models are bivariate stochastic processes $\{(Y_k, Z_k)\}_{k \geq 1}$ where the state sequence $\{Z_k\}_{k \geq 1}$ is a Markov chain which is only partially observed through the sequence $\{Y_k\}_{k \geq 1}$. Conditionally on the state sequence $\{Z_k\}_{k \geq 1}$ the observations are independent and for all $\ell \geq 1$ the conditional distribution of Y_ℓ given $\{Z_k\}_{k \geq 1}$ depends on Z_ℓ only. These models are used in a large variety of disciplines such as financial econometrics, biology, signal processing, see [21] and the references therein. In general state space models, bayesian filtering and smoothing problems, i.e. the computation of the posterior distributions of a sequence of states (Z_k, \dots, Z_p) for $1 \leq k \leq p \leq \ell$ given observations (Y_1, \dots, Y_ℓ) , are challenging tasks. Filtering refers to the estimation of the distributions of the hidden state Z_k given the observations (Y_1, \dots, Y_k) up to time k , while fixed-interval smoothing stands for the estimation of the distribution of sequence of states (Z_k, \dots, Z_p) given observations (Y_1, \dots, Y_ℓ) with $1 \leq k \leq p < \ell$.

When the state and observation models are linear and Gaussian, filtering can be solved explicitly using the Kalman filter [52]. Exact solutions of the fixed-horizon smoothing problem can be obtained using either the Rauch-Tung-Striebel smoother [82] or the Bryson-Frazier "two-filter" smoother [11]. This paper focuses on Conditionally Linear and Gaussian Models (CLGM) given for $k \geq 1$ by:

$$Z_k = d_{a_k} + T_{a_k} Z_{k-1} + H_{a_k} \varepsilon_k, \quad (5.1)$$

where

- $(\varepsilon_k)_{k \geq 1}$ is a sequence of independent and identically distributed (i.i.d.) m -dimensional Gaussian vectors with zero mean and identity covariance.

- $(a_k)_{k \geq 0}$ is a homogeneous Markov chain taking values in a finite space $\{1, \dots, J\}$, called *regimes*.
- $(H_j)_{1 \leq j \leq J}$ are positive-definite matrices.
- Z_0 is a m -dimensional Gaussian random with mean μ_1 and variance Σ_1 independent of $(\varepsilon_k)_{k \geq 1}$.

At each time step $k \geq 1$, the observation Y_k is given by:

$$Y_k = c_{a_k} + B_{a_k} Z_k + G_{a_k} \eta_k, \quad (5.2)$$

where

- $(\eta_k)_{k \geq 0}$ is a i.i.d. sequence of p -dimensional Gaussian vectors, independent of $(\varepsilon_k)_{k \geq 1}$ and Z_1 .
- $(G_j)_{1 \leq j \leq J}$ are $p \times p$ positive-definite matrices and $(B_j)_{1 \leq j \leq J}$ are $p \times m$ matrices.

CLGM play an important role in many applications; see [84] and the references therein for an up-to-date account. A crucial feature of these models is that, conditional on the regime sequence (a_1, \dots, a_n) , both the state equation and the observation equation are linear and Gaussian, which implies that conditional on the sequence of regimes and the observations, the filtering and the smoothing distribution of the 'continuous' states (Z_1, \dots, Z_n) can be computed explicitly.

To exploit this specific structure, it has been suggested in the pioneering papers of [16, 28] to solve the filtering problem by combining Sequential Monte Carlo (SMC) to sample the regimes with the Kalman filter to compute the conditional distribution of the states sequence $(Z_j)_{j=1}^n$ conditional on the regimes and the observations. This is a specific instance of Rao-Blackwellized Monte Carlo filters, often referred to as the *Mixture Kalman Filter*. Improvements of these early filtering techniques have been introduced in [29, 86].

The use of Rao Blackwellization to solve the smoothing problem has been proved to be more challenging and has received satisfactory solutions only recently.

The first forward-backward smoother proposed in the literature [35] was not fully Rao Blackwellized as it required to sample the hidden linear states in the backward pass.

An alternative approach, based on the so-called 'structural approximation' of the model suggested in an early paper by [56], was proposed by [3] to avoid to sample a 'continuous' state in the backward pass. This approximation is rather ad-hoc and the resulting smoother is not consistent when the number of particles goes to infinity. The inaccuracy introduced by the approximation might be difficult to control.

The first fully Rao Blackwellized SMC smoother which should lead to consistent approximations when the number of particles grows to infinity was proposed by [10] and extends the Bryson-Frazier smoother for Gaussian linear state space models using the generalized two-filter formula with Rao Blackwellization steps for the forward and the backward filters. This two-filter approach combines a forward filter with a backward information filter which are approximated numerically using SMC for the regime sequence and Kalman filtering techniques for the hidden linear states.

More recently, [67, 85, 65] introduced a Rao Blackwellized smoother based on the forward-backward decomposition of the FFBS algorithm with Rao Blackwellization steps both in the forward and backward time directions. The update of the smoothing distribution of the regime given the observations shares some striking similarities with the Rauch-Tung-Striebel smoothing procedure, which is at the heart of the FFBS procedure. The Rao-Blackwellization requires to update backward in time the smoothing distribution of the states given the regimes and the observations, which is achieved by using a *à la Kalman* backward update.

In this paper, we propose to improve the performance of the algorithms introduced in [10] and in [67, 85, 65] by using additional Rao-Blackwellization steps allowing to sample "new" particles in the backward pass. This approach may be seen as an extension of the ideas of [33] for Rao-Blackwellized smoothers. In [10], for all $1 \leq i \leq n$, the sampled forward and backward sequences are merged to approximate the posterior distribution of (a_i, z_i) . We propose to approximate the marginal smoothing distribution of (a_i, z_i) by merging the sampled forward and backward tra-

jectories at times $i - 1$ and $i + 1$ and integrating out all possible paths between time $i - 1$ and time i and between time i and time $i + 1$ instead of sampling random variables. Similarly, in the backward pass of [67, 85, 65], a regime \tilde{a}_i is sampled at time $1 \leq i \leq n - 1$ using the particles produced by the forward filter at time i . In this case, particle rejuvenation may be introduced by using the forward weighted samples at time $i - 1$ and extending these trajectories at time i with a Kalman filter for all possible values of the regime. Then, \tilde{a}_i is sampled in $\{1, \dots, J\}$ using an appropriately adapted weight.

The paper is organized as follows. The proposed modification of the fully Rao Blackwellized Monte Carlo smoothers are presented in Section 5.2. The performance of all these methods is illustrated in Section 5.3 with simulated data. In Section 5.4, an application to commodity markets is presented; the performance of our procedure is illustrated with crude oil data. A detailed derivation of the algorithms is provided in the Appendix.

5.2 Rao Blackwellized smoothing algorithm

This section details the Sequential Monte Carlo algorithms which can be used to approximate the conditional distribution of the states (a_1, \dots, a_n) or the marginal distributions of (a_i, z_i) given the observations (Y_1, \dots, Y_n) . For all $m \times m$ matrix let $|A|$ be the determinant of A . If A is a positive-definite matrix, for all $z \in \mathbb{R}^m$ define

$$\|z\|_A^2 := z' A^{-1} z.$$

Let $m(a_i, z_{i-1}; z_i)$ be the probability density of the conditional distribution of Z_i given (a_i, Z_{i-1}) and $g(a_i, z_i; y_i)$ be the probability density of the conditional distribution of Y_i given (a_i, Z_i) :

$$m(a_i, z_{i-1}; z_i) := |2\pi \bar{H}_{a_i}|^{-1/2} \exp \left\{ -\frac{1}{2} \|z_i - d_{a_i} - T_{a_i} z_{i-1}\|_{\bar{H}_{a_i}}^2 \right\}, \quad (5.3)$$

$$g(a_i, z_i; y_i) := |2\pi \bar{G}_{a_i}|^{-1/2} \exp \left\{ -\frac{1}{2} \|y_i - c_{a_i} - B_{a_i} z_i\|_{\bar{G}_{a_i}}^2 \right\}, \quad (5.4)$$

where

$$\bar{G}_j := G_j G_j', \quad \bar{H}_j := H_j H_j'.$$

All the algorithms considered in this paper are based on forward-backward or two-filter decompositions of the smoothing distributions and share the same forward filter presented in Section 5.2.1.

5.2.1 Forward filter

The SMC approximation $p^N(a_{1:i}, z_i | y_{1:i})$ of $p(a_{1:i}, z_i | y_{1:i})$ may be obtained using a standard Rao Blackwellized algorithm. The procedure produces a sequence of trajectories $(a_{1:i}^k)_{1 \leq k \leq N}$ associated with normalized importance weights $(\omega_i^k)_{1 \leq k \leq N}$ ($\sum_{k=1}^N \omega_i^k = 1$) used to define the following approximation of $p(a_{1:i}, z_i | y_{1:i})$:

$$p^N(a_{1:i}, z_i | y_{1:i}) = \sum_{k=1}^N \delta_{a_{1:i}^k}(a_{1:i}) \omega_i^k p(z_i | a_{1:i}^k, y_{1:i}). \quad (5.5)$$

In this equation, the conditional distribution of the hidden state z_i given the observations $y_{1:i}$ and a trajectory $a_{1:i}^k$ is a Gaussian distribution whose mean μ_i^k and variance P_i^k may be obtained by using the Kalman filter's update.

Initialization

At time $i = 1$, write, for all $1 \leq j \leq J$,

$$\mu_{1|0}^j = c_j + B_j \mu_1 \quad \text{and} \quad P_{1|0}^j = B_j \Sigma_1 B_j' + \bar{G}_j.$$

$(a_1^k)_{1 \leq k \leq N}$ are sampled independently in $\{1, \dots, J\}$ with probabilities proportional to

$$\pi_j p(a_1 = j | y_1) \propto \pi_j |P_{1|0}^j|^{-1/2} \exp \left\{ -(y_1 - \mu_{1|0}^j)' (P_{1|0}^j)^{-1} (y_1 - \mu_{1|0}^j) / 2 \right\}.$$

Then, μ_1^k and P_1^k are computed using a Kalman filter:

$$\begin{aligned} K_1^k &= \Sigma_1 B_{a_1^k}' \left(B_{a_1^k} \Sigma_1 B_{a_1^k}' + \bar{G}_{a_1^k} \right)^{-1}, \\ \mu_1^k &= \mu_1 + K_1^k \left(Y_1 - c_{a_1^k} - B_{a_1^k} \mu_1 \right), \\ P_1^k &= \left(I - K_1^k B_{a_1^k} \right) \Sigma_1. \end{aligned}$$

Each particle a_1^k is associated with the importance weight $\omega_1^k = 1/N$.

Iterations

Several procedures may be used to extend the trajectories $(a_{1:i-1}^k)_{1 \leq k \leq N}$ at time i . For all sampled trajectories $\{a_{1:i-1}^k\}_{1 \leq k \leq N}$ and all $1 \leq j \leq J$, [16] used the incremental weights:

$$\gamma_i^{j,k} = p(y_i | a_i = j, a_{1:i-1}^k, y_{1:i-1}) Q(a_{i-1}^k, j).$$

The conditional distribution of Y_i given $a_{1:i-1}^k$, a_i and $Y_{1:i-1}$ is a Gaussian distribution with mean $c_{a_i} + B_{a_i} \mu_{i|i-1}^k$ and variance $B_{a_i} P_{i|i-1}^k B_{a_i}' + \bar{G}_{a_i}$ so that:

$$\gamma_i^{j,k} \propto Q(a_{i-1}^k, j) |B_j P_{i|i-1}^{j,k} B_j' + \bar{G}_j|^{-1/2} \exp \left\{ -\frac{1}{2} \left\| y_i - c_j - B_j \mu_{i|i-1}^{j,k} \right\|_{B_j P_{i|i-1}^{j,k} B_j' + \bar{G}_j}^2 \right\},$$

where $\mu_{i|i-1}^{j,k} = d_j + T_j \mu_{i-1}^k$ and $P_{i|i-1}^{j,k} = T_j P_{i-1}^k T_j' + \bar{H}_j$. In [16], for all $1 \leq k \leq N$, an ancestral path is chosen with probabilities proportional to $(\omega_{i-1}^k)_{1 \leq k \leq N}$. Then, the new regime a_i^k is sampled in $\{1, \dots, J\}$ with probabilities proportional to $(\gamma_i^{j,k})_{1 \leq j \leq J}$. A drawback of this method is that only ancestral paths that have been selected using the importance weights $(\omega_{i-1}^k)_{1 \leq k \leq N}$ are extended at time i . Following [4], this may be improved by considering all the offsprings of all ancestral trajectories $\{a_{1:i-1}^k\}_{1 \leq k \leq N}$. Each ancestral path has J offsprings at time i , it is thus necessary to choose a given number of trajectories at time i (for instance N) among the NJ possible paths. To obtain the weight associated with each offspring write the following approximation of $p(a_{1:i} | y_{1:i})$ based on the weighted samples at time $i-1$:

$$\begin{aligned} p^N(a_{1:i} | y_{1:i}) &\propto \sum_{k=1}^N \omega_{i-1}^k Q(a_{i-1}^k, a_i) p(y_i | a_{1:i-1}^k, a_i, y_{1:i-1}) \delta_{a_{1:i-1}^k}(a_{1:i-1}), \\ &\propto \sum_{k=1}^N \sum_{j=1}^J \omega_{i-1}^k \gamma_i^{j,k} \delta_{(a_{1:i-1}^k, j)}(a_{1:i}). \end{aligned}$$

Therefore, each ancestral trajectory of the form $(a_{1:i-1}^k, j)$, $1 \leq k \leq N$, $1 \leq j \leq J$, is associated with the normalized weight $\tilde{\omega}_i^{j,k} \propto \omega_{i-1}^k \gamma_i^{j,k}$. Several random selection schemes have been proposed to discard some of the possible offsprings to maintain an average number of N particles at each time step. Following [4], we might choose between the Kullback-Leibler Optimal Selection (KL-OS) or the Chi-Squared Optimal Selection (CS-OS) to associate a new weight to each of the NJ trajectories.

If the new weight is 0, then the corresponding particle can be removed.

KL-OS: λ is chosen as the solution of :

$$\sum_{k=1}^N \sum_{j=1}^J \min(\tilde{\omega}_i^{j,k}/\lambda, 1) = N.$$

For all $1 \leq j \leq J$ and $1 \leq k \leq N$, if $\tilde{\omega}_i^{j,k} \geq \lambda$ then the new weight $\tilde{\Omega}_i^{j,k}$ is $\tilde{\Omega}_i^{j,k} = \tilde{\omega}_i^{j,k}$ and if $\tilde{\omega}_i^{j,k} < \lambda$:

$$\tilde{\Omega}_i^{j,k} = \begin{cases} \lambda & \text{with probability } \tilde{\omega}_i^{j,k}/\lambda, \\ 0 & \text{with probability } 1 - \tilde{\omega}_i^{j,k}/\lambda. \end{cases}$$

CS-OS: λ is chosen as the solution of :

$$\sum_{k=1}^N \sum_{j=1}^J \min(\sqrt{\tilde{\omega}_i^{j,k}/\lambda}, 1) = N.$$

For all $1 \leq j \leq J$ and $1 \leq k \leq N$, if $\tilde{\omega}_i^{j,k} \geq \lambda$ then the new weight $\tilde{\Omega}_i^{j,k}$ is $\tilde{\Omega}_i^{j,k} = \tilde{\omega}_i^{j,k}$ and if $\tilde{\omega}_i^{j,k} < \lambda$:

$$\tilde{\Omega}_i^{j,k} = \begin{cases} \sqrt{\tilde{\omega}_i^{j,k}} \lambda & \text{with probability } \sqrt{\tilde{\omega}_i^{j,k}/\lambda}, \\ 0 & \text{with probability } 1 - \sqrt{\tilde{\omega}_i^{j,k}/\lambda}. \end{cases}$$

Then, in both cases, all particles such that $\tilde{\Omega}_i^{j,k} = 0$ are discarded and for all the other trajectories defined as an ancestral path $(a_{1:i-1}^k)$ extended by $a_i^k = j$, the new corresponding weight ω in (5.5) is given by the normalized weight $\tilde{\Omega}_i^{j,k}$.

