Université de Montréal

Essays on numerically efficient inference in nonlinear and non-Gaussian state space models, and commodity market analysis

par
Gbowan Barnabé Djegnéné

Département de sciences économiques
Faculté des arts et des sciences

Thèse présentée à la Faculté des arts et des sciences
en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)
en sciences économiques

Juin, 2013

Université de Montréal
Faculté des arts et des sciences

Cette thèse intitulée:

Essays on numerically efficient inference in nonlinear and non-Gaussian state space models, and commodity market analysis

présentée par:

Gbowan Barnabé Djegnéné

a été évaluée par un jury composé des personnes suivantes:

Andriana Bellou,	président-rapporteur
William McCausland,	directeur de recherche
Gérard Gaudet,	codirecteur
Francisco Ruge-Murcia,	membre du jury
Joshua Chan,	examinateur externe
Michel Delfour,	représentant du doyen de la FAS

Thèse acceptée le: . . . . . .

RÉSUMÉ

Les deux premiers articles élaborent des procédures de simulation du vecteur d'état et d'estimation des paramètres dans des modèles espace-états non linéaires et non-Gaussiens. Nous proposons des spécifications des modèles espace-états qui offrent plus de flexibilité dans la modélisation des relations dynamiques avec variables latentes. Les procédures d'estimation des paramètres dans ces modèles sont une extension de la méthode HESSIAN de McCausland[2012]. Ainsi, elles utilisent une approximation de la densité à posteriori du vecteur d'état qui permet de : simuler directement de la loi à posteriori du vecteur d'état, de simuler en seul bloc le vecteur d'état et de le simuler conjointement avec le vecteur de paramètres, et de ne pas admettre l'introduction d'inconnues additionnelles. Ces propriétés permettent d'obtenir des simulateurs à posteriori avec une efficacité numérique relative très élevée. Les procédures d'estimation élaborées sont génériques. Elles ouvrent ainsi une voie pour une analyse des modèles espace-états non linéaires et non-Gaussiens sans une grande contribution du modélisateur.

Le troisième article est une contribution dans l'analyse des marchés agricoles. Les firmes privées coexistent avec les coopératives de fermiers dans les marchés agricoles en Afrique subsaharienne. Les firmes privées accaparent les plus grandes parts de marché, alors que certains modèles théoriques prédisent leur disparition une fois confrontées aux coopératives agricoles. Par ailleurs, certaines observations et études empiriques lient la forte incidence d'une coopérative dans une région à la confiance interpersonnelle entre les personnes de cette région, et par conséquent la confiance de ces personnes envers les coopératives existantes. Nous proposons un modèle théorique qui cadre mieux avec ces observations empiriques. Un modèle où la réputation de la coopérative est un facteur déterminant de l'équilibre de marché dans la compétition sur le prix à la livraison entre celle-ci et une firme privée.

Mots clés : Espace-État, Non-linéaire, non-Gaussien, MCMC, Efficacité numérique, Volatilité stochastique, Durée stochastique, Coopérative, Réputation.

JEL Classification : C11, C15, C58, C63, Q13.

ABSTRACT

The first two articles build procedures to simulate vector of univariate states and estimate parameters in nonlinear and non Gaussian state space models. We propose state space specifications that offer more flexibility in modeling dynamic relationship with latent variables. Our procedures are extension of the HESSIAN method of McCausland[2012]. Thus, they use approximation of the posterior density of the vector of states that allow to : simulate directly from the state vector posterior distribution, to simulate the states vector in one bloc and jointly with the vector of parameters, and to not allow data augmentation. These properties allow to build posterior simulators with very high relative numerical efficiency. Generic, they open a new path in nonlinear and non Gaussian state space analysis with limited contribution of the modeler.

The third article is an essay in commodity market analysis. Private firms coexist with farmers' cooperatives in commodity markets in subsaharan african countries. The private firms have the biggest market share while some theoretical models predict they disappearance once confronted to farmers cooperatives. Elsewhere, some empirical studies and observations link cooperative incidence in a region with interpersonal trust, and thus to farmers trust toward cooperatives. We propose a model that sustain these empirical facts. A model where the cooperative reputation is a leading factor determining the market equilibrium of a price competition between a cooperative and a private firm.

Keywords : State-space, Nonlinear, Non-Gaussian, MCMC, Numerical Efficiency, Stochastic Volatility, Stochastic Duration, Cooperatives, Reputation.

JEL Classification : C1, C15, C58, C63, Q13

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

LISTE DES ANNEXES

MÈRE ! J'AI FINI, PRESQUE ...

À l'école primaire, la communauté villageoise organisait chaque fin d'année une cérémonie de collation des grades.

Au menu principal, la redoutable cérémonie d'appel des élèves par ordre de mérite devant l'assemblée.

Les parents couvraient alors de cadeaux leur progéniture selon leur rang.

À chacune de mes classes, j'étais souvent parmi les premiers appelés, pour ne pas dire le premier.

À la cérémonie de la quatrième année, sentant probablement qu'à ce rythme je finirai par la ruiner financièrement, ma mère me fit cette promesse :

> Fils, ton véritable cadeau, tu l'auras lorsque tu auras fini d'apprendre.