5.2.2 Algorithm based on the structural approximation of [56], [3]

The method to approximate the smoothing distributions of the regimes given all the observations proposed by [3] relies on an approximation of $p(a_i|a_{i+1:n}, y_{1:n})$, which is hardly justified and may introduce some bias. Assume first that the particles and importance weights $(a_i^k, \omega_i^k)_{1 \leq k \leq N}$, $1 \leq i \leq n$ approximating the filtering distributions are available. Then, note that, for all $1 \leq i \leq n$,

$$p(a_{i:n}|y_{1:n}) = p(a_i|a_{i+1:n}, y_{1:n})p(a_{i+1:n}|y_{1:n}).$$

Then, the approximation of [56, 3] replaces $p(a_i|a_{i+1:n}, y_{1:n})$ by

$$p(a_i|a_{i+1}, y_{1:i}) \propto Q(a_i, a_{i+1})p(a_i|y_{1:i}).$$

Plugging the particle approximation of $p(a_i|y_{1:i})$ in this expression provides the following procedure to sample trajectories approximately distributed according to $p(a_{1:n}|y_{1:n})$:

- \tilde{a}_n is first sampled in $\{a_n^1, \dots, a_n^N\}$ with probabilities proportional to $(\omega_n^k)_{1 \leq k \leq N}$.
- Then, for all $1 \leq i \leq n-1$, \tilde{a}_i is sampled in $\{a_i^1, \dots, a_i^N\}$ with probabilities proportional to $(\omega_i^k Q(a_i^k, \tilde{a}_{i+1}))_{1 \leq k \leq N}$.

If $(\tilde{a}_{1:n}^k)_{1 \leq k \leq \tilde{N}}$ are independent copies of $\tilde{a}_{1:n}$, the SMC approximation of [56] of the joint smoothing distribution of the regimes is:

$$p_N^{\text{Kim}}(a_{1:n}|Y_{1:n}) = \frac{1}{\tilde{N}} \sum_{k=1}^{\tilde{N}} \delta_{\tilde{a}_{1:n}^k}(a_{1:n}).$$

5.2.3 FFBS based algorithm of [67, 85, 65]

[67, 85, 65] proposed a Rao-Blackwellized procedure to sample the regime backward in time following the same steps as in the Forward Filtering Backward Smoothing algorithm [49, 28] without using the ad-hoc approximation of [56]. The algorithm relies on the decomposition given, for all $1 \leq i \leq n-1$, by:

$$p(a_{1:n}|y_{1:n}) = p(a_{1:i}|a_{i+1:n}, y_{1:n})p(a_{i+1:n}|y_{1:n}),$$

This decomposition is similar to the Rauch-Tung-Striebel decomposition of the filtering distribution. The first factor on the right hand side of the previous equation is nevertheless more difficult to handle because it itself relies on all the observations. Nevertheless, as noted by [85], this term can be computed recursively by considering the following obvious decomposition:

$$p(a_{1:i}|a_{i+1:n}, y_{1:n}) \propto p(y_{i+1:n}, a_{i+1:n}|a_{1:i}, y_{1:i})p(a_{1:i}|y_{1:i}).$$

The second factor in the last equation may be approximated using the ancestral trajectories $(a_{1:i}^k)_{1 \leq k \leq N}$ and the associated importance weights $(\omega_i^k)_{1 \leq k \leq N}$ produced by the forward filter. Therefore, $p(a_{1:i}|a_{i+1:n}, y_{1:n})$ may be approximated by:

$$p^N(a_{1:i}|a_{i+1:n}, y_{1:n}) = \sum_{k=1}^N \tilde{\omega}_{i|n}^k \delta_{a_{1:i}^k}(a_{1:i}), \quad \text{with} \quad \tilde{\omega}_{i|n}^k \propto \omega_i^k p(y_{i+1:n}, a_{i+1:n} | a_{1:i}^k, y_{1:i}).$$

Then, a trajectory $\tilde{a}_{1:n}$ approximatively distributed according to $p(a_{1:n}|y_{1:n})$ may be sampled following these steps:

- Set $\tilde{a}_n = a_n^k$ with probabilities proportional to $(\omega_n^k)_{1 \leq k \leq N}$.
- For all $1 \leq i \leq n-1$, set $\tilde{a}_i = a_i^k$ with probabilities proportional to $(\tilde{\omega}_{i|n}^k)_{1 \leq k \leq N}$.

This algorithm requires to compute the quantity $p(y_{i+1:n}, a_{i+1:n} | a_{1:i}^k, y_{1:i})$. As noted by [67], this predictive quantity is available analytically using Kalman filtering techniques. However, this has to be done for each trajectory $(a_{1:i}^k)_{1 \leq k \leq N}$, which leads to an algorithm with a prohibitive computational complexity. [67] proposed a procedure computationally less intensive by conditioning with respect to z_i and then marginalizing with respect to this variable:

$$p(y_{i+1:n}, a_{i+1:n} | a_{1:i}^k, y_{1:i}) = \int p(y_{i+1:n}, a_{i+1:n} | z_i, a_i^k) p(z_i | a_{1:i}^k, y_{1:i}) dz_i, \quad (5.6)$$

This is similar to the two-filter decomposition of the smoothing distribution, see Section 5.2.4. By [67],

$$p(y_{i+1:n}, a_{i+1:n} | z_i, a_i) \propto Q(a_i, a_{i+1}) \exp \left\{ - (z_i' \Omega_i(a_{i+1:n}) z_i - 2 \lambda_i'(a_{i+1:n}) z_i) / 2 \right\},$$

where the proportionality is with respect to (a_i, z_i) and

$$p(y_{i:n}, a_{i+1:n} | z_i, a_i) \propto \exp \left\{ - (z_i' \hat{\Omega}_i(a_{i:n}) z_i - 2 \hat{\lambda}_i'(a_{i:n}) z_i) / 2 \right\},$$

where the proportionality is with respect to z_i . These quantities may be computed recursively backward in time with:

$$\begin{aligned} \hat{\Omega}_n(a_n) &= B_{a_n}' \bar{G}_{a_n}^{-1} B_{a_n}, \\ \hat{\lambda}_n(a_n) &= B_{a_n}' \bar{G}_{a_n}^{-1} (y_n - c_{a_n}). \end{aligned}$$

Then, for $1 \leq i \leq n-1$, define $m_{i+1} = \hat{\lambda}_{i+1} - \hat{\Omega}_{i+1} d_{a_{i+1}}$ and $M_{i+1} = H'_{a_{i+1}} \hat{\Omega}_{i+1} H_{a_{i+1}} + I$ and write

$$\begin{aligned}\Omega_i(a_{i+1:n}) &= T'_{a_{i+1}} (I - \hat{\Omega}_{i+1}(a_{i+1:n})H_{a_{i+1}}M_{i+1}^{-1}H'_{a_{i+1}}) \hat{\Omega}_{i+1}(a_{i+1:n})T_{a_{i+1}}, \\ \lambda_i(a_{i+1:n}) &= T'_{a_{i+1}} (I - \hat{\Omega}_{i+1}(a_{i+1:n})H_{a_{i+1}}M_{i+1}^{-1}H'_{a_{i+1}})m_{i+1}.\end{aligned}$$

As $p(y_{i:n}, a_{i+1:n} | z_i, a_i) = p(y_i | z_i, a_i)p(y_{i+1:n}, a_{i+1:n} | z_i, a_i)$,

$$\begin{aligned}\hat{\Omega}_i(a_{i:n}) &= \Omega_i(a_{i+1:n}) + B'_{a_i} \bar{G}_{a_i}^{-1} B_{a_i}, \\ \hat{\lambda}_i(a_{i:n}) &= \lambda_i(a_{i+1:n}) + B'_{a_i} \bar{G}_{a_i}^{-1} (y_i - c_{a_i}).\end{aligned}$$

Then, by (5.6),

$$p(y_{i+1:n}, a_{i+1:n} | a_{1:i}^k, y_{1:i}) \propto Q(a_i^k, a_{i+1}) |\Lambda_i|^{-1/2} \exp\{\eta_i/2\}, \quad (5.7)$$

where the proportionality is with respect to $a_{1:i}^k$ and

$$\begin{aligned}\Lambda_i^k(a_{i+1:n}) &= (\Gamma_i^k)' \Omega_i(a_{i+1:n}) \Gamma_i^k + I, \\ \eta_i^k(a_{i+1:n}) &= \|\mu_i^k\|_{\Omega_i^{-1}(a_{i+1:n})}^2 - 2\lambda_i'(a_{i+1:n})\mu_i^k - \|(\Gamma_i^k)'(\lambda_i(a_{i+1:n}) - \Omega_i(a_{i+1:n})\mu_i^k)\|_{\Lambda_i^k(a_{i+1:n})}^2,\end{aligned}$$

where $P_i^k = \Gamma_i^k(\Gamma_i^k)'$. Therefore,

$$\tilde{\omega}_{i|n} \propto \omega_i^k Q(a_i^k, a_{i+1}) |\Lambda_i^k(a_{i+1:n})|^{-1/2} \exp\{\eta_i^k(a_{i+1:n})/2\}.$$

If $(\tilde{a}_{1:n}^k)_{1 \leq k \leq \tilde{N}}$ are independent copies of $\tilde{a}_{1:n}$, the SMC approximation of [67] of the joint smoothing distribution of the regime is:

$$p_{\tilde{N}}^{\text{Lbscg}}(a_{1:n} | Y_{1:n}) = \frac{1}{\tilde{N}} \sum_{k=1}^{\tilde{N}} \delta_{\tilde{a}_{1:n}^k}(a_{1:n}).$$

Additional particle rejuvenation step

An improved version of this FFBS algorithm may be defined for all $2 \leq i \leq n-1$ by writing :

$$\begin{aligned}p(a_{1:i} | a_{i+1:n}, y_{1:n}) &\propto p(y_{i+1:n}, a_{i+1:n} | a_{1:i}, y_{1:i}) p(a_{1:i} | y_{1:i}), \\ &\propto p(y_{i+1:n}, a_{i+1:n} | a_{1:i}, y_{1:i}) \int p(a_{1:i-1}, z_{i-1} | y_{1:i-1}) Q(a_{i-1}, a_i) \\ &\quad m(a_i, z_{i-1}; z_i) g(a_i, z_i; y_i) dz_{i-1:i}.\end{aligned}$$

Replacing $p(a_{1:i-1}, z_{i-1} | y_{1:i-1})$ in the integral by the particle approximation obtained during the forward pass and using Kalman filtering techniques for each trajectory $\{a_{1:i-1}^k\}_{1 \leq k \leq N}$ and each $a_i \in \{1, \dots, J\}$ yields:

$$\int p^N(a_{1:i-1}, z_{i-1} | y_{1:i-1}) Q(a_{i-1}, a_i) m(a_i, z_{i-1}; z_i) g(a_i, z_i; y_i) dz_{i-1:i} \propto \sum_{k=1}^N \omega_{i|i-1}^k(a_i) \delta_{a_{1:i-1}^k}(a_{1:i-1}),$$

where

$$\omega_{i|i-1}^k(a_i) = \omega_{i-1}^k Q(a_{i-1}^k, a_i) |\Sigma_{i|i-1}^k(a_i)|^{-1/2} e^{-(y_i - y_{i|i-1}^k(a_i))' (\Sigma_{i|i-1}^k(a_i))^{-1} (y_i - y_{i|i-1}^k(a_i)) / 2},$$

$$y_{i|i-1}^k(a_i) = c_{a_i} + B_{a_i} (d_{a_i} + T_{a_i} \mu_{i-1}^k) \text{ and } \Sigma_{i|i-1}^k(a_i) = B_{a_i} (T_{a_i} P_{i-1}^k T_{a_i}' + \bar{H}_{a_i}) B_{a_i}' + \bar{G}_{a_i}.$$

On the other hand, for all $1 \leq k \leq N$, $p(y_{i+1:n}, a_{i+1:n} | a_{1:i-1}^k, a_i, y_{1:i})$ is computed as in (5.7) with all possible values $a_i \in \{1, \dots, J\}$ and not only the regime of the filtering pass $\{a_i^k\}_{1 \leq k \leq N}$. This means that a Kalman filter must be used for each trajectory $a_{1:i-1}^k$ which may be extended by $a_i \in \{1, \dots, J\}$. Denote by $\mu_{i|i-1}^k(a_i)$ and $P_{i|i-1}^k(a_i)$ the mean and covariance matrix of the law of z_i given $(a_{1:i-1}^k, a_i)$. Then,

$$p(y_{i+1:n}, a_{i+1:n} | a_{1:i-1}^k, a_i, y_{1:i}) = Q(a_i, a_{i+1}) |\Lambda_{i|i-1}^k(a_{i:n})|^{-1/2} \exp\{\eta_{i|i-1}^k(a_{i:n})/2\},$$

where the proportionality is with respect to $(a_{1:i-1}^k, a_i)$ and

$$\begin{aligned} \Lambda_{i|i-1}^k(a_{i:n}) &= (\Gamma_{i|i-1}^k(a_i))' \Omega_i(a_{i+1:n}) \Gamma_{i|i-1}^k(a_i) + I, \\ \eta_{i|i-1}^k(a_{i:n}) &= \|\mu_{i|i-1}^k(a_i)\|_{\Omega_i^{-1}(a_{i+1:n})}^2 - 2\lambda_i'(a_{i+1:n}) \mu_{i|i-1}^k(a_i) \\ &\quad - \|(\Gamma_{i|i-1}^k(a_i))' (\lambda_i(a_{i+1:n}) - \Omega_i(a_{i+1:n}) \mu_{i|i-1}^k(a_i))\|_{\Lambda_i(a_{i+1:n})}^2, \end{aligned}$$

where $P_{i|i-1}^k(a_i) = \Gamma_{i|i-1}^k(a_i) (\Gamma_{i|i-1}^k(a_i))'$. $p(a_{1:i} | a_{i+1:n}, y_{1:n})$ is then approximated by :

$$p^N(a_{1:i} | a_{i+1:n}, y_{1:n}) \propto \sum_{k=1}^N \omega_{i|i-1}^k(a_i) Q(a_i, a_{i+1}) |\Lambda_{i|i-1}^k(a_{i:n})|^{-1/2} \exp\{\eta_{i|i-1}^k(a_{i:n})/2\} \delta_{a_{1:i-1}^k}(a_{1:i-1}).$$

By integrating over all possible paths $a_{1:i-1}$, \tilde{a}_i is sampled in $\{1, \dots, J\}$.

5.2.4 Rao Blackwellized Two-filter Smoother (RBTF)

Contrary to the previous methods, two-filter based smoothers are designed to compute approximations of marginal smoothing distributions (usually the posterior distribution of one or two consecutive regimes given all the observations). [10] introduced the following decomposition of the smoothing distributions for all $2 \leq i \leq n$:

$$p(a_i, z_i | y_{1:n}) \propto p(a_i, z_i | y_{1:i-1}) p(y_{i:n} | a_i, z_i). \quad (5.8)$$

The first term in the right hand side may be approximated using the forward filter, however, as the function $(a_i, z_i) \mapsto p(y_{i:n} | a_i, z_i)$ is not a probability density function, it is not possible to use SMC samples to approximate it. The backward filter uses artificial densities to introduce a surrogate target density function which may be approximated recursively using SMC methods. Then, the forward and backward weighted samples are combined using (5.8) to approximate $p(a_i, z_i | y_{1:n})$. **Additional particle rejuvenation step**

for $2 \leq i \leq n-1$, our proposed algorithm relies on the explicit marginalization:

$$p(a_i, z_i | y_{1:n}) = \sum_{a_{i-1}} \sum_{a_{i+1}} \int_{z_{i-1}} \int_{z_{i+1}} \psi_i^n(a_{i-1:i+1}, z_{i-1:i+1}) dz_{i-2} dz_{i+1}, \quad (5.9)$$

where $\psi_i^n(a_{i-1:i+1}, z_{i-1:i+1})$ is the smoothing distribution of the hidden regimes and states between time indices $i-1$ and $i+1$. Note that the EM algorithm requires the approximation of $p(a_{i-1}, z_{i-1}, a_{i+1}, z_{i+1} | y_{1:n})$ in the E-step, this may be obtained following the same steps by marginalizing explicitly the linear states at time $i-2$ and $i+1$, see Section 5.3. Intermediate computations follow the same steps as for the approximation of $p(a_i, z_i | y_{1:n})$. ψ_i^n may be decomposed as follows:

$$\begin{aligned} \psi_i^n(a_{i-1:i+1}, z_{i-1:i+1}) &\propto p(y_{i+1:n} | a_{i+1}, z_{i+1}) p(a_{i-1}, z_{i-1} | y_{1:i-1}) Q(a_{i-1}, a_i) m(a_i, z_{i-1}; z_i) \\ &\quad \times g(a_i, z_i; y_i) Q(a_i, a_{i+1}) m(a_{i+1}, z_i; z_{i+1}), \end{aligned}$$

where the proportionality is with respect to $(a_{i-1:i+1}, z_{i-1:i+1})$. Then, by (5.9), the smoothing distribution $p(a_i, z_i | y_{1:n})$ may be written as

$$p(a_i, z_i | y_{1:n}) \propto r_i(a_i, z_i, y_{1:i-1}) g(a_i, z_i; y_i) t_i(a_i, z_i, y_{i+1:n}), \quad (5.10)$$

where m and g are defined in (5.3) and (5.4) and

$$r_i(a_i, z_i, y_{1:i-1}) = p(a_i, z_i | y_{1:i-1}), \quad (5.11)$$

$$= \sum_{a_{i-1}} \int_{z_{i-1}} p(a_{i-1}, z_{i-1} | y_{1:i-1}) m(a_i, z_{i-1}; z_i) Q(a_{i-1}, a_i) dz_{i-1}, \quad (5.12)$$

$$t_i(a_i, z_i, y_{i+1:n}) = \sum_{a_{i+1}} \int_{z_{i+1}} m(a_{i+1}, z_i; z_{i+1}) Q(a_i, a_{i+1}) p(y_{i+1:n} | a_{i+1}, z_{i+1}) dz_{i+1}. \quad (5.13)$$

In the forward pass described in Section 5.2.1, a set of possible sequences of regimes $a_{1:i-1}^k$ associated with importance weights ω_{i-1}^k , $1 \leq k \leq N$ are sampled to approximate $p(a_{i-1}, z_{i-1} | y_{1:i-1})$. Plugging this approximation into (5.12) provides a normalized approximation $r_i^N(a_i, z_i, y_{1:i-1})$ of $r_i(a_i, z_i, y_{1:i-1})$ integrating over all possible choices of (a_{i-1}, z_{i-1}) :

$$r_i^N(a_i, z_i, y_{1:i-1}) = \sum_{a_{i-1}} \int_{z_{i-1}} p^N(a_{i-1}, z_{i-1} | y_{1:i-1}) m(a_i, z_{i-1}; z_i) Q(a_{i-1}, a_i) dz_{i-1}.$$

Define

$$\begin{aligned} \Omega_{i-1}^k(a_i) &= T_{a_i} P_{i-1}^k T_{a_i}' + \bar{H}_{a_i}, \quad \mu_{i-1}^k(a_i) = d_{a_i} + T_{a_i} \mu_{i-1}^k, \quad r_{i-1}^k(a_i) = (\Omega_{i-1}^k(a_i))^{-1} \mu_{i-1}^k(a_i), \\ \omega_{f,i}^k(a_i) &= \omega_{i-1}^k Q(a_{i-1}^k, a_i) |2\pi \Omega_{i-1}^k(a_i)|^{-1/2} \exp\left\{-\frac{1}{2} \|\mu_{i-1}^k(a_i)\|_{\Omega_{i-1}^k(a_i)}^2\right\}. \end{aligned}$$

Then,

$$r_i^N(a_i, z_i, y_{1:i-1}) = \sum_{k=1}^N \omega_{f,i}^k(a_i) \exp\left\{-\frac{1}{2} \|z_i\|_{\Omega_{i-1}^k(a_i)}^2 + z_i' r_{i-1}^k(a_i)\right\}. \quad (5.14)$$

The backward pass described below produces a sequence of states $\tilde{a}_{i+1:n}^\ell$ associated with importance weights $\tilde{\omega}_{i+1:n}^\ell$, $1 \leq \ell \leq N$ which are used to approximate $p(y_{i+1:n} | a_{i+1}, z_{i+1})$. Plugging this approximation into (5.13) provides a non normalized approximation $t_i^N(a_i, z_i, y_{i+1:n})$ of $t_i(a_i, z_i, y_{i+1:n})$ integrating over all possible choices (a_{i+1}, z_{i+1}) .

These steps are then combined to form a non normalized SMC approximation of $p(a_i, z_i | y_{1:n})$ using (5.10). As $t_i^N(a_i, z_i, y_{i+1:n})$ is not normalized, the normalization of the SMC approximation of $p(a_i, z_i | y_{1:n})$ is obtained by integrating over the states a_i, z_i , when $r_i(a_i, z_i, y_{1:i-1})$ and $t_i(a_i, z_i, y_{i+1:n})$ are replaced by $r_i^N(a_i, z_i, y_{1:i-1})$ and

$t_i^N(a_i, z_i, y_{i+1:n})$ in (5.10). Our procedure allows to construct sequence of regimes with non-degenerated importance weights in the combination step. This procedure improves significantly [10] where no marginalization of $p(a_i, z_i | y_{1:n})$ over the states at times $i - 1$ and $i + 1$ is performed and where the proposed forward and backward paths are directly merged. This method often leads to importance weights which are close to be numerically degenerated.

As the function $(a_{i+1}, z_{i+1}) \mapsto p(y_{i+1:n} | a_{i+1}, z_{i+1})$ is not a probability density function, it is not possible to use SMC samples to approximate it. Following [10], for any probability densities $(\gamma_i)_{1 \leq i \leq n}$, define the following joint probability densities:

$$\tilde{p}_n(a_n, z_n, y_n) := \gamma_n(a_n, z_n)g(a_n, z_n; y_n), \quad \tilde{p}_n(y_n) := \sum_{a_n=1}^J \int \gamma_n(a_n, z_n)g(a_n, z_n; y_n)dz_n,$$

and, for all $1 \leq i \leq n - 1$,

$$\begin{aligned} \tilde{p}_i(a_{i:n}, z_{i:n}, y_{i:n}) &:= \gamma_i(a_i, z_i)p(y_{i:n} | a_{i:n}, z_{i:n})p(a_{i+1:n}, z_{i+1:n} | a_i, z_i), \\ \tilde{p}_i(y_{i:n}) &:= \sum_{a_{i:n}=1}^J \int \gamma_i(a_i, z_i)p(y_{i:n} | a_{i:n}, z_{i:n})p(a_{i+1:n}, z_{i+1:n} | a_i, z_i)dz_{i:n}. \end{aligned}$$

By Lemma 1, these probability densities may be used to approximate the quantities $p(y_{i:n} | a_i, z_i)$, $1 \leq i \leq n$, in (5.13).

Lemma 1. For all $1 \leq i \leq n - 1$,

$$\tilde{p}_i(a_i, z_i | y_{i:n}) = p(y_{i:n} | a_i, z_i)\gamma_i(a_i, z_i)/\tilde{p}_i(y_{i:n}), \quad (5.15)$$

$$\tilde{p}_i(a_i, z_i | y_{i:n}) = \gamma_i(a_i, z_i) \sum_{a_{i+1:n}=1}^J \frac{\tilde{p}_i(a_{i:n} | y_{i:n})p(y_{i:n} | a_{i:n}, z_i)}{\int \gamma_i(a_i, z')p(y_{i:n} | a_{i:n}, z')dz'}. \quad (5.16)$$

Proof. The proof is postponed to Appendix A.2.1. \square

A set of weighted trajectories $(\tilde{a}_{i:n}^\ell)_{1 \leq \ell \leq N}$ with importance weights $(\tilde{\omega}_i^\ell)_{1 \leq \ell \leq N}$, $1 \leq i \leq n$, may be sampled recursively backward in time to produce a SMC approximation of $\tilde{p}(a_{i:n} | y_{i:n})$.

- For $1 \leq \ell \leq N$, sample $\tilde{a}_n^j \sim \tilde{q}_n(\cdot)$ and set:

$$\tilde{\omega}_n^\ell \propto \frac{\int \gamma_n(\tilde{a}_n^\ell, z')g(\tilde{a}_n^\ell, z'; y_n)dz'}{\tilde{q}_n(\tilde{a}_n^\ell)}.$$

- For all $1 \leq i \leq n-1$, resample the set $(\tilde{a}_{i+1:n}^\ell)_{1 \leq j \leq N}$ using the normalized weights $(\tilde{\omega}_{i+1}^\ell)_{1 \leq j \leq N}$. Then, for $1 \leq \ell \leq N$, sample $\tilde{a}_i^j \sim \tilde{q}_i(\tilde{a}_{i+1:n}^\ell, \cdot)$ and set:

$$\tilde{\omega}_i^\ell \propto \frac{q(\tilde{a}_i^\ell, \tilde{a}_{i+1}^\ell) \int \gamma_i(\tilde{a}_i^\ell, z') p(y_{i:n} | \tilde{a}_{i:n}^\ell, z') dz'}{\tilde{q}_i(\tilde{a}_{i+1:n}^\ell, \tilde{a}_i^\ell) \int \gamma_{i+1}(\tilde{a}_{i+1}^\ell, z') p(y_{i+1:n} | \tilde{a}_{i+1:n}^\ell, z') dz'}.$$

To obtain uniformly weighted samples at each time step, in the numerical experiments we use:

$$\tilde{q}_n(\cdot) = \int \gamma_n(\cdot, z') g(\cdot, z'; y_n) dz' \quad \text{and} \quad \tilde{q}_i(\tilde{a}_{i+1:n}^\ell, \cdot) = \frac{q(\cdot, \tilde{a}_{i+1}^\ell) \int \gamma_i(\cdot, z') p(y_{i:n} | (\cdot, \tilde{a}_{i+1:n}^\ell), z') dz'}{\int \gamma_{i+1}(\tilde{a}_{i+1}^\ell, z') p(y_{i+1:n} | \tilde{a}_{i+1:n}^\ell, z') dz'}.$$

By (5.15) and (5.16), we get

$$\begin{aligned} p(y_{i:n} | a_i, z_i) &= \frac{\tilde{p}_i(y_{i:n}) \tilde{p}_i(a_i, z_i | y_{i:n})}{\gamma_i(a_i, z_i)}, \\ &= \tilde{p}_i(y_{i:n}) \sum_{a_{i+1:n}=1}^J \frac{\tilde{p}_i(a_{i:n} | y_{i:n}) p(y_{i:n} | a_{i:n}, z_i)}{\int \gamma_i(a_i, z') p(y_{i:n} | a_{i:n}, z') dz'}, \end{aligned}$$

which suggests the following particle approximation $p_N^{\text{two-filt}}(y_{i:n} | a_i, z_i)$ of $p(y_{i:n} | a_i, z_i)$:

$$p_N^{\text{two-filt}}(y_{i:n} | a_i, z_i) = \tilde{p}_i(y_{i:n}) \sum_{\ell=1}^N \frac{\tilde{\omega}_i^\ell p(y_{i:n} | \tilde{a}_{i:n}^\ell, z_i)}{\int \gamma_i(\tilde{a}_i^\ell, z') p(y_{i:n} | \tilde{a}_{i:n}^\ell, z') dz'} \delta_{\tilde{a}_i^\ell}(a_i).$$

The conditional likelihood of the observations given the sequence of states $p(y_{i:n} | a_{i:n}, z_i)$ can be computed explicitly using a Gaussian backward smoother; these computations are summarized in Lemma 2.

Lemma 2. For all $1 \leq i \leq n$,

$$p(y_{i:n} | a_{i:n}, z_i) = \exp \left\{ -\frac{1}{2} \tilde{c}_i(a_{i:n}) - \frac{1}{2} \|z_i\|_{\tilde{P}_i(a_{i:n})}^2 + z_i' \tilde{v}_i(a_{i:n}) \right\}, \quad (5.17)$$

where

$$\tilde{c}_n(a_n) = p \log(2\pi) + \log |\overline{G}_{a_n}| + \|y_n - c_{a_n}\|_{\overline{G}_{a_n}}^2, \quad (5.18)$$

$$\tilde{P}_n^{-1}(a_n) = B_{a_n}' \overline{G}_{a_n}^{-1} B_{a_n}, \quad (5.19)$$

$$\tilde{v}_n(a_n) = B_{a_n}' \overline{G}_{a_n}^{-1} (y_n - c_{a_n}) \quad (5.20)$$

and, for all $1 \leq i \leq n-1$,

$$\tilde{c}_i(a_{i:n}) = \tilde{c}_{i|i+1}(a_{i:n}) + \rho \log(2\pi) + \log |\overline{G}_{a_i}| + \left\| y_i - c_{a_i} \right\|_{\overline{G}_{a_i}}^2, \quad (5.21)$$

$$\tilde{P}_i^{-1}(a_{i:n}) = \tilde{P}_{i|i+1}^{-1}(a_{i:n}) + B'_{a_i} \overline{G}_{a_i}^{-1} B_{a_i}, \quad (5.22)$$

$$\tilde{v}_i(a_{i:n}) = \tilde{v}_{i|i+1}(a_{i:n}) + B'_{a_i} \overline{G}_{a_i}^{-1} (y_i - c_{a_i}), \quad (5.23)$$

with

$$\begin{aligned} \Delta_i(a_{i+1:n}) &= \left(I_m + H'_{a_{i+1}} \tilde{P}_{i+1}^{-1}(a_{i+1:n}) H_{a_{i+1}} \right)^{-1}, \\ \tilde{r}_{i|i+1}(a_{i+1:n}) &= \tilde{v}_{i+1}(a_{i+1:n}) + \overline{H}_{a_{i+1}}^{-1} d_{a_{i+1}}, \\ \tilde{c}_{i|i+1}(a_{i+1:n}) &= \tilde{c}_{i+1}(a_{i+1:n}) + \log |\overline{H}_{a_{i+1}}| + d'_{a_{i+1}} \overline{H}_{a_{i+1}}^{-1} d_{a_{i+1}} - \log |H_{a_{i+1}} \Delta_i(a_{i+1:n}) H'_{a_{i+1}}| \\ &\quad - \tilde{r}'_{i|i+1}(a_{i+1:n}) H_{a_{i+1}} \Delta_i(a_{i+1:n}) H'_{a_{i+1}} \tilde{r}_{i|i+1}(a_{i+1:n}), \\ \tilde{P}_{i|i+1}^{-1}(a_{i+1:n}) &= T'_{a_{i+1}} \left(I_m - \overline{H}_{a_{i+1}}^{-1} H_{a_{i+1}} \Delta_i(a_{i+1:n}) H'_{a_{i+1}} \right) \overline{H}_{a_{i+1}}^{-1} T_{a_{i+1}}, \\ \tilde{v}_{i|i+1}(a_{i+1:n}) &= T'_{a_{i+1}} \overline{H}_{a_{i+1}}^{-1} \left[-d_{a_{i+1}} + H_{a_{i+1}} \Delta_i(a_{i+1:n}) H'_{a_{i+1}} \left(\tilde{v}_{i+1}(a_{i+1:n}) + \overline{H}_{a_{i+1}}^{-1} d_{a_{i+1}} \right) \right]. \end{aligned}$$