La quête de ce cadeau m'a transporté des plages ensoleillées de mon village natal, à l'adorable hiver québécois.

Si j'ai bonne mémoire, c'était un jeudi, juste deux mois après mon arrivée à Montréal.

J'étais sur le pas de ma porte, allant à la cérémonie de remise officielle de la Bourse de la Banque Laurentienne qui m'avait permis d'initier ce volet-ci de cette quête.

Le téléphone sonna. Au bout du fil, l'ainé de ma famille :

> 'Bonjour petit frère. Ton cadeau, tu ne l'auras plus !'

Mais au fait, ce cadeau je l'ai reçu dès l'instant où ma mère fit cette promesse.

Elle sema en moi cette quête inlassable de la connaissance, une quête presque infinie !

Merci **Neah Ephrasie Antoinette Blah**.

REMERCIEMENTS

Je n'aurai pu mener à terme ces travaux de recherche sans le concours remarquable de diverses personnes, autant sur le plan académique que social. C'est ici le lieu de les remercier du plus profond du cœur.

Mes premiers remerciements vont à l'endroit de Monsieur Koffi N'Guessan, Directeur de l'École Nationale Supérieure de Statistique et d'Économie Appliquée (ENSEA), d'Abidjan, Côte d'Ivoire. C'est grâce à son concours que j'ai débuté dans l'enseignement supérieur et entrepris mes premiers travaux de recherche. Je le remercie surtout pour sa passion à former des jeunes africains en économie et statistique.

Aux Professeurs William McCausland et Gérard Gaudet, je vous exprime ma gratitude pour avoir accepté de diriger mes travaux de recherche. Merci pour votre investissement personnel à la conclusion de ces travaux. Certaines étapes n'auraient pu être franchies sans vos immenses connaissances. Avec vous, j'ai appris autant sur le plan scientifique que humain. Vous m'avez appris à transformer des idées en savoir. Merci surtout pour votre patience, car le chemin fut long et sinueux.

Je ne saurai oublier les autres professeurs du Département de Sciences Économique pour leurs enseignements de qualité et le personnel administratif pour leur impressionnant soutien tout au long de ces années de recherche.

Je remercie le Département de Sciences Économiques, la Banque Laurentienne ainsi que le Centre Interuniversitaire de Recherche en Économie Quantitative (CIREQ) pour leurs concours financiers qui m'ont permis de mener ces travaux de recherche à terme et de participer à diverses conférences enrichissantes.

On dit souvent qu'une thèse est un moment de solitude. Non, ce n'est pas vrai. Nous bénéficions dans ces moments de l'amour de notre famille et de nos ami(e)s qui nous permet de tenir dans les moments de doute. C'est pourquoi, je voudrais dans ces lignes remercier du plus profond de mon coeur, mon épouse Zaguehi Edwige Tagro, ainsi que notre fils Babo Penyel Djegnene, pour leur patience et leur amour. Je remercie mon frère aîné Babo Charles, pour sa contribution immense à ma formation. Je ne saurai oublier dans ce volet, mes amis Paul Angoua et Isabelle Dorcas Akaffou. Ils étaient là dans les

moments les plus difficiles pour m'aider à remonter la pente.

À toutes et tous, merci. Que Dieu vous bénisse et vous garde.

INTRODUCTION GÉNÉRALE

Ma thèse est composée de trois articles. Les deux premiers articles proposent des procédures de simulation du vecteur d'état et d'estimation des paramètres dans les modèles espace-états nonlinéaires et non-Gaussiens. Ces modèles sont utiles pour représenter les interactions dynamiques avec variables latentes. Ils permettent notamment l'analyse de la volatilité des prix des actifs financiers ou l'analyse de la structure à terme des prix de commodité. Le troisième article propose un modèle théorique sur le rôle de la réputation d'une coopérative de fermiers dans l'issue de la compétition entre celle-ci et une firme privée dans un marché agricole. Cette structure de ma thèse est le reflet d'une ambition première formulée dans mon projet de candidature à l'admission au doctorat. Cette ambition consistait d'une part à analyser la dynamique des actifs financiers et les stratégies de gestion de risque associées, et d'autre part à déterminer un mode d'agrégation optimale des petits fermiers leur permettant l'usage des instruments financiers pour gérer leur exposition au risque.

Le premier article propose une procédure de simulation du vecteur d'état et d'estimation des paramètres dans un modèle espace-état avec effet de type levier. Dans ces modèles, le vecteur d'état est Gaussien et les vecteurs d'observation sont non linéaires dans les variables d'état avec des distributions non Gaussiennes. Autre caractéristique importante, sachant l'état courant, le vecteur d'observation courant et l'innovation courante de l'état sont dépendants. Le modèle de volatilité stochastique avec effet levier de Harvey et Shephard [1996] est le plus connu ayant de telles caractériques, d'où la dénomination.

Élaborer des procédures d'estimation des paramètres dans cette classe de modèle est très difficile. Le calcul de la vraisemblance exige une intégration du vecteur d'état sur un espace de très grande dimension. Un calcul analytique de la vraisemblance est donc impossible. La simulation du vecteur d'état s'avère alors nécessaire pour approximer les valeurs de la vraisemblance ou faire de l'estimation Bayésienne à posteriori.