Proof. The proof is postponed to Appendix A.2.1. \square

By Lemma 2, the SMC approximation $p_N^{\text{two-filt}}(y_{i:n}|a_i, z_i)$ of $p(y_{i:n}|a_i, z_i)$ is then given by

$$p_N^{\text{two-filt}}(y_{i:n}|a_i, z_i) = \tilde{p}_i(y_{i:n}) \sum_{\ell=1}^N \frac{\delta_{\tilde{a}_i^\ell}(a_i) \tilde{\omega}_i^\ell}{\int \gamma_i(\tilde{a}_i^\ell, z') p(y_{i:n}|\tilde{a}_{i:n}^\ell, z') dz'} \exp \left\{ -\frac{1}{2} \|z_i\|_{\tilde{P}_i^\ell}^2 + z_i' \tilde{v}_i^\ell - \frac{1}{2} \tilde{c}_i^\ell \right\}, \quad (5.24)$$

where $(\tilde{P}_i^\ell)^{-1} := \tilde{P}_i^{-1}(\tilde{a}_{i:n}^\ell)$, $\tilde{v}_i^\ell := \tilde{v}_i(\tilde{a}_{i:n}^\ell)$, $\tilde{c}_i^\ell := \tilde{c}_i(\tilde{a}_{i:n}^\ell)$. Define

$$\begin{aligned} \Delta_i^\ell &:= \left(I + H'_{\tilde{a}_{i+1}^\ell} (\tilde{P}_{i+1}^\ell)^{-1} H_{\tilde{a}_{i+1}^\ell} \right)^{-1}, \\ \delta_i^\ell &:= \tilde{v}_{i+1}^\ell + \overline{H}_{\tilde{a}_{i+1}^\ell}^{-1} (d_{\tilde{a}_{i+1}^\ell} + T_{\tilde{a}_{i+1}^\ell} z_i). \end{aligned}$$

Then, by (5.13), the SMC approximation $t_i^N(a_i, z_i, y_{i+1:n})$ of $t_i(a_i, z_i, y_{i+1:n})$ is given by:

$$\begin{aligned} t_i^N(a_i, z_i, y_{i+1:n}) &= \sum_{a_{i+1}=1}^J \int_{z_{i+1}} m(a_{i+1}, z_i; z_{i+1}) Q(a_i, a_{i+1}) p^N(y_{i+1:n} | a_{i+1}, z_{i+1}) dz_{i+1}, \\ &= \tilde{p}_{i+1}(y_{i+1:n}) \sum_{\ell=1}^N C_i^{-1}(\tilde{a}_{i+1:n}^\ell) Q(a_i, \tilde{a}_{i+1}^\ell) \tilde{\omega}_{i+1}^\ell |\bar{H}_{\tilde{a}_{i+1}^\ell}^{-1}|^{-1/2} |H_{\tilde{a}_{i+1}^\ell}^\ell \Delta_i^\ell H_{\tilde{a}_{i+1}^\ell}^{\prime\ell}|^{1/2} \\ &\quad \times \exp \left\{ \frac{1}{2} (\delta_i^\ell)' H_{\tilde{a}_{i+1}^\ell}^\ell \Delta_i^\ell H_{\tilde{a}_{i+1}^\ell}^{\prime\ell} \delta_i^\ell - \frac{1}{2} \left\| d_{\tilde{a}_{i+1}^\ell}^\ell + T_{\tilde{a}_{i+1}^\ell}^\ell z_i \right\|_{\bar{H}_{\tilde{a}_{i+1}^\ell}^\ell}^2 \right\}, \end{aligned} \quad (5.25)$$

$$= \sum_{\ell=1}^N \tilde{\omega}_{b,i}^\ell(a_i) \exp \left\{ -\frac{1}{2} \|z_i\|_{\tilde{s}_{i+1}^\ell}^2 + z_i' \tilde{s}_{i+1}^\ell \right\}, \quad (5.26)$$

where

$$\begin{aligned} C_i(\tilde{a}_{i+1:n}^\ell) &:= \exp \left\{ \tilde{c}_{i+1}^\ell / 2 \right\} \int_{z_{i+1}} \gamma_{i+1}(\tilde{a}_{i+1}^\ell, z) \tilde{p}(y_{i+1:n} | \tilde{a}_{i+1:n}^\ell, z) dz, \\ \tilde{\omega}_{b,i}^\ell(a_i) &= \tilde{p}_{i+1}(y_{i+1:n}) C_i(\tilde{a}_{i+1:n}^\ell)^{-1} Q(a_i, \tilde{a}_{i+1}^\ell) \tilde{\omega}_{i+1}^\ell |\bar{H}_{\tilde{a}_{i+1}^\ell}^{-1}|^{-1/2} |H_{\tilde{a}_{i+1}^\ell}^\ell \Delta_i^\ell H_{\tilde{a}_{i+1}^\ell}^{\prime\ell}|^{1/2} \\ &\quad \times \exp \left\{ -d_{\tilde{a}_{i+1}^\ell}^{\prime\ell} \bar{H}_{\tilde{a}_{i+1}^\ell}^{-1} d_{\tilde{a}_{i+1}^\ell}^\ell / 2 \right\} \exp \left\{ (\tilde{v}_{i+1}^\ell + \bar{H}_{\tilde{a}_{i+1}^\ell}^{-1} d_{\tilde{a}_{i+1}^\ell}^\ell)' H_{\tilde{a}_{i+1}^\ell}^\ell \Delta_i^\ell H_{\tilde{a}_{i+1}^\ell}^{\prime\ell} (\tilde{v}_{i+1}^\ell + \bar{H}_{\tilde{a}_{i+1}^\ell}^{-1} d_{\tilde{a}_{i+1}^\ell}^\ell) / 2 \right\}, \\ (\tilde{S}_{i+1}^\ell)^{-1} &= T_{\tilde{a}_{i+1}^\ell}^{\prime\ell} \bar{H}_{\tilde{a}_{i+1}^\ell}^{-1} (T_{\tilde{a}_{i+1}^\ell}^\ell - H_{\tilde{a}_{i+1}^\ell}^\ell \Delta_i^\ell H_{\tilde{a}_{i+1}^\ell}^{\prime\ell} \bar{H}_{\tilde{a}_{i+1}^\ell}^{-1} T_{\tilde{a}_{i+1}^\ell}^\ell), \\ \tilde{s}_{i+1}^\ell &= T_{\tilde{a}_{i+1}^\ell}^{\prime\ell} \bar{H}_{\tilde{a}_{i+1}^\ell}^{-1} (H_{\tilde{a}_{i+1}^\ell}^\ell \Delta_i^\ell H_{\tilde{a}_{i+1}^\ell}^{\prime\ell} (\tilde{v}_{i+1}^\ell + \bar{H}_{\tilde{a}_{i+1}^\ell}^{-1} d_{\tilde{a}_{i+1}^\ell}^\ell) - d_{\tilde{a}_{i+1}^\ell}^\ell). \end{aligned}$$

In the experiments (see section 5.4), $\gamma_i(a_i, z_i)$ is set as a mixture of Gaussian distributions. Note that for such a choice, the integral $\int \gamma_i(a_i, z') p(y_{i:n} | a_{i:n}, z') dz'$ may be computed explicitly; see Lemma 3. This choice differs from [10] where it is advocated to set γ_i as the product of two independent densities $\gamma_i^a(a_i)$ and $\gamma_i^z(z_i)$. The accuracy of the algorithm relies heavily on a proper tuning of this artificial density.

Lemma 3. For all $1 \leq i \leq n$,

$$\begin{aligned} \int \phi_{\mu_i, \Sigma_i}(z_i) p(y_{i:n} | a_{i:n}, z_i) dz_i &= \exp \left\{ -\frac{1}{2} \log |\Sigma_i| - \frac{1}{2} \mu_i' \Sigma_i^{-1} \mu_i \right\} \\ &\quad \times \exp \left\{ -\frac{1}{2} \tilde{c}_i(a_{i:n}) + \frac{1}{2} \log |\tilde{\Omega}_i(a_{i:n})| + \frac{1}{2} \tilde{z}_i'(a_{i:n}) \tilde{\Omega}_i(a_{i:n}) \tilde{z}_i(a_{i:n}) \right\}, \end{aligned}$$

where $\phi_{\mu, \Sigma}$ is the probability density function of a m dimensional Gaussian random variable with mean μ and variance matrix Σ and

$$\tilde{\Omega}_i(a_{i:n}) := \left(\Sigma_i^{-1} + \tilde{P}_i^{-1}(a_{i:n}) \right)^{-1} \quad \text{and} \quad \tilde{z}_i(a_{i:n}) := \Sigma_i^{-1} \mu_i + \tilde{v}_i(a_{i:n})$$

and where c_i , \tilde{P}_i and v_i are given in Lemma 2.

Proof. The proof is postponed to Appendix A.2.1. □

5.3 Simulated data

This section highlights the improvements brought by the additional Rao-Blackwellization steps for the two-filter and the FFBS approximations of the marginal smoothing distributions in the case where the number of states is $J = 2$. In each setting the transition matrix Q is such that the probability of switching from one regime to the other is small, as expected for the WTI crude oil data; see Section 5.4. First, the FFBS algorithm is applied to a simple one-dimensional model with:

$$\begin{aligned} \pi(1) = \pi(2) = 1/2; \quad d_1 = 0.5 \quad d_2 = 0; \quad c_1 = c_2 = 0, \\ Q = \begin{pmatrix} 0.99 & 0.01 \\ 0.03 & 0.97 \end{pmatrix} \quad T_1 = T_2 = 1 \quad \bar{H}_1 = \bar{H}_2 = 0.05, \\ B_1 = B_2 = 1 \quad \bar{G}_1 = 0.3 \quad \bar{G}_2 = 0.1. \end{aligned}$$

The two FFBS algorithms are used with $N = \tilde{N} = 50$ and the estimated posterior probabilities $\mathbb{P}(a_k = 1 | Y_{1:n})$ are compared to the estimation obtained with the proposed FFBS algorithm with rejuvenation and 2000 particles. Figure 5.1 displays the mean estimation error over 100 independent Monte Carlo runs. The estimation error is defined as the difference between the the estimation obtained with the proposed FFBS algorithm with rejuvenation and 2000 particles and the estimation given by the two FFBS algorithms with $N = \tilde{N} = 50$. In addition, Figure 5.2 displays the empirical variance of the estimation for the two algorithms to illustrate that both estimators have similar variability.

Then, the performance of the two-filter based algorithms are compared in two different settings with $n = 1500$ observations.

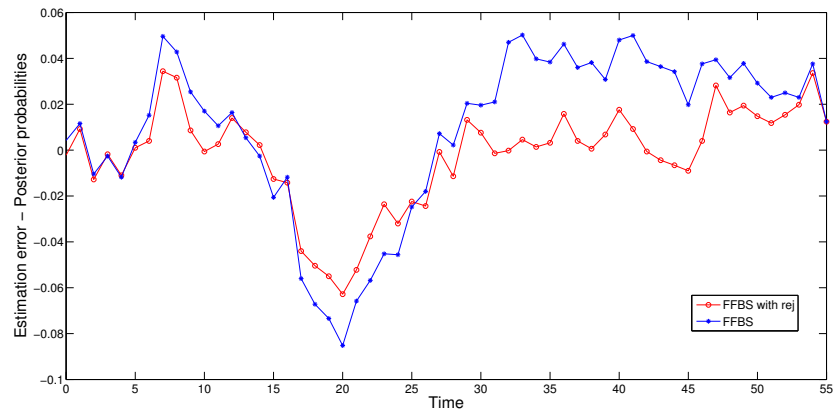


Figure 5.1: Posterior probabilities estimation error for the two FFBS algorithms.

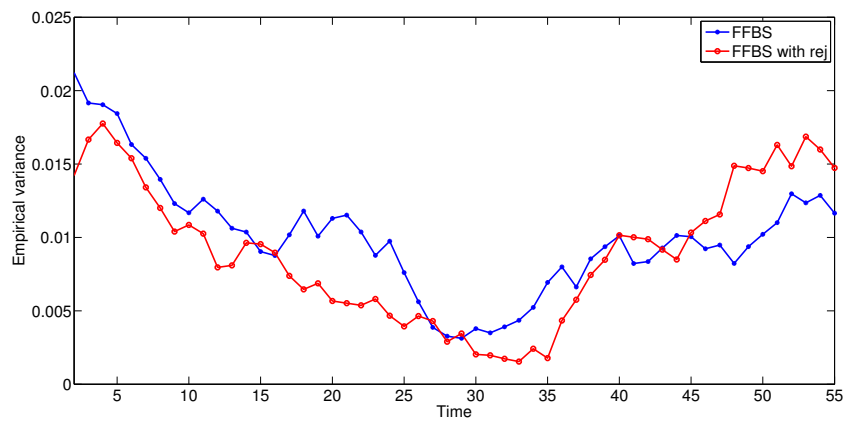


Figure 5.2: Empirical variances of the estimation for the two FFBS algorithms.

Experiment 1:

$$\pi(1) = \pi(2) = 1/2; d_1 = \begin{pmatrix} 0 \\ 0.5 \end{pmatrix} d_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}; c_1 = (0 \ 0 \ 0 \ 0)' \quad c_2 = (0 \ 0 \ 0 \ 0)',$$

$$Q = \begin{pmatrix} 0.99 & 0.01 \\ 0.03 & 0.97 \end{pmatrix} \quad T = \begin{pmatrix} 1 & -0.02 \\ 0 & 0.96 \end{pmatrix} \quad \bar{H}_1 = \begin{pmatrix} 0.25 & 0.01 \\ 0.01 & 0.01 \end{pmatrix} \quad \bar{H}_2 = \begin{pmatrix} 0.3 & 0 \\ 0 & 0.3 \end{pmatrix},$$

$$B = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 0.5 & 0.5 & 0.1 \end{pmatrix}' \quad \bar{G} = \text{diag}(0.1, 0.1, 0.1, 0.1) \quad \mu_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \Sigma_1 = \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix}.$$

The approximations of the smoothing distributions are computed with $N = 10$. RBTF and the algorithm of [10] are compared to the approximations provided by RBTF with 1000 particles.

Experiment 2: In this second experiment, we use the same parameters except that we increase by a factor 10 the values in \bar{G} . The smoothing distributions are estimated with $N = 15$ particles. Both methods are compared to the approximations provided by RBTF with 1000 particles. The specific form of B_1 (first row with entries equal to one) and the fact that $T_1 = T_2$, $B_1 = B_2$ and $\bar{G}_1 = \bar{G}_2$ are also motivated by the application to WTI crude oil data. It is assumed that the differences between the two regimes are given by the state's dynamics (d_1 , d_2 , \bar{H}_1 and \bar{H}_2). Figures 5.3 and 5.4 displays the approximations of the posterior distributions of the regimes given by the algorithms using:

$$\gamma_i(a_i, z_i) = \sum_{k=1}^N Q(a_{i-1}^k, a_i) \omega_{i-1}^k \exp \left\{ -\frac{1}{2} \log |\bar{H}_{a_{i-1}^k}| - \frac{1}{2} \left\| z_i - d_{a_{i-1}^k} - T \mu_{i-1}^k \right\|_{\bar{H}_{a_{i-1}^k}}^2 \right\}.$$

In both cases, RBTF behaves slightly better than [10] which does not detect some regime switchings (for instance at time 780 in Figure 5.3 and at time 300 in Figure 5.4) or gives false detections (for instance at time 10 in Figure 5.4).