Pour la classe de modèle considérée, il est impossible de simuler le vecteur d'état directement à partir de sa densité conditionnelle sachant les observations et les paramètres.

Nous appellerons cette densité, la densité cible. Pour palier cela, il faut utiliser une densité auxiliaire qui servira de densité proposée dans une méthode Markov chain Monte Carlo, ou de densité préférentielle dans un échantillonnage préférentiel. Nous construisons dans cet article une telle densité auxiliaire comme approximation de la densité cible.

La méthode utilisée pour construire la densité auxiliaire est similaire à la méthode HESSIAN de McCausland[2012], élaborée pour des modèles sans effet de type levier. McCausland[2012] utilise les dérivées, par rapport aux variables d'états, de la log densité conditionnelle de l'observation courante sachant l'état courant pour construire l'approximation de la densité cible. Nous utilisons les dérivées partielles, par rapport aux variables d'état, de la log densité conditionnelle de l'observation courante sachant l'état courant et l'innovation courante de l'état pour construire l'approximation de la densité cible.

Notre méthode hérite de la méthode HESSIAN, certaines propriétés qui la démarquent des autres propositions existantes dans la littérature pour estimer les paramètres dans les modèles espace-état avec effet de type levier. Elle simule le vecteur d'état directement de la loi de la densité cible. Nous ne rendons pas le modèle linéaire en vue de simuler d'une distribution approximative, et pondérer l'échantillon par la suite pour obtenir des estimateurs de la loi cible. Le vecteur d'état est simulé en un seul bloc dans les méthodes MCMC. De plus, le vecteur d'état et le vecteur des paramètres sont simulés conjointement en un seul bloc. Ces premières propriétés permettent d'élaborer une procédure de simulation à posteriori du vecteur d'état et des paramètres avec une efficacité numérique relative très élevée.

Dernière propriété, et pas la moindre, la procédure pour construire la densité auxiliaire est générique. Sa mise en œuvre pour un modèle particulier, nécessite simplement de calculer les dérivées partielles de la log densité conditionnelle de l'observation courante sachant l'état courant et l'innovation courante de l'état. Nous obtenons les valeurs exactes des dérivées partielles sans nécessairement calculer leurs expressions analytiques. Nous pouvons calculer les dérivées partielles pour des fonctions élémentaires et par la suite les combiner avec des routines appliquant la formule de Faa-Di-Bruno pour le calcul des dérivées partielles de fonctions composées. Même si nous n'utilisons pas des dérivées numériques, il est possible d'y avoir recours. Les dérivées numériques dimi-

nuent la qualité de l'approximation sans compromettre la convergence des simulations à posteriori.

Nous illustrons notre méthode avec le modèle de volatilité stochastique avec effet de levier. Nous permettons aux innovations de l'observation d'être Gaussienne ou t de Student. Avec l'échantillonnage préférentiel,notre procédure atteint une efficacité numérique relative de presque 100% pour tous les paramètres. Ce qui est très élevé pour une méthode générique comparée aux méthodes modèles spécifiques de Omori et al[2007] et Omori et Nakajima[2009].

Le second article propose une méthode de simulation de l'état et d'estimation des paramètres dans un modèle espace-état nonlinéaire et non-Gaussien. À la différence du premier article, la transition entre états peut se faire de façon nonlinéaire et l'innovation de l'état peut être non Gaussienne. La variable d'observation conserve sa flexibilité antérieure dans sa liaison avec l'état et dans sa distribution. Nous maintenons toujours l'hypothèse de la dépendance conditionnelle entre l'observation courante et l'innovation courante de l'état. Cette formulation est beaucoup plus flexible que la plupart des spécifications existantes d'un modèle espace-état. Elle incorpore ainsi les modèles espace-états avec effet de type levier.

Nous généralisons la méthode HESSIAN décrite dans le premier article pour simuler l'état latent et estimer les paramètres du modèle. L'approximation de la densité cible est construite à partir des dérivées partielles, par rapport aux variables d'état, de la log densité jointe de l'observation courante et de l'innovation courante, sachant l'état courant. Il n'est pas requis de factoriser cette densité jointe comme le produit de la densité de l'innovation courante de l'état sachant l'état courant et de la densité de l'observation sachant l'état courant et l'innovation courante de l'état.

La procédure pour construire l'approximation de la densité cible est une fois de plus générique. Les propriétés de la méthode HESSIAN de base sont préservées de sorte que nous élaborons des simulateurs à posteriori avec une efficacité numérique relative très élevée.

Nous illustrons la méthode HESSIAN Généralisée avec le modèle de durée conditionnelle stochastique introduit par Feng et al [2004]. Nous atteignons dans la simulation

des paramètres de ce modèle des efficacités numériques autour de 100%, alors que la méthode décrite dans Strickland et al.[2006], pour des modèles similaires, a une efficacité numérique qui n'excède pas 5%. Pour ce modèle, nous proposons aussi une méthode de traitement des transactions simultanées qui permet une meilleure approximation de la densité prédictive générée par les paramètres estimés pour les faibles valeurs de durée.