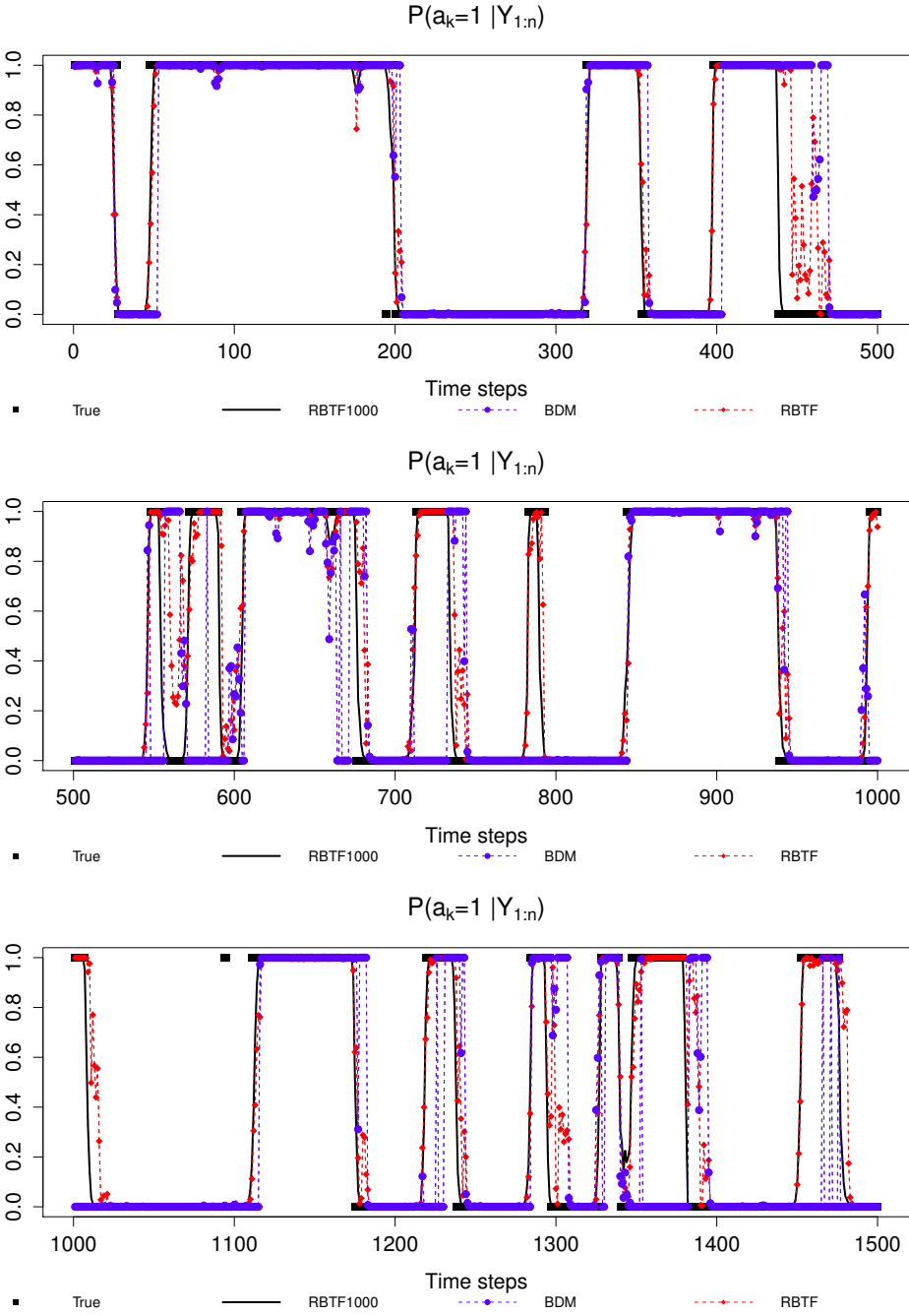


Figure 5.3: Experiment 1 - estimated probabilities by [10] (blue) and RBTF (red) with $N = 10$ particles, RBTF with $N = 1000$ particles (black line) and true regimes (squared black).

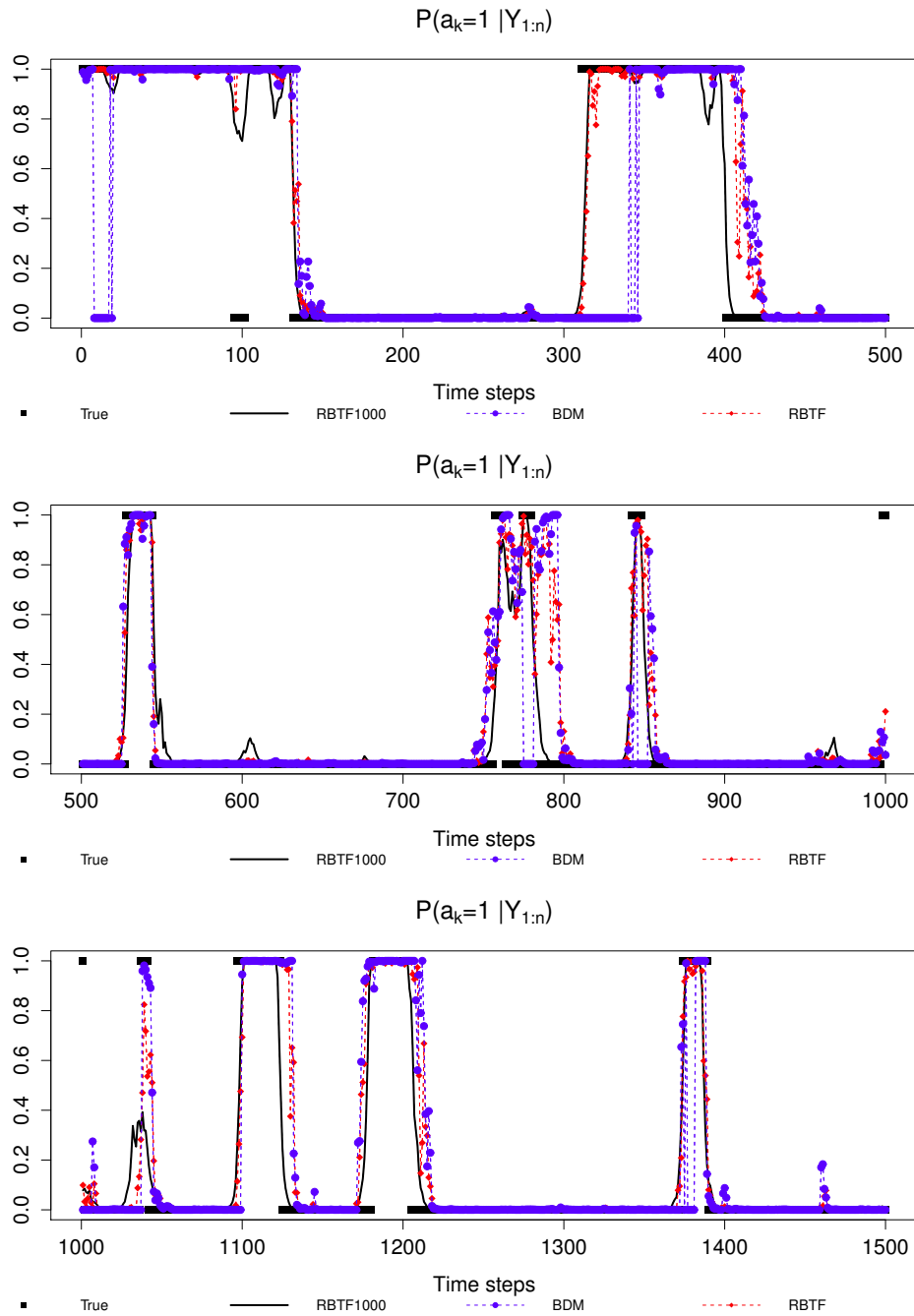


Figure 5.4: Experiment 2 - estimated smoothing probabilities by [10] (blue) and RBTF (red) with $N = 15$ particles, RBTF with $N = 1000$ particles (black line) and true regimes (squared black).

5.4 Application to CME crude oil (WTI)

5.4.1 Model

Modeling commodity prices is a crucial step to value contingent claims related to energy markets and to optimize storage or extraction strategies. In [39, 87], the authors proposed a model where the spot price of a commodity ($S_t, t \geq 0$) depends on a second factor ($\delta_t, t \geq 0$), referred to as the instantaneous convenience yield. This factor plays the role of dividends in equity markets and models the benefit of holding the physical commodity or the storage and maintenance costs required to keep the commodity. In this model, this convenience yield is described as an Ornstein-Uhlenbeck process:

$$\begin{aligned} dS_t &= (r - \delta_t)S_t dt + \sigma S_t dW_t^1, \\ d\delta_t &= \kappa(\alpha - \delta_t)dt + \eta dW_t^2, \quad d\langle W_t^1, W_t^2 \rangle = \rho dt, \end{aligned}$$

where the parameters $(r, \sigma, \kappa, \alpha, \eta, \rho)$ are constant and $((W_t^1, W_t^2), t \geq 0)$ are standard Brownian motions. This model appears to be too restrictive as energy markets are not likely to revert to a single equilibrium value. This assumption is relaxed using Markov switching models to allow several possible regimes for the spot price and the convenience yield. Following [2], the spot price and convenience yield are described in this paper as:

$$\begin{aligned} dS_t &= (r - \delta_t)S_t dt + \sigma_{a_t} S_t dW_t^1, \\ d\delta_t &= \kappa(\alpha_{a_t} - \delta_t)dt + \eta_{a_t} dW_t^2, \quad d\langle W_t^1, W_t^2 \rangle = \rho_{a_t} dt, \end{aligned}$$

where $(a_t)_{t \geq 0}$ is a finite state space Markov process.

This model allows to exhibit fundamental features of commodity future prices, which typically display different regimes of volatility and/or convenience yield. A two-regime model is already sufficient to produce stylized effects such as contango (increase of future prices) and backwardation (decrease of future prices). Assuming that the switching rate between regimes is negligible compared to the inverse of the discretization period, the discretized version of the spot price and convenience yield

$Z_i = (\ln S_i, \delta_i)$ (the sampling period is taken to be 1) is modeled as a CLGM. The explicit integration of this SDE detailed in Lemma 4 yields the following discrete time model for $(Z_i)_{i \geq 0}$:

$$Z_i = d_{a_{i-1}} + TZ_{i-1} + H_{a_{i-1}} \varepsilon_i,$$

where (with $\bar{H}_{a_{i-1}} := H_{a_{i-1}} H'_{a_{i-1}}$ and $\tau = t_i - t_{i-1}$):

$$d_{a_{i-1}} := \begin{pmatrix} \left[\mu - \alpha_{a_{i-1}} - \sigma_{a_{i-1}}^2/2 \right] \tau + \alpha_{a_{i-1}} [1 - e^{-\kappa\tau}] / \kappa \\ \alpha_{a_{i-1}} [1 - e^{-\kappa\tau}] \end{pmatrix},$$

$$T := \begin{pmatrix} 1 & -[1 - e^{-\kappa\tau}] / \kappa \\ 0 & e^{-\kappa\tau} \end{pmatrix},$$

$$\begin{aligned} \bar{H}_{a_{i-1}}(1, 1) &= \sigma_{a_{i-1}}^2 \tau + \eta_{a_{i-1}}^2 \left\{ \tau + (1 - e^{-2\kappa\tau}) / (2\kappa) - 2(1 - e^{-\kappa\tau}) / \kappa \right\} / \kappa^2 \\ &\quad - 2\rho_{a_{i-1}} \eta_{a_{i-1}} \sigma_{a_{i-1}} \left\{ \tau t_i - (1 - e^{-\kappa\tau}) / \kappa \right\} / \kappa, \end{aligned}$$

$$\bar{H}_{a_{i-1}}(1, 2) = \left(\rho_{a_{i-1}} \eta_{a_{i-1}} \sigma_{a_{i-1}} - \eta_{a_{i-1}}^2 / \kappa \right) (1 - e^{-\kappa\tau}) / \kappa + \eta_{a_{i-1}}^2 (1 - e^{-2\kappa\tau}) / (2\kappa^2),$$

$$\bar{H}_{a_{i-1}}(2, 1) = \bar{H}_{a_{i-1}}(1, 2), \quad \bar{H}_{a_{i-1}}(2, 2) = \eta_{a_{i-1}}^2 (1 - e^{-2\kappa\tau}) / (2\kappa).$$

The observations are Wednesday future contracts of the West Texas Intermediate crude oil (WTI) traded in the Chicago Mercantile Exchange (CME) from 11 January 1995 to 13 November 2013. The contracts are numbered F_1, F_2, \dots, F_{36} where F_1 (or front month) is the earliest delivery future contract, F_2 is the second earliest delivery future contract and so on. Among these 36 contracts, the four future contracts: F_1, F_4, F_6, F_{13} are used since their trading volumes and their impacts on the Term Structures are the most important (F_1 is the most liquid contract, F_{13} characterizes the gap between prices over a one year period, F_4 and F_6 are intermediate future contracts that are mostly traded). As in [2], we consider that each future contract has a fixed time to maturity: F_1, F_4, F_6, F_{13} have time to maturity 4, 16, 26, 56 weeks. Our time series contains $n = 975$ weekly data with 534 in backwardation and 441 in contango (the backwardation effect is more frequent with crude oil data). At each time $t_i = i\tau$, with $\tau = 0.0192$, the observations of the $p = 4$ future prices are $Y_i := (\ln(F_{i\tau t, m_1}^{(market)}), \dots, \ln(F_{i\tau t, m_p}^{(market)}))'$, where $F_{t_i, m}$ is the future price at t_i for a maturity m weeks. A closed form solution for $F_{t_i, m}$ may be written:

$$F_{t_i, m} := \exp(A_m(a_i) + B_m Z_i),$$

where $B_0 = \begin{pmatrix} 1 & 0 \end{pmatrix}$ and $B_m = B_{m-1}T$ so that $B_m = \begin{pmatrix} 1 & -(1 - e^{-\kappa m \tau})/\kappa \end{pmatrix}$ and for all $1 \leq j \leq J$, $A_0(j) = 0$, and

$$A_m(j) = \ln \left(\sum_{k=1}^J Q(j, k) \exp(A_{m-1}(k)) \right) + B_{m-1}d_j + \frac{1}{2}B_{m-1}\bar{H}_j B'_{m-1}.$$

Therefore, the observations of the logfuture prices are given, for all $1 \leq i \leq n$, by:

$$Y_i = c_{a_i} + BZ_i + G\eta_i,$$

where ξ_i is a standard multivariate Gaussian random variable and:

$$c'_j = [A_{m_1}(j), \dots, A_{m_p}(j)], \quad B' = [B'_{m_1}, \dots, B'_{m_p}], \quad G = \text{diag}(g_1, \dots, g_d).$$

The model depends of the parameters:

$$\theta := \{\pi, Q, \mu_1, \Sigma_1, \kappa, (\alpha_j)_{1 \leq j \leq J}, (\sigma_j)_{1 \leq j \leq J}, (\eta_j)_{1 \leq j \leq J}, (\rho_j)_{1 \leq j \leq J}, (g_\ell)_{1 \leq \ell \leq d}\}.$$

The aim of this section is to estimate θ and the posterior probabilities $\mathbb{P}(a_k = j | Y_{1:n})$, $1 \leq k \leq n$, $1 \leq j \leq J$. Given the observations $Y_{1:n}$, the EM algorithm introduced in [23] maximizes the incomplete data log-likelihood $\theta \mapsto \ell_\theta^n$ defined by

$$\ell_\theta^n(Y_{1:n}) := \log \left(\sum_{a_1=1}^J \dots \sum_{a_n=1}^J \int p_\theta(a_{1:n}, z_{1:n}, Y_{1:n}) dz_{1:n} \right),$$

where the complete data likelihood p_θ is given by

$$p_\theta(a_{1:n}, z_{1:n}, Y_{1:n}) := \pi(a_1) \phi_{\mu_1, \Sigma_1}(z_1) g_\theta(a_1, z_1; Y_1) \prod_{i=2}^n Q(a_{i-1}, a_i) m_\theta(a_i, z_{i-1}; z_i) g_\theta(a_i, z_i; Y_i),$$

where ϕ_{μ_1, Σ_1} is the Gaussian probability density function with mean μ_1 and variance Σ_1 . Denote by $\mathbb{E}_\theta[\cdot | Y_{1:n}]$ the conditional expectation given $Y_{1:n}$ when the parameter value is set to θ . The EM algorithm iteratively builds a sequence $\{\theta_p\}_{p \geq 0}$ of parameter estimates following the two steps:

1. **E-step**: compute $\theta \mapsto Q(\theta, \theta_p) := \mathbb{E}_{\theta_p} [\log p_\theta(X_{1:n}, Y_{1:n}) | Y_{1:n}]$;
2. **M-step**: choose θ_{p+1} as a maximizer of $\theta \mapsto Q(\theta, \theta_p)$.