Notre troisième article s'intéresse à l'organisation des marchés agricoles. Tel qu'indiqué plus haut, l'intérêt porté à ces marchés a découlé d'un projet de transfert de connaissances auprès des petits fermiers pour l'usage des instruments financiers pour gérer leur exposition au risque. La libéralisation des marchés agricoles en Afrique subsaharienne a généré d'une part l'introduction des firmes privées sur le marché de la collecte locale avec situation d'oligopsone, et d'autre part un transfert de la volatilité des prix internationaux aux petits fermiers. Juguler ces deux effets néfastes de la libéralisation impliquait une organisation des fermiers en des coopératives agricoles viables. Mais, après plus d'une décennie de promotion, les coopératives agricoles ont réalisé peu de progrès. Pourtant, le modèle théorique de Albeak et Schultz(1998) et dans une certaine mesure Sexton(1990) prédisent la disparition d'une firme privée compétissant avec une coopérative de fermiers dans un marché agricole.

L'observation du comportement des animateurs des coopératives locales a suggéré que la confiance suscitée par les coopératives, autrement dit leur reputation pourrait expliquer leurs difficultés de croissance. Par ailleurs, nous pouvons récenser dans la littérature des contributions empiriques ou théoriques liant densité de coopératives et confiance. Ainsi, James et Sykuta [2004], sur la base d'une enquête auprès des producteurs de maïs et soja du Missouri, relèvent que la confiance et la perception de l'honnêteté sont les principaux facteurs expliquant la décision d'un fermier d'appartenir à une coopérative que de vendre sa production à une firme privée. Miguel, Gertler, et Levine [2005] utilisent la densité des coopératives dans une région comme proxy de la confiance régnant dans cette région.

Nous proposons un modèle théorique où la réputation d'une coopérative est un facteur déterminant de l'équilibre d'une compétition prix entre celle-ci et une firme privée dans un marché agricole. Nous empruntons à Sexton[1990] la structure économique du

modèle. Nous y remplaçons la dimension spatiale par une dimension de confiance entre les fermiers et la coopérative. L'échéancier de paiement proposé par la coopérative au fermier fait de la confiance envers la coopérative un élément important de sa décision de la patronner. En effet, le paiement de la production du fermier par la coopérative se fait au moins en deux étapes. Un prix à la livraison lorsque la coopérative prend possession de la matière première. Un prix résiduel, ou dividende, versé en fin d'exercice. Si le prix à la livraison est certain, le 'dividende' l'est moins. Ce dernier peut être détourné ou l'incompétence de la coopérative peut ne pas générer de 'dividende'. La relation de confiance entre le producteur et la coopérative, ou la réputation de cette dernière, joue alors un rôle prépondérant dans la décision du producteur d'appartenir à la coopérative. Toutes choses égales par ailleurs, une bonne réputation de la coopération accroit la part de marché de celle-ci tout en contraignant la firme privée à pratiquer un prix à la livraison plus élevé. Une telle prédiction cadre mieux avec l'observation empirique que celles des modèles de Sexton [1990], Albaek et Schultz[1998] et Karantininis et Zago[2001] sur le même sujet.

CHAPITRE 1

THE HESSIAN METHOD FOR MODELS WITH LEVERAGE-LIKE EFFECTS

Abstract

We propose a new method for simulation smoothing in state space models with univariate states and leverage-like effects. Given a vector θ of parameters, the state sequence $\alpha = (\alpha_1, \ldots, \alpha_n)^\top$ is Gaussian and the sequence $y = (y_1^\top, \ldots, y_n^\top)^\top$ of observed vectors may be conditionally non-Gaussian. By leverage-like effect, we mean conditional dependence between the observation y_t and the contemporaneous *innovation* of the state equation, not just the contemporaneous state α_t. We use this term since stochastic volatility models with the leverage effect are a leading example.

Our method is an extension of the HESSIAN method described in McCausland [16], which only works for models without leverage-like effects, models in which the density $f(y_t|\theta, \alpha)$ depends only on θ and α_t. Like that method, ours is based on a close approximation $g(\alpha|\theta, y)$ of the conditional density $f(\alpha|\theta, y)$. One can use $g(\alpha|\theta, y)$ for importance sampling or Markov chain Monte Carlo (MCMC). With a suitable approximation $g(\theta|y)$ of $f(\theta|y)$, we can use $g(\theta, \alpha|y) = g(\theta|y)g(\alpha|\theta, y)$ as an importance or proposal density for the joint posterior distribution of parameters and states. Applications include the approximation of likelihood function values and the marginal likelihood, and Bayesian posterior simulation. We construct the approximation $g(\alpha|\theta, y)$ for Gaussian and Student's t stochastic volatility models with leverage. For both models, we make a joint proposal of the state and parameter vectors. Unlike Omori et al. [20] and Nakajima and Omori [18], we do not augment the data by adding mixture indicators or heavy tail scaling factors. For the numerical estimation of posterior means of parameters, our generic procedure is more numerically efficient than the model specific procedures of those papers — using randomised pseudo-Monte Carlo importance sampling, we obtain relative numerical efficiencies close to 100% for all parameters and both Gaussian and Student's t stochastic volatility models. For many parameters, this is considerably higher than the numerical efficiency of the method of Omori et al. [20], for a model with Gaussian in-

novations, and the method of Nakajima and Omori [18], for a model with Student's t innovations.

Keywords : State space models, Nonlinear, Non-Gaussian, MCMC, Numerical Efficiency, Stochastic Volatility

JEL Classification : C11, C15, C58, C63.

1.1 Introduction

State space models govern the interaction of observable data $y = (y_1^\top, \ldots, y_t^\top, \ldots, y_n^\top)^\top$ and latent states $\alpha = (\alpha_1, \ldots, \alpha_t, \ldots, \alpha_n)^\top$, given a vector θ of parameters. They are very useful in capturing dynamic relationships, especially where there are changing, but latent, economic conditions : the states may be unobserved state variables in macroeconomic models, log volatility in asset markets or time varying model parameters.

Simulation smoothing methods have proven useful for approximating likelihood function values and Bayesian posterior simulation. They involve simulating the conditional distribution of states given data and parameters. We will call this distribution the *target* distribution. Simulation typically entails importance sampling or Markov chain Monte Carlo (MCMC). We show examples of both in Section 1.4.

State space models with conditional dependence between the observed value y_t and the contemporaneous *innovation* of the state equation, not just the contemporaneous state α_t, are of particular interest. The best known examples are stochastic volatility models with an asymmetric volatility effect known as the leverage effect. In the model introduced by Harvey and Shephard [12], the latent states α_t are log volatilities, given by

$$\alpha_1 = \bar{\alpha} + \frac{\sigma}{\sqrt{1-\phi^2}} u_0, \qquad \alpha_{t+1} = (1-\phi)\bar{\alpha} + \phi\alpha_t + \sigma u_t, \tag{1.1}$$

and observed returns y_t are given by

$$y_t = \exp(\alpha_t/2) v_t, \tag{1.2}$$

where the (u_t, v_t) are serially independent with

$$u_0 \sim N(0,1), \qquad \begin{bmatrix} u_t \\ v_t \end{bmatrix} \sim \text{i.i.d. N} \left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right), \tag{1.3}$$

and $(\sigma, \phi, \rho, \bar{\alpha})$ is a vector of parameters. If $\rho = 0$, y_t and the contemporaneous innovation σu_t are conditionally independent given α_t. When $\rho \neq 0$, they are conditionally

dependent and we call this conditional independence a leverage effect.

Others have extended this model. Jacquier et al. [13] and Omori et al. [20] consider inference in stochastic volatility models with leverage and heavy-tailed conditional return distributions. This and other empirical work has shown convincingly that stochastic volatility models with leverage are more realistic descriptions of stock returns than models without.

Leverage-like effects may be useful in other models as well. There is little reason beyond computational convenience to rule them out. Feng et al. [6] show that conditional dependence is more realistic in stochastic conditional duration models.

Designing inferential methods for such models has proven difficult, however, and methods with high numerical efficiency have been model specific. Nine years passed between Kim et al. [15], introducing the auxiliary mixture model approach for stochastic volatility models without leverage, and Omori et al. [20], extending it to models with leverage.

We extend the HESSIAN method of McCausland [16], which does simulation smoothing for models without leverage-like effects. That method used multiple derivatives of $\log f(y_t|\theta, \alpha_t)$ with respect to α_t to construct a close approximation to the target distribution. In models with leverage-like effects, the conditional distribution of y_t given α depends not only on α_t but also α_{t+1}. To obtain a similar standard of approximation that McCausland [16] does, we need multiple partial derivatives of $\log f(y_t|\theta, \alpha_t, \alpha_{t+1})$ with respect to α_t and α_{t+1}. Using these derivatives to construct an approximation of the target density requires more effort, largely because when there are leverage-like effects, all non-zero elements of the Hessian of the log target density depend on α, not just the diagonal elements.

Our method inherits the following features of the original method :

1. It involves direct simulation of states from their posterior distribution using a proposal or importance distribution approximating the target distribution. This is unlike auxiliary mixture model approaches, in which a model is first transformed into a linear model, and then any non-Gaussian distributions in the transformed model are approximated by finite Gaussian mixtures. Kim et al. [15], Chib et al.

[3], Omori et al. [20] use this auxiliary mixture model approach for stochastic volatility models ; Stroud et al. [26], Frühwirth-Schnatter and Wagner [8] and Frühwirth-Schnatter et al. [9] use it for other non-linear non-Gaussian state space models. Using the direct approach, we avoid model-specific transformations, data augmentation, and the need to weight or apply additional accept-reject steps to correct for approximation error.