All the conditional expectations involved in $Q(\theta, \theta_p)$ are approximated using our RBTF algorithm to define the SMC approximation $\theta \mapsto Q^N(\theta, \theta_p)$ of $\theta \mapsto Q(\theta, \theta_p)$. As the function $\theta \mapsto Q^N(\theta, \theta_p)$ cannot be maximized analytically, the M-step is performed numerically using the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) introduced in [47]. This derivative-free optimization procedure is known to perform well in complex multimodal optimization settings, see e.g. [46].

5.4.2 Numerical results

The initial transition probability in CMA-ES is chosen as $Q(1,1) = 0.98$, $Q(2,2) = 0.97$ and $\pi(1) = \pi(2) = 0.5$ where the number 1 represents the backwardation regime and 2 represents the contango regime. The other parameters are initialized as shown in the following table:

κ	α_1	α_2	σ_1	σ_2	η_1	η_2	ρ_1	ρ_2	g_1	g_2	g_3	g_4
5.0	0.1	-0.05	0.4	0.4	0.5	0.5	0.75	0.65	0.1	0.1	0.1	0.1

Figure 5.5: Initial values for the EM algorithm.

The number of particles is set to $N = 100$, $\tau = 1/52$. The interest rate is $r = 0.0296$ as in [2]. The initial guess for the mean and variance of the initial state are

$$\mu_1 = \left(\ln F_{1,4}^{(\text{market})} \quad r - \frac{\ln F_{1,16}^{(\text{market})} - \ln F_{1,4}^{(\text{market})}}{(16-4)\tau} \right) \quad \text{and} \quad \Sigma_1 = \begin{pmatrix} 0.05 & 0 \\ 0 & 0.05 \end{pmatrix}.$$

The CMA-ES algorithm is used with an initial standard deviation for the parameters $\sigma_{cmaes} = 0.005$, a number of selected search points $\mu_{cmaes} = 20$ and a population size $\lambda_{cmaes} = 100$. The algorithm is stopped after 10000 iterations. In Gibson-Schwartz model [39], a stronger backwardation effect implies a greater value for α for the same values of the other parameters. For the CME WTI Crude Oil, backwardation effect is more frequent than contango effect so that α_1 should be greater than α_2 . Therefore, this condition is imposed for all simulations in the CMA-ES algorithm. The results after 2500 iterations of the EM algorithm are given in Figure 5.1.

CHAPTER 5. REGIME SWITCHING HIDDEN MARKOV MODELS

Parameter	κ	σ_1	σ_2	η_1	η_2	ρ_1	ρ_2
Value	2.6378	0.3733	0.3485	0.5892	0.3814	0.8709	0.6761
Std. Dev	0.1999	0.005438	0.002884	0.0483	0.0349	0.0064	0.0052

Parameter	α_1	α_2	g_1	g_2	g_3	g_4	$Q(1,1)$	$Q(2,2)$
Value	0.0889	-0.0281	2.3e-2	1.0e-4	3.0e-4	2.3e-2	0.9917	0.9880
Std. Dev	0.004248	0.00149	1.9e-4	2.6e-4	2.3e-4	2.1e-04	6.7e-4	9.6e-4

Table 5.1: Final estimates after 3500 iterations.

As expected, we obtain $\sigma_1 \geq \sigma_2$, $\alpha_1 \geq \alpha_2$, $\eta_1 \geq \eta_2$ and $\rho_1 \geq \rho_2$ at convergence of the EM algorithm. Moreover, $Q(1,1) > Q(2,2)$ is corresponding to the prediction that we did from the data description. From the results, $\sigma_1 \geq \sigma_2$, $\alpha_1 \geq \alpha_2$ indicates the first regime (backwardation) characterized by a higher value in both volatility and equilibrium level of convenience yield, and the second regime (contango) characterized by a lower value in both volatility and equilibrium convenience yield level. This in accordance with the theory of storage that the volatility of the commodity spot price is high when the inventory is low, and the convenience yield is all the higher as inventory is low.

Figure 5.6 compares the evolution of future 1M (the nearest contracts) to the *term structure* observed from CME WTI crude oil, defined as the difference of future 13M and future 1M (to avoid seasonality). The figure shows that it is not necessary to have an inverse relationship between the price of the nearest contract and the term structure. But when a significant drop in the price of the nearest contract occurs, the term structure increases (i.e. in contango).

The correlation between the spot price and the convenience yield is positive and high in both two regimes. This is an accordance to what as been observed in most commodity market; see [39]. The slope of future curve decreases in function of maturity.

Figure 5.7 and 5.8 displays the the estimated posterior probabilities of the regimes and the observed future slope. When the future curve is in backwardation (resp. contango), the model is expected to be in the first regime (resp. second regime),

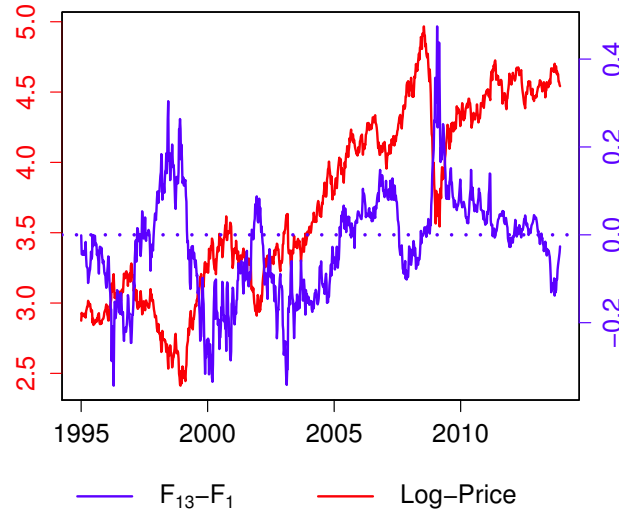


Figure 5.6: Log-price (red line) and slope of future curves (blue line)

except for the period where the slope of the future curve is too small and in the period from December 2008 to April 2009 (beginning of the crisis).

5.5 Conclusions

In this paper, we proposed a survey of Rao Blackwellized Sequential Monte Carlo methods to approximate smoothing distributions in conditionally linear and Gaussian state spaces. These methods combine particle based approximations to sample the sequence of hidden regimes while the hidden linear states are explicitly integrated conditional on the sequence of regimes and observations. They are based on the forward backward decomposition of the smoothing distributions using either two-filter based algorithms or forward filtering backward smoothing techniques mimicking the Rauch-Tung-Striebel smoother. The paper presents these algorithms in a common unifying framework and also provides a survey of different techniques that could be used in the forward filtering pass which may improve significantly the usual mixture Kalman filter. The filtering distributions are approximated at each time step

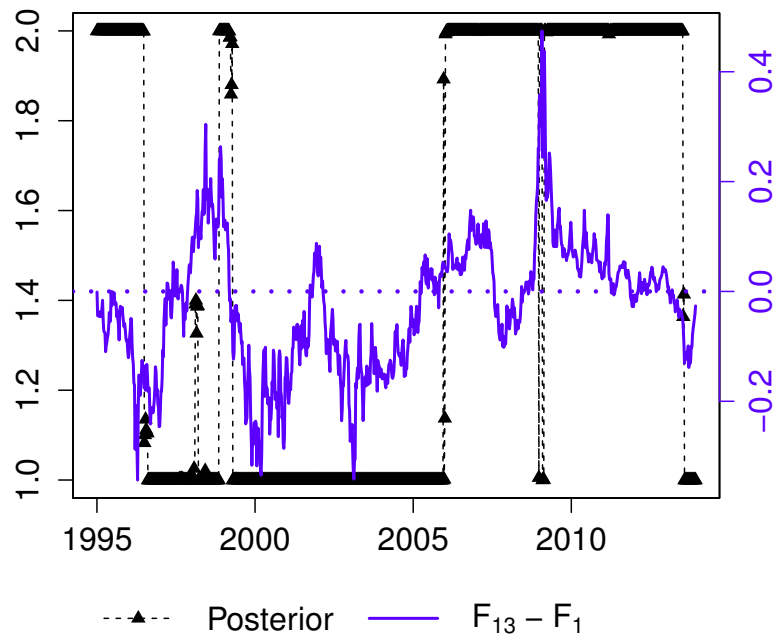


Figure 5.7: Posterior probability (triangle black) and slope of future curves (blue line)

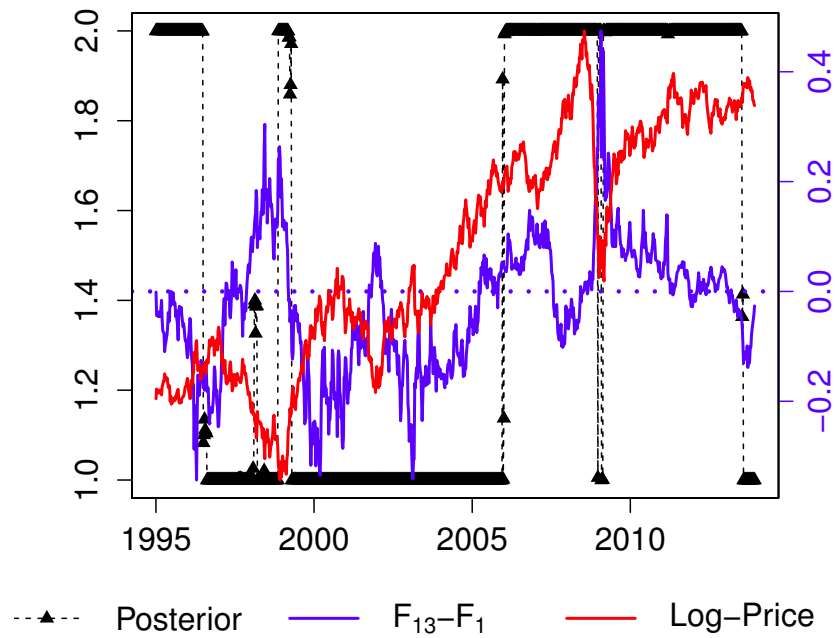


Figure 5.8: Posterior probability (triangle black) and slope of future curves (blue line)

by considering all possible offsprings of all ancestral trajectories before discarding degenerated paths instead of resampling the ancestral paths before propagating them at the next time step. We also investigated the benefit of additional Rao Blackwellization steps which allow to sample new regimes at each time step conditional on the forward and backward particles. This rejuvenation step uses explicit integration of the hidden linear states before merging the forward and backward filters for two-filter based algorithms or before sampling new states backward in time for FFBS based methods. The paper displays some Monte Carlo experiments with simulated data to illustrate that this additional rejuvenation step improves the performance of the smoothing algorithms with no additional computational costs or variance. They are also applied to commodity markets using WTI crude oil data.

Appendix A

Appendices

A.1 Appendix of Chapter 4

A.1.1 Exponential deviation inequalities for the forward filter and the backward information filter

The following result is proved in [24].

Lemma 7. Assume that a_N , b_N , and b are random variables defined on the same probability space such that there exist positive constants β , B , C , and M satisfying

- (i) $|a_N/b_N| \leq M$, \mathbb{P} -a.s. and $b \geq \beta$, \mathbb{P} -a.s.,
- (ii) For all $\epsilon > 0$ and all $N \geq 1$, $\mathbb{P}[|a_N| > \epsilon] \leq B e^{-CN(\epsilon/M)^2}$,
- (iii) For all $\epsilon > 0$ and all $N \geq 1$, $\mathbb{P}[|b_N - b| > \epsilon] \leq B e^{-CN\epsilon^2}$.

Then, for all $\epsilon > 0$,

$$\mathbb{P}\left[\left|\frac{a_N}{b_N}\right| > \epsilon\right] \leq B \exp\left\{-CN\left(\frac{\epsilon\beta}{2M}\right)^2\right\}.$$

Proposition 8 provides an exponential deviation inequality for the forward filter and is proved in [24].

Proposition 8. Assume that A1 and A2 hold for some $T > 0$. Then, for all $s \geq 1$, there exist $0 < B_s, C_s < \infty$ such that for all $N \geq 1$, $\epsilon > 0$, and all $h \in \mathbb{F}_b(\mathcal{X}, \mathcal{X})$,

$$\mathbb{P}\left(\left|\Omega_s^{-1} \sum_{i=1}^N \omega_s^i h(\xi_s^i) - \phi_{\mathcal{X},s}[h]\right| \geq \epsilon\right) \leq B_s e^{-C_s N \epsilon^2 / \text{osc}(h)^2}.$$

Proposition 9 provides an exponential inequality for the backward information filter $\psi_{\gamma,t|T}$ and its unnormalized approximation. Its proof is similar to the proof [24, Theorem 5] and is omitted.

Proposition 9. Assume that A1 and A3 hold for some $T > 0$. Then, for all $0 \leq t \leq T$, there exist $0 < B_{t|T}, C_{t|T} < \infty$ such that for all $N \geq 1$, $\epsilon > 0$, and all $h \in \mathbb{F}_b(\mathcal{X}, \mathcal{X})$,

$$\mathbb{P}\left[\left|\check{\Omega}_{t|T}^{-1} \sum_{i=1}^N \check{\omega}_{t|T}^i h(\check{\xi}_{t|T}^i) - \psi_{\gamma,t|T}[h]\right| \geq \epsilon\right] \leq B_{t|T} e^{-C_{t|T} N \epsilon^2 / \text{osc}(h)^2}.$$

A.1.2 Asymptotic normality of the forward filter and the backward information filter

Proposition 10 provides a CLT for the weighted particles $\{(\omega_s^i, \xi_s^i)\}_{i=1}^N$ approximating the filtering distribution $\phi_{\mathcal{X},s}$ and is proved for instance in [20].

Proposition 10. Assume that A1 and A2 hold for some $T > 0$. Then, for all $0 \leq s \leq T$ and all $h \in \mathbb{F}_b(\mathcal{X}, \mathcal{X})$,

$$N^{1/2} \left(\sum_{i=1}^N \frac{\omega_s^i}{\Omega_s} h(\xi_s^i) - \phi_{\mathcal{X},s}[h] \right) \xrightarrow{\mathcal{D}}_{N \rightarrow \infty} \mathcal{N}\left(0, \Gamma_{\mathcal{X},s}[h - \phi_{\mathcal{X},s}[h]]\right),$$

where

$$\begin{aligned} \Gamma_{\mathcal{X},0}[h] &:= \frac{\int \rho_0(dx_0) \omega_0^2(x_0) h^2(x_0)}{\left(\int \rho_0(dx_0) \omega_0(x_0)\right)^2} \quad \text{and for all } s \geq 1, \\ \Gamma_{\mathcal{X},s}[h] &:= \frac{\Gamma_{\mathcal{X},s-1} \left[\int q(\cdot, x_s) g_s(x_s) h(x_s) dx_s \right]}{\phi_{\mathcal{X},s-1}^2 \left[\int q(\cdot, x_s) g_s(x_s) dx_s \right]} \\ &\quad + \frac{\phi_{\mathcal{X},s-1} \left[\int \omega_s(\cdot, x_s) q(\cdot, x_s) g_s(x_s) h^2(x_s) dx_s \right] \phi_{\mathcal{X},s-1}[\vartheta_s]}{\phi_{\mathcal{X},s-1}^2 \left[\int q(\cdot, x_s) g_s(x_s) dx_s \right]}. \end{aligned}$$

Proposition 11 provides a CLT for the weighted particles $\{(\check{\omega}_{t|T}^j, \check{\xi}_{t|T}^j)\}_{j=1}^N$ approximating the backward information filter. Its proof follows the same lines as the proof of Proposition 10 and is omitted for brevity.