2. It involves drawing the entire state sequence as a single MCMC block. This leads to efficiency improvements when there is posterior serial dependence. While drawing the entire state sequence using a multivariate Gaussian proposal distribution is impractical, we make it possible by constructing a much closer approximation of the target distribution. Many articles have used multivariate Gaussian proposal distributions to update the state vector, but usually only for about 10–50 observations at a time, not the entire sample. These include Shephard and Pitt [24], Watanabe and Omori [27], Strickland et al. [25], Jungbacker and Koopman [14] and Omori and Watanabe [19]. The Efficient Importance Sampling (EIS) method of Richard and Zhang [21] features draws of the entire state sequence as a block, but since their approximate target distribution is constructed using the random numbers used to draw from it, EIS estimators of likelihood function values do not have the simulation consistency or lack of simulation bias that true importance sampling estimators do. See the discussion in McCausland [16] for more details.
3. Since the approximation is so close, we can draw parameters and states together as a single block. We do this using a joint proposal distribution combining our approximation of the conditional posterior distribution of states given parameters with an approximation of the marginal posterior distribution of parameters. Drawing states and parameters in a single block leads to further efficiency improvements because of posterior dependence between states and parameters. In this way, we achieve numerical efficiencies comparable to model-specific auxiliary mixture model approaches, which also often feature joint draws of parameter and states. The examples of Section 1.4 suggest that our method is even more efficient than these approaches, partly because we avoid data augmentation and the need to cor-

rect for approximation error. Being able to draw all parameters and states jointly in an untransformed model also opens up new opportunities — it allows for importance sampling, variance reduction using randomised pseudo Monte Carlo, and very efficient approximations of the marginal likelihood, as we see in Section 1.4.

4. We construct our approximation of the target distribution in a generic way. The only model-specific computation is the evaluation of derivatives of the log measurement density. Existing, well tested, and publicly available generic code uses the routines for computing model-specific derivatives in order to do simulation smoothing for that model. Exact evaluation of derivatives does not require finding analytic expressions — we can use generic routines to combine derivative values according to Leibniz' rule for multiple derivatives of products and Faà di Bruno's rule for multiple derivatives of composite functions. Although we do not do so here, we could also resort to numerical derivatives — there would a cost in numerical efficiency, but simulation consistency would not be compromised. The Student's t distribution and other scale mixtures of normals are often used in stochastic volatility models, partly because they work well in auxiliary mixture model approaches using data augmentation for the mixing random variables. A generic approach allows for other, possibly skewed, measurement distributions.

5. It is based on operations using the sparse Hessian matrix of the log target density, rather than on the Kalman filter. Articles using the former approach include Rue [22], for linear Gaussian Markov random fields, Chan and Jeliazkov [2] and McCausland et al. [17], for linear Gaussian state space models, and Rue et al. [23] for non-linear non-Gaussian Markov random fields. The Integrated Nested Laplace Approximation (INLA) method described in the last article has spawned a large applied literature. Articles using the Kalman filter include Carter and Kohn [1], Frühwirth-Schnatter [7], de Jong and Shephard [4] and Durbin and Koopman [5] for linear Gaussian state space models. Auxiliary mixture model methods for non-linear or non-Gaussian models tend to use the Kalman filter, but this is not an essential feature of auxiliary mixture model methods.

We will now be more precise about the class of state space models we consider. The state and measurement equations are

$$\begin{aligned} \alpha_1 = d_0 + u_0, \qquad \alpha_{t+1} = d_t + \phi_t \alpha_t + \omega_t^{-1/2} u_t, \\ f(y|\alpha) = \left[\prod_{t=1}^{n-1} f(y_t|\alpha_t, \alpha_{t+1}) \right] f(y_n|\alpha_n), \end{aligned} \tag{1.4}$$

where $\alpha = (\alpha_1, \ldots, \alpha_n)$ is a vector of univariate latent states α_t, the u_t are independent Gaussian random variables with mean 0 and unit variance, the y_t are observable random vectors, and the $f(y_t|\alpha_t, \alpha_{t+1})$ are measurement density or mass functions. We do not require them to be Gaussian, linear or univariate. We say that models of this form exhibit a leverage-like effect whenever $f(y_t|\alpha_t, \alpha_{t+1})$ depends on α_{t+1}. This will be the case when the observable vector y_t and the contemporaneous state innovation $u_t = \alpha_{t+1} - d_t - \phi_t \alpha_t$ are conditionally dependent given the contemporaneous state α_t.

Throughout most of the paper, we condition on d_t, ϕ_t, ω_t and any other parameters on which the $f(y_t|\alpha_t, \alpha_{t+1})$ might depend, and suppress notation for this conditioning. In Section 1.4, where we consider joint inference for parameters and states, we are explicit about this conditioning.

It is easy to see that the model in equations (1.1), (1.2) and (1.3) is of the form given by (1.4). We use (1.1) to write

$$u_t = [\alpha_{t+1} - (1 - \phi)\bar{\alpha} - \phi \alpha_t]/\sigma,$$

then use the standard formula for conditional Gaussian distributions to obtain

$$y_t|\alpha \sim \mathrm{N}\left((\rho/\sigma) \exp(\alpha_t/2)(\alpha_{t+1} - (1 - \phi)\bar{\alpha} - \phi \alpha_t), (1 - \rho^2) \exp(\alpha_t) \right). \tag{1.5}$$

In Section 1.2 we describe our approximation $g(\alpha|y)$ of the target density $f(\alpha|y)$. We show how to evaluate it and how to draw from the distribution with density $g(\alpha|y)$. In Section 1.3 we apply tests of program correctness to the code we use to compute $g(\alpha|y)$ and draw from the approximate distribution. These tests are similar to those des-

cribed in Geweke [11]. Section 1.4 illustrates our methods using stochastic volatility models with leverage, with Gaussian and Student's t measurement innovations. Section 1.5 concludes.

1.2 An approximation of the target density

In this section we define our approximation $g(\alpha|y)$ of the target density $f(\alpha|y)$. We do not provide a closed form expression for $g(\alpha|y)$, but instead show how to evaluate and sample from $g(\alpha|y)$ using $O(n)$ operations. The density $g(\alpha|y)$ is proper and fully normalized.

Our approximation is not model specific. We construct $g(\alpha|y)$ for a particular state space model using a suitable description of the model, consisting of the following quantities and computational routines.

We specify the state dynamics by providing $\bar{\Omega}$ and $\bar{c}$, the precision and covector of the marginal distribution of α, the state sequence. This gives the distribution of α as $\alpha \sim N(\bar{\Omega}^{-1}\bar{c}, \bar{\Omega}^{-1})$. The precision, unlike the variance, is a tri-diagonal matrix, with $O(n)$ elements. Appendix I.1 describes how to compute $\bar{\Omega}$ and $\bar{c}$ in terms of the d_t, ϕ_t and ω_t.

We specify the measurement distributions by supplying routines to compute, for $t = 1, \ldots, n-1$, the functions

$$\psi_t(\alpha_t, \alpha_{t+1}) \doteq \log f(y_t|\alpha_t, \alpha_{t+1}), \quad \psi_n(\alpha_n) = \log f(y_n|\alpha_n), \tag{1.6}$$

and the partial derivatives

$$\psi_t^{(p,q)}(\alpha_t, \alpha_{t+1}) \doteq \frac{\partial^{p+q}\psi_t(\alpha_t, \alpha_{t+1})}{\partial\alpha_t^p \partial\alpha_{t+1}^q}, \quad \psi_n^{(p)}(\alpha_n) = \frac{\partial^p \psi(\alpha_n)}{\partial\alpha_n^p}, \tag{1.7}$$

for orders p and q up to certain values P and Q. For convenience, Table 1.I summarizes this and other important notation.

The routines to compute the $\psi_t(\alpha_t, \alpha_{t+1})$ and $\psi_n(\alpha_n)$ must give exact results, as they are used to evaluate $f(\alpha|y)$ up to a normalization factor. The partial derivatives, however,

may be numerical derivatives or other approximations. Approximation error may make $g(\alpha|y)$ a cruder approximation of $f(\alpha|y)$ and thus diminish the numerical precision of IS or MCMC. But we will still be able to evaluate and simulate $g(\alpha|y)$ without error, and so it does not compromise simulation consistency.

Like the target density, the approximation $g(\alpha|y)$ has the Markov property, allowing us to decompose it as

$$g(\alpha|y) = g(\alpha_n|y) \prod_{t=n-1}^{1} g(\alpha_t|\alpha_{t+1}, y). \tag{1.8}$$

Each factor is a proper fully normalized density function closely approximating the corresponding factor of $f(\alpha|y)$. Whether we need to evaluate $g(\alpha|y)$, simulate it or both, the decomposition allows us to do so sequentially, for t descending from n to 1.

Approximations rely on Taylor series expansions, some exact and some approximate, of various functions, including $b_{t|t+1}(\alpha_{t+1})$ and $\mu_{t|t+1}(\alpha_{t+1})$, the mode and mean of the conditional distribution of α_t given α_{t+1} and y. Some expansions are computed during a forward pass, around the mode $(a_1, \ldots, a_n)$ of the target distribution, a static point of expansion. So for example, we compute $B_{t|t+1}(\alpha_{t+1})$ and $M_{t|t+1}(\alpha_{t+1})$ as approximate Taylor series expansions of $b_{t|t+1}(\alpha_{t+1})$ and $\mu_{t|t+1}(\alpha_{t+1})$ around a_{t+1}.

During the backward pass, we compute approximate Taylor series expansions of $h_n(\alpha_n) \doteq \log f(\alpha_n|y)$ and $h_t(\alpha_t; \alpha_{t+1}) \doteq \log f(\alpha_t|\alpha_{t+1}, y)$, $t = n-1, \ldots, 1$, which we will treat as univariate functions of α_t with parameter α_{t+1}. Here, the point of expansion is a moving target, depending on α_{t+1}. The expansion is fifth order, allowing a much better than Gaussian (second order) approximation.

The densities $g(\alpha_t|\alpha_{t+1}, y)$ are members of the class of perturbed Gaussian distributions described in Appendix G of McCausland [16]. Parameters of the perturbed Gaussian distribution give a mode of the distribution and the second through fifth derivatives of $\log g(\alpha_t|\alpha_{t+1}, y)$ at that mode. Choosing parameters amounts to approximating $b_{t|t+1}(\alpha_{t+1})$, the mode of $f(\alpha_t|\alpha_{t+1}, y)$, and the second through fifth derivatives of $\log f(\alpha_t|\alpha_{t+1}, y)$ there.