Proposition 11. Assume that A1 and A3 hold. Then, for all $0 \leq t \leq T$ and all $h \in \mathbb{F}_b(X, \mathcal{X})$,

$$N^{1/2} \left(\sum_{j=1}^N \frac{\check{\omega}_{t|T}^j}{\check{\Omega}_{t|T}} h(\check{\xi}_{t|T}^j) - \psi_{\gamma, t|T}[h] \right) \xrightarrow{D}_{N \rightarrow \infty} \mathcal{N} \left(0, \check{\Gamma}_{\gamma, t|T} [h - \psi_{\gamma, t|T}[h]] \right),$$

where

$$\begin{aligned} \check{\Gamma}_{\gamma, T|T}[h] &:= \frac{\int \check{\rho}_T(dx_T) \check{\omega}_{T|T}^2(x_T) h^2(x_T)}{\left(\int \check{\rho}_T(dx_T) \check{\omega}_{T|T}(x_T) \right)^2} \quad \text{and for all } t \leq T-1, \\ \check{\Gamma}_{\gamma, t|T}[h] &:= \frac{\check{\Gamma}_{\gamma, t+1|T} \left[\int \gamma_t(x_t) g_t(x_t) q(x_t, \cdot) \gamma_{t+1}^{-1}(\cdot) h(x_t) dx_t \right]}{\psi_{\gamma, t+1|T}^2 \left[\int \gamma_t(x_t) g_t(x_t) q(x_t, \cdot) \gamma_{t+1}^{-1}(\cdot) dx_t \right]} \\ &\quad + \frac{\psi_{\gamma, t+1|T} \left[\int \check{\omega}_t(x_t, \cdot) q(x_t, \cdot) g_t(x_t) \gamma_t(x_t) \gamma_{t+1}^{-1}(\cdot) h^2(x_t) dx_t \right] \psi_{\gamma, t+1|T} \left[\mathcal{D}_{t|T} \gamma_{t+1}^{-1} \right]}{\psi_{\gamma, t+1|T}^2 \left[\int \gamma_t(x_t) g_t(x_t) q(x_t, \cdot) \gamma_{t+1}^{-1}(\cdot) dx_t \right]}. \end{aligned}$$

A.2 Appendix of Chapter 5

A.2.1 Proofs of lemmas 1-3

Lemmas 1, 2 and 3 are close to [10, Proposition 5, Proposition 6]. The proofs are detailed in this appendix for completeness.

APPENDIX A. APPENDICES

Proof of Lemma 1. For all $1 \leq i \leq n-1$,

$$\begin{aligned}
 p_{\theta}(y_{i:n}|a_i, z_i) &= \sum_{a_{i+1:n}} \int p_{\theta}(y_{i:n}, a_{i+1:n}, z_{i+1:n}|a_i, z_i) dz_{i+1:n}, \\
 &= \sum_{a_{i+1:n}} \int p_{\theta}(a_{i+1:n}, z_{i+1:n}|a_i, z_i) p_{\theta}(y_{i:n}|a_{i:n}, z_{i:n}) dz_{i+1:n}, \\
 &= \frac{\tilde{p}_{\theta,i}(y_{i:n})}{\gamma_i^{\theta}(a_i, z_i)} \sum_{a_{i+1:n}} \int \frac{\gamma_i^{\theta}(a_i, z_i)}{\tilde{p}_{\theta,i}(y_{i:n})} p_{\theta}(a_{i+1:n}, z_{i+1:n}|a_i, z_i) p_{\theta}(y_{i:n}|a_{i:n}, z_{i:n}) dz_{i+1:n}, \\
 &= \frac{\tilde{p}_{\theta,i}(y_{i:n})}{\gamma_i^{\theta}(a_i, z_i)} \sum_{a_{i+1:n}} \int \tilde{p}_{\theta,i}(a_{i:n}, z_{i:n}|y_{i:n}) dz_{i+1:n}, \\
 &= \frac{\tilde{p}_{\theta,i}(y_{i:n})}{\gamma_i^{\theta}(a_i, z_i)} \tilde{p}_{\theta,i}(a_i, z_i|y_{i:n}),
 \end{aligned}$$

which concludes the proof of (5.15). To prove (5.16) write,

$$\begin{aligned}
 \tilde{p}_{\theta,i}(a_{i:n}, z_i|y_{i:n}) &= \frac{\gamma_i^{\theta}(a_i, z_i)}{\tilde{p}_{\theta}(y_{i:n})} \int p_{\theta}(y_{i:n}|a_{i:n}, z_{i:n}) p_{\theta}(a_{i+1:n}, z_{i+1:n}|a_i, z_i) dz_{i+1:n}, \\
 &= \frac{\gamma_i^{\theta}(a_i, z_i)}{\tilde{p}_{\theta,i}(y_{i:n})} \int \frac{p_{\theta}(y_{i:n}|a_{i:n}, z_i) p_{\theta}(z_{i+1:n}|y_{i:n}, a_{i:n}, z_i)}{p_{\theta}(z_{i+1:n}|a_{i:n}, z_i)} p_{\theta}(a_{i+1:n}, z_{i+1:n}|a_i, z_i) dz_{i+1:n}, \\
 &= \frac{\gamma_i^{\theta}(a_i, z_i)}{\tilde{p}_{\theta,i}(y_{i:n})} p_{\theta}(y_{i:n}|a_{i:n}, z_i) p_{\theta}(a_{i+1:n}|a_i).
 \end{aligned}$$

Therefore,

$$\tilde{p}_{\theta,i}(a_i, z_i|y_{i:n}) = \frac{\gamma_i^{\theta}(a_i, z_i)}{\tilde{p}_{\theta,i}(y_{i:n})} \sum_{a_{i+1:n}} p_{\theta}(y_{i:n}|a_{i:n}, z_i) p_{\theta}(a_{i+1:n}|a_i)$$

and the proof is completed upon noting that

$$\tilde{p}_{\theta,i}(a_{i:n}|y_{i:n}) = \frac{p_{\theta}(a_{i+1:n}|a_i)}{\tilde{p}_{\theta,i}(y_{i:n})} \int \gamma_i^{\theta}(a_i, z) p_{\theta}(y_{i:n}|a_{i:n}, z) dz$$

□

Proof of Lemma 2. The result is proved by backward induction. (5.18), (5.19) and (5.20) follow directly from (5.2). Assume that for a given $1 \leq i \leq n-1$, $p(y_{i+1:n}|a_{i+1:n}, z_{i+1})$ is given by (5.17). Write

$$p(y_{i:n}|a_{i:n}, z_i) = \int p(z_{i+1}|z_i, a_i) p(y_i|z_i, a_i) p(y_{i+1:n}|a_{i+1:n}, z_{i+1}) dz_{i+1},$$

with

$$\begin{aligned}
 p(z_{i+1}|z_i, a_i) &= \exp \left\{ -\frac{m}{2} \log(2\pi) - \frac{1}{2} \log |\overline{H}_{a_i}| - \frac{1}{2} \|z_{i+1} - d_{a_i} - Tz_i\|_{\overline{H}_{a_i}}^2 \right\}, \\
 p(y_i|z_i, a_i) &= \exp \left\{ -\frac{p}{2} \log(2\pi) - \frac{1}{2} \log |\overline{G}| - \frac{1}{2} \|y_i - c_{a_i} - Bz_i\|_{\overline{G}}^2 \right\}, \\
 p(y_{i+1:n}|a_{i+1:n}, z_{i+1}) &= \exp \left\{ -\frac{1}{2} c_{i+1}(a_{i+1:n}) - \frac{1}{2} \|z_{i+1}\|_{\tilde{P}_{i+1}(a_{i+1:n})}^2 + z'_{i+1} \nu_{i+1}(a_{i+1:n}) \right\}.
 \end{aligned}$$

Let Δ_i and δ_i be given by:

$$\begin{aligned}
 \Delta_i(a_{i:n}) &:= \left(I_m + H'_{a_i} \tilde{P}_{i+1}^{-1}(a_{i+1:n}) H_{a_i} \right)^{-1}, \\
 \delta_i(a_{i:n}) &:= \nu_{i+1}(a_{i+1:n}) + \overline{H}_{a_i}^{-1} (d_{a_i} + Tz_i).
 \end{aligned}$$

Then, $\overline{H}_{a_i}^{-1} + \tilde{P}_{i+1}^{-1}(a_{i+1:n}) = \left(H_{a_i} \Delta_i(a_{i:n}) H'_{a_i} \right)^{-1}$ and (5.21), (5.22) and (5.23) follows from

$$\begin{aligned}
 &\int \exp \left\{ -\frac{1}{2} \|z_{i+1}\|_{H_{a_i} \Delta_i(a_{i:n}) H'_{a_i}}^2 + z'_{i+1} \delta_i(a_{i:n}) \right\} dz_{i+1} \\
 &= \exp \left\{ \frac{1}{2} \log(2\pi) + \frac{1}{2} \log |H_{a_i} \Delta_i(a_{i:n}) H'_{a_i}| + \frac{1}{2} \delta_i'(a_{i:n})' H_{a_i} \Delta_i(a_{i:n}) H'_{a_i} \delta_i(a_{i:n}) \right\}.
 \end{aligned}$$

□

Proof of Lemma 3. By Lemma 2,

$$\begin{aligned}
 \phi_{\mu_i, \Sigma_i}(z_i) p(y_{i:n}|a_{i:n}, z_i) &= \exp \left\{ -\frac{m}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma_i| - \frac{1}{2} \|z_i - \mu_i\|_{\Sigma_i}^2 \right\} \\
 &\quad \times \exp \left\{ -\frac{1}{2} c_i(a_{i:n}) - \frac{1}{2} \|z_i\|_{\tilde{P}_i(a_{i:n})}^2 + z'_i \nu_i(a_{i:n}) \right\}.
 \end{aligned}$$

The proof is completed noting that

$$\begin{aligned}
 &\int \exp \left\{ -\frac{1}{2} z'_i \tilde{\Omega}_i^{-1}(a_{i:n}) z_i + z'_i (\Sigma_i^{-1} \mu_i + \nu_i(a_{i:n})) \right\} dz_i \\
 &= \exp \left\{ \frac{m}{2} \log(2\pi) + \frac{1}{2} \log |\tilde{\Omega}_i(a_{i:n})| + \frac{1}{2} \left[\Sigma_i^{-1} \mu_i + \nu_i(a_{i:n}) \right]' \tilde{\Omega}_i(a_{i:n}) \left[\Sigma_i^{-1} \mu_i + \nu_i(a_{i:n}) \right] \right\}.
 \end{aligned}$$

□

A.2.2 SDE explicit integration

Lemma 4. *Let $(X_t, \delta_t)_{t \geq 0}$ be solutions to the following SDE:*

$$\begin{aligned} dX_t &= (\mu - \delta_t - \sigma^2/2)dt + \sigma dW_t^1, \\ d\delta_t &= \kappa(\alpha - \delta_t)dt + \eta dW_t^2, \end{aligned}$$

$(W_t^1)_{t \geq 0}$ and $(W_t^2)_{t \geq 0}$ are standard Brownian motions such that $d\langle W_t^1, W_t^2 \rangle = \rho dt$. Then, for all $t \geq 0$ and $h > 0$,

$$\begin{pmatrix} X_{t+h} \\ \delta_{t+h} \end{pmatrix} = d_h + T_h \begin{pmatrix} X_t \\ \delta_t \end{pmatrix} + H_h \varepsilon,$$

where ε is a standard 2-dimensional Gaussian random variable and (with $\bar{H}_h := H_h' H_h$),

$$d := \begin{pmatrix} [\mu - \alpha - \sigma^2/2]h + \alpha[1 - e^{-\kappa h}]/\kappa \\ \alpha[1 - e^{-\kappa h}] \end{pmatrix}, \quad T_h := \begin{pmatrix} 1 & -[1 - e^{-\kappa h}]/\kappa \\ 0 & e^{-\kappa h} \end{pmatrix},$$

$$\begin{aligned} \bar{H}_h(1,1) &:= \sigma^2 h + \eta^2 \left\{ h + (1 - e^{-2\kappa h})/(2\kappa) - 2(1 - e^{-\kappa h})/\kappa \right\} / \kappa^2 \\ &\quad - 2\rho\eta\sigma \left\{ h - (1 - e^{-\kappa h})/\kappa \right\} / \kappa, \\ \bar{H}_h(1,2) &:= (\rho\eta\sigma - \eta^2/\kappa)(1 - e^{-\kappa h})/\kappa + \eta^2(1 - e^{-2\kappa h})/(2\kappa^2), \\ \bar{H}_h(2,1) &:= \bar{H}_h(1,2), \\ \bar{H}_h(2,2) &:= \eta^2(1 - e^{-2\kappa h})/(2\kappa). \end{aligned}$$

Proof. For all $t \geq 0$,

$$X_t = X_0 + (\mu - \sigma^2/2)t - \int_0^t \delta_s ds + \sigma W_t^1$$

and, as $(\delta_t)_{0 \leq t \leq T}$ is an Ornstein-Uhlenbeck process,

$$\delta_t = \delta_0 e^{-\kappa t} + \alpha(1 - e^{-\kappa t}) + \int_0^t \eta e^{\kappa(s-t)} dW_s^2.$$

Then,

$$\begin{aligned}\int_0^t \delta_s ds &= (\delta_0 - \alpha)(1 - e^{-\kappa t})/\kappa + \alpha t + \eta \int_0^t \int_0^s e^{\kappa(u-s)} dW_u^2 ds, \\ &= (\delta_0 - \alpha)(1 - e^{-\kappa t})/\kappa + \alpha t + (\eta/\kappa) \int_0^t (1 - e^{-\kappa(t-s)}) dW_s^2.\end{aligned}$$

Defining $\tilde{W}_t^1 := -(\eta/\kappa) \int_0^t (1 - e^{-\kappa(t-s)}) dW_s^2 + \sigma W_t^1$ and $\tilde{W}_t^2 := \int_0^t \eta e^{\kappa(s-t)} dW_s^2$, this yields:

$$\begin{aligned}X_t &= X_0 + (\mu - \sigma^2/2)t + (\alpha - \delta_0)(1 - e^{-\kappa t})/\kappa - \alpha t + \tilde{W}_t^1, \\ \delta_t &= \delta_0 e^{-\kappa t} + \alpha(1 - e^{-\kappa t}) + \tilde{W}_t^2.\end{aligned}$$

The proof is concluded upon noting that \tilde{W}_t^1 and \tilde{W}_t^2 are centered Gaussian random variables such that:

- $\text{Var}[\tilde{W}_t^1] = \sigma^2 t + \eta^2 \left\{ t + (1 - e^{-2\kappa t})/(2\kappa) - 2(1 - e^{-\kappa t})/\kappa \right\} / \kappa^2 - 2\rho\eta\sigma \left\{ t - (1 - e^{-\kappa t})/\kappa \right\} / \kappa,$
- $\text{Var}[\tilde{W}_t^2] = \eta^2 (1 - e^{-2\kappa t}) / (2\kappa),$
- $\text{Cov}[\tilde{W}_t^1, \tilde{W}_t^2] = (\rho\eta\sigma - \eta^2/\kappa) (1 - e^{-\kappa t}) / \kappa + \eta^2 (1 - e^{-2\kappa t}) / (2\kappa^2).$

□

Bibliography

- [1] Aiube, T.K.N. Baidya, F. and Tito, E. (2008). Analysis of commodity prices with the particle filter. *Energy Economics*, 30(2):597–605.
- [2] Almansour, A. (2016). Convenience yield in commodity price modeling: A regime switching approach. *Energy Economics*, 53:238–247.
- [3] Barber, D. (2006). Expectation correction for smoothed inference in switching linear dynamical systems. *Journal of Machine Learning Research*, 7:2515–2540.
- [4] Barembruch, S., Garivier, A., and Moulines, E. (2008). On optimal sampling for particle filtering in digital communication. *IEEE 9th Workshop on Signal Processing Advances in Wireless*, pages 634–638.
- [5] Baum, L. and Petrie, T. (1966). Ann. math. statist. *Statistical inference for probabilistic functions of finite state Markov chains*, 37:1554–1563.
- [6] Black, F. (1976). The pricing of commodity contracts. *Journal of Financial Economics*, 3(1-2):167–179.
- [7] Brennan, M. (1958). The supply of storage. *The American Economic Review*, 48(1):50–72.
- [8] Brennan, M. and Schwartz, E. (1985). Evaluating natural resource investments. *Journal of Business*, 58(2):135–57.
- [9] Bresler, Y. (1986). Two filter formulae for discrete time nonlinear bayesian smoothing. *International journal of control*, 43(2):629–641.

BIBLIOGRAPHY

- [10] Briers, M., Doucet, A., and Maskell, S. (2010). Smoothing algorithms for state-space models. *Annals Institute Statistical Mathematics*, 62(1):61–89.
- [11] Bryson, A. and Frazier, M. (1963). Smoothing for linear and nonlinear dynamic systems. *Proc. Optimum Sys. Synthesis Conf.*
- [12] Cai, J. (1994). A markov model of switching-regime arch. *Journal of Business & Economic Statistics*, 12(3):309–316.
- [13] Cappé, O. (2011). Online EM algorithm for hidden Markov models. *Journal of Computational and Graphical Statistics*, 20(73):728–749.
- [14] Cappé, O., Moulines, E., and Rydén, T. (2005). *Inference in Hidden Markov Models*. Springer.
- [15] Casassus, J. and Collin-Dufresne, P. (2005). Stochastic convenience yield implied from commodity futures and interest rates. *The Journal of Finance*, 60(5):2283–2331.
- [16] Chen, R. and Liu, J. (2000). Mixture kalman filters. *Journal of the Royal Statistical Society B*, 62:493–508.
- [17] Chiarella, C., Clewlow, L., and Kang, B. (2009). Modelling and estimating the forward price curve in the energy market. *Quantitative Finance Research Center*.
- [18] Clewlow, L. and Strickland, C. (1999a). A multi-factor model for energy derivatives. *QFRC Research Paper Series 28*.
- [19] Clewlow, L. and Strickland, C. (1999b). Valuing energy options in a one factor model fitted to forward prices. *QFRC Research Paper Series 10*.
- [20] Del Moral, P. (2004). *Feynman-Kac Formulae. Genealogical and Interacting Particle Systems with Applications*. Springer.
- [21] Del Moral, P. (2013). *Mean field simulation for Monte Carlo integration*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability.

-
- [22] Del Moral, P., Doucet, A., and Singh, S. (2010). A Backward Particle Interpretation of Feynman-Kac Formulae. *ESAIM M2AN*, 44(5):947–975.
- [23] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. B*, 39(1):1–38 (with discussion).
- [24] Douc, R., Garivier, A., Moulines, E., and Olsson, J. (2011). Sequential Monte Carlo smoothing for general state space hidden Markov models. *Ann. Appl. Probab.*, 21(6):2109–2145.
- [25] Douc, R. and Moulines, E. (2008). Limit theorems for weighted samples with applications to sequential Monte Carlo methods. *Ann. Statist.*, 36(5):2344–2376.
- [26] Douc, R., Moulines, E., and Stoffer, D. (2014). *Nonlinear time series: theory, methods and applications with R examples*. CRC Press.
- [27] Doucet, A., de Freitas, N., and Gordon, N., editors (2001a). *Sequential Monte Carlo methods in practice*. Springer.
- [28] Doucet, A., Godsill, S., and Andrieu, C. (2000). On sequential Monte-Carlo sampling methods for Bayesian filtering. *Stat. Comput.*, 10:197–208.
- [29] Doucet, A., Gordon, N., and Krishnamurthy, V. (2001b). Particle filters for state estimation of jump Markov linear systems. *IEEE Transactions on Signal Processing*, 49(3):613–624.
- [30] Dubarry, C. and Le Corff, S. (2011). Fast computation of smoothed additive functionals in general state-space models. *IEEE Statistical Signal Processing Workshop (SSP)*, pages 197–200.
- [31] Dubarry, C. and Le Corff, S. (2013). Non-asymptotic deviation inequalities for smoothed additive functionals in nonlinear state-space models. *Bernoulli*, 19(5B):2222–2249.
- [32] Fama, E.-F. and French, K.-R. (1988). Business cycles and the behavior of metals prices. *Journal of Finance*, 43(5):1075–1093.

BIBLIOGRAPHY

- [33] Fearnhead, P. and Clifford, P. (2003). On-line inference for hidden Markov models via particle filters. *J.R. Statist. Soc. B*, 65:887–899.
- [34] Fearnhead, P., Wyncoll, D., and Tawn, J. (2010). A sequential smoothing algorithm with linear computational cost. *Biometrika*, 97(2):447–464.
- [35] Fong, W., Godsill, S.-J., Doucet, A., and West, M. (2002). Monte Carlo smoothing with application to audio signal enhancement. *IEEE Transactions on Signal Processing*, 50(2):438–449.
- [36] Fong, W. and See, K. (2002). A markov switching model of the conditional volatility of crude oil futures prices. *Energy Economics*, 24(1):71–95.
- [37] Fort, G. and Moulines, E. (2003). Convergence of the Monte Carlo expectation maximization for curved exponential families. *Ann. Statist.*, 31(4):1220–1259.
- [38] Gabillon, J. (1991). The Term Structures of Oil Futures Prices. *Management Science*, 46:893–911.
- [39] Gibson, R. and Schwartz, E. (1990). Stochastic convenience yield and the pricing of oil contingent claims. *Journal of finance*, 45(3):959–976.
- [40] Godsill, S. J., Doucet, A., and West, M. (2004). Monte Carlo smoothing for non-linear time series. *J. Am. Statist. Assoc.*, 50:438–449.
- [41] Gordon, N., Salmond, D., and Smith, A. (1993). Novel approach to nonlinear/non-Gaussian bayesian state estimation. *IEE Proc. F, Radar Signal Process*, 140:107–113.
- [42] Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42(1):27–62.
- [43] Hamilton, J. (1988). Rational-expectations econometric analysis of changes in regime: An investigation of the term structure of interest rates. *Journal of Economic Dynamics and Control*, 12(2-3):385–423.

-
- [44] Hamilton, J. D. and Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, 64(1-2):307–333.
- [45] Hansen, N. (2006). The CMA evolution strategy: a comparing review. In *Towards a new evolutionary computation*, pages 75–102. Springer.
- [46] Hansen, N. and Kern, S. (2004). Evaluating the CMA Evolution Strategy on Multimodal Test Functions. *Eighth International Conference on Parallel Problem Solving from Nature*, 72:337–354.
- [47] Hansen, N. and Ostermeier, A. (2001). Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195.
- [48] Hilliard, J.-E. and Reis, J. (1998). Valuation of Commodity Futures and Options Under Stochastic Convenience Yields, Interest Rates, and Jump Diffusions in the Spot. *The journal of financial and quantitative analysis*, 33(1):61–86.
- [49] Hürzeler, M. and Künsch, H. R. (1998). Monte Carlo approximations for general state-space models. *J. Comput. Graph. Statist.*, 7:175–193.
- [50] Jacob, P., Murray, L., and Rubenthaler, S. (2013). Path storage in the particle filter. *Statistics and Computing*, pages 1–10.
- [51] Kaldor, N. (1939). Speculation and economic stability. *Review of Economic Studies*, 7(1):1–27.
- [52] Kalman, R. (1960). A new approach to linear filtering and prediction problems. *J. Basic Engng*, 82:35–45.
- [53] Kantas, N., Doucet, A., Singh, S., Maciejowski, J., and Chopin, N. (2015). On particle methods for parameter estimation in state-space models. *Statist. Sci.*, 30(3):328–351.
- [54] Keynes, J.-M. (1930). *A Treatise on Money: The Pure Theory of Money and The Applied Theory of Money*. Paperback.

BIBLIOGRAPHY

- [55] Kiesel, R., Schindlmayr, G., and Borger, R. (2009). A two-factor model for the electricity forward market. *Quantitative Finance*, 9:279–287.
- [56] Kim, C. (1994). Dynamic linear models with markov-switching. *Journal of Econometrics*, 60(1-2):1–22.
- [57] Kitagawa, G. (1987). Non-gaussian state-space modeling of nonstationary time series. *Journal of the American Statistical Association*, 82(400):1032–1041.
- [58] Kitagawa, G. (1994). The two-filter formula for smoothing and an implementation of the gaussian-sum smoother. *Annals of the Institute of Statistical Mathematics*, 46(4):605–623.
- [59] Kitagawa, G. (1996). Monte-Carlo filter and smoother for non-Gaussian non-linear state space models. *J. Comput. Graph. Statist.*, 1:1–25.
- [60] Kitagawa, G. and Sato, S. (2001). Monte carlo smoothing and self-organizing state-space model. In Doucet, A., De Freitas, N., and Gordon, N., editors, *Sequential Monte Carlo methods in Practice*. Springer.
- [61] Künsch, H. R. (2005). Recursive Monte Carlo filters: Algorithms and theoretical analysis. *Ann. Statist.*, 33(5):1983–2021.
- [62] Le Corff, S. and Fort, G. (2013a). Convergence of a particle-based approximation of the block online Expectation Maximization algorithm. *ACM Transactions on Modeling and Computer Simulation*, 23(1):2.
- [63] Le Corff, S. and Fort, G. (2013b). Online Expectation Maximization based algorithms for inference in hidden Markov models. *Electronic Journal of Statistics*, 7:763–792.
- [64] Lee, A. and Whiteley, N. (2016). Variance estimation and allocation in the particle filter. *arXiv:1509.00394*.
- [65] Lindsten, F., Bunch, P., Godsill, S., and Schon, T. (2013). Rao-Blackwellized particle smoothers for mixed linear/nonlinear state-space models. *Proceedings*

-
- of the 38th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).*
- [66] Lindsten, F., Bunch, P., Sarkka, S., Schon, T., and Godsill, S. (2015a). Rao-Blackwellized particle smoothers for conditionally linear Gaussian models. *IEEE Journal of Selected Topics in Signal Processing*, 10(2):353–365.
- [67] Lindsten, F., Bunch, P., Sarkka, S., Schon, T., and Godsill, S. (2015b). Rao-Blackwellized particle smoothers for conditionally linear Gaussian models. *ArXiv:1505.06357*.
- [68] Liu, P. and Kang, T. (2011). The stochastic behavior of commodity prices with heteroskedasticity in the convenience yield. *J. Empir. Financ.*, 18(2):211–224.
- [69] Lucia, J.-J. and Schwartz, E.-S. (2001). Electricity Prices and Power Derivatives: Evidence from the Nordic Power Exchange. *Review of Derivatives Research*, 5(1):5–50.
- [70] Manoliu, M. and Tompaidis, S. (2002). Energy Futures Prices: Term Structure Models with Kalman Filter Estimation. *Applied Mathematical Finance*, 49:21–43.
- [71] Miltersen, K.-R. and Schwartz, E.-S. (1998). Pricing of options on commodity futures with stochastic term structures of convenience yields and interest rates. *Journal of financial and quantitative analysis*, 33(1):33–59.
- [72] Mirantes, A., Poblacin, J., and Serna, G. (2013). The stochastic seasonal behavior of energy commodity convenience yields. *Energy Economics*, 40:155–166.
- [73] Mongillo, G. and Deneve, S. (2008). Online learning with hidden Markov models. *Neural Computation*, 20(7):1706–1716.
- [74] Mount, Y., Ning, T. D., and Cai, X. (2006). Predicting price spikes in electricity markets using a regime-switching model with time-varying parameters. *Energy Economics*, 28(1):62–80.

BIBLIOGRAPHY

- [75] Naifar, N. and Dohaiman, M. S. A. (2013). Nonlinear analysis among crude oil prices, stock markets' return and macroeconomic variables. *International Review of Economics & Finance*, 27:416–431.
- [76] Olsson, J., Cappe, O., Douc, R., and Moulines, E. (2008). Sequential monte carlo smoothing with application to parameter estimation in nonlinear state space models. *Bernoulli*, 14(1):155–179.
- [77] Olsson, J. and Strojby, J. (2011). Particle-based likelihood inference in partially observed diffusion processes using generalised Poisson estimators. *Electronic Journal of Statistics*, 5:1090–1122.
- [78] Olsson, J. and Westerborn, J. (2015). Efficient particle-based online smoothing in general hidden Markov models: the PaRIS algorithm. *ArXiv:1412.7550*.
- [79] Persing, A. and Jasra, A. (2013). Likelihood computation for hidden markov models via generalized two-filter smoothing. *Statistics & Probability Letters*, 83(5):1433–1442.
- [80] Pitt, M. K. and Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters. *J. Am. Statist. Assoc.*, 94(446):590–599.
- [81] Poyiadjis, G., Doucet, A., and Singh, S. (2011). Particle approximations of the score and observed information matrix in state space models with application to parameter estimation. *Biometrika*, 98:65–80.
- [82] Rauch, A., Striebel, C., and Tung, F. (1965). Maximum likelihood estimates of linear dynamic systems. *American Institute of Aeronautics and Astronautics journal*, 3(8):1445–1450.
- [83] Richter, M.-C. and Sorensen, C. (2002). Stochastic Volatility and Seasonality in Commodity Futures and Options: The Case of Soybeans. *Berlin Meetings Presented Paper*.
- [84] Sarkka, S. (2013). *Bayesian filtering and smoothing*. Cambridge University Press.

- [85] Sarkka, S., Bunch, P., and Godsill, S. (2012). A backward-simulation based Rao-Blackwellized particle smoother for conditionally linear Gaussian models. *Proceedings of the 16th IFAC Symposium on System Identification (SYSID)*.
- [86] Schon, T., Gustafsson, F., and Nordlund, P.-J. (2005). Marginalized particle filters for mixed linear/nonlinear state-space models. *IEEE Transactions on Signal Processing*, 53(7):2279–2289.
- [87] Schwartz, E.-S. (1997). The Stochastic Behavior of Commodity Prices: Implications for Valuation and Hedging. *Journal of finance*, 52(3):923–973.
- [88] Schwartz, E.-S. and Smith, J.-E. (2000). Short-Term Variations and Long-Term Dynamics in Commodity Prices. *Management Science*, 46:893–911.
- [89] Sorensen, C. (2002). Modeling seasonality in agricultural commodity futures. *Journal of Futures Markets*, 22(5):393–426.
- [90] Trolle, A.-B. and Schwartz, E.-S. (2009). Unspanned Stochastic Volatility and the Pricing of Commodity Derivatives. *Review of Financial Studies*, 22:4423–4446.
- [91] Vo, M. (2009). Regime-switching stochastic volatility: Evidence from the crude oil market. *Energy Economics*, 31(5):779–788.
- [92] Working, H. (1949). The theory of price of storage. *The American Economic Review*, 39(6):1254–1262.