In Appendix I.3.1, we derive this exact result for the first derivative of h_t :

$$\begin{aligned} h_t^{(1)}(\alpha_t;\alpha_{t+1}) = \bar{c}_t - \bar{\Omega}_{t-1,t}\mu_{t-1|t}(\alpha_t) - \bar{\Omega}_{t,t}\alpha_t - \bar{\Omega}_{t,t+1}\alpha_{t+1} \\ + x_{t-1|t}(\alpha_t) + \psi_t^{(1,0)}(\alpha_t,\alpha_{t+1}), \quad t=2,\ldots,n-1, \end{aligned} \tag{1.9}$$

where $\mu_{t-1|t}(\alpha_t) \doteq E[\alpha_{t-1}|\alpha_t, y]$ and $x_{t-1|t}(\alpha_t) \doteq E[\psi_{t-1}^{(0,1)}(\alpha_{t-1},\alpha_t)|\alpha_t, y]$. We also give analogous results for the cases $t=1$ and $t=n$.

We cannot evaluate $\mu_{t-1|t}(\alpha_t)$, $x_{t-1|t}(\alpha_t)$ or their derivatives exactly. Nor can we evaluate the mode $b_{t|t+1}(\alpha_{t+1})$ exactly. Instead, we provide polynomial approximations $M_{t-1|t}(\alpha_t)$, $B_{t|t+1}(\alpha_{t+1})$ and $X_{t-1|t}(\alpha_t)$ of $\mu_{t-1|t}(\alpha_t)$, $b_{t|t+1}(\alpha_{t+1})$ and $x_{t-1|t}(\alpha_t)$. We use these to approximate the value $b_{t|t+1}(\alpha_{t+1})$ and the derivatives $h_t^{(r)}(\alpha_t;\alpha_{t+1})$, $r = 1,\ldots,5$. $M_{t-1|t}(\alpha_t)$ and $X_{t-1|t}(\alpha_t)$ are approximate Taylor expansions of $\mu_{t-1|t}(\alpha_t)$ and $x_{t-1|t}(\alpha_t)$ around a_t. $B_{t|t+1}(\alpha_{t+1})$ is an approximate Taylor expansion of $b_{t|t+1}(\alpha_{t+1})$ around a_{t+1}.

We draw α, evaluate $g(\alpha|y)$, or both using the following steps. We first compute the mode $a=(a_1,\ldots,a_n)$ of the target distribution using the method described in Appendix B of McCausland [16]. In a forward pass we compute the coefficients of the polynomials $B_{t|t+1}(\alpha_{t+1})$, $M_{t-1|t}(\alpha_t)$, and $X_{t-1|t}(\alpha_t)$, for $t=1,\ldots,n-1$. Finally, we compute, for $t=n,\ldots,1$, $B_{t|t+1}(\alpha_{t+1})$ and $H_t^{(r)}(B_{t|t+1}(\alpha_{t+1});\alpha_{t+1})$, using these values as the parameters of the perturbed Gaussian distribution. With these values sets, we can draw α_t, evaluate $g(\alpha_t|\alpha_{t+1},y)$ or both. In the rest of this section, we describe these steps in more detail. Full detail is left to various appendices.

1.2.1 Precomputation

We first compute the precision $\bar{\Omega}$ and covector $\bar{c}$ of the Gaussian prior distribution of states as a function of d_t, ϕ_t and ω_t in (1.4). We then compute the mode a of the target distribution. This gives, as bi-products, several quantities used later. This includes the precision $\bar{\bar{\Omega}}$ and covector $\bar{\bar{c}}$ of a Gaussian approximation $\mathrm{N}(\bar{\bar{\Omega}}^{-1}\bar{\bar{c}}, \bar{\bar{\Omega}}^{-1})$ of the target density. It also gives the conditional variances $\Sigma_t \doteq \mathrm{Var}[\alpha_t|\alpha_{t+1}]$, $t=1,\ldots,n-1$, and $\Sigma_n \doteq \mathrm{Var}[\alpha_n]$ implied by this Gaussian approximation.

This precomputation is similar to that described in Appendix B of McCausland [16]. Little modification is required, and we give details in Appendix I.1.

1.2.2 A Forward Pass

In order to describe the forward pass, it will be helpful to introduce a sequence of multivariate Gaussian conditional distributions. We define, for $t = 1, \ldots, n-1$, $(a_{1|t+1}(\alpha_{t+1}), \ldots, a_{t|t+1}(\alpha_{t+1}))$ as the conditional mode of $(\alpha_1, \ldots, \alpha_t)$ given α_{t+1} and y, and $\bar{\bar{\Omega}}_{1:t|t+1}$ as the negative Hessian matrix of $\log f(\alpha_1, \ldots, \alpha_t | \alpha_{t+1}, y)$ with respect to $(\alpha_1, \ldots, \alpha_t)$, evaluated at $(a_{1|t+1}, \ldots, a_{t|t+1})$. Thus we can view the distribution $N((a_{1|t+1}, \ldots, a_{t|t+1}), \bar{\bar{\Omega}}^{-1}_{1:t|t+1})$ as an approximation of the conditional distribution of $(\alpha_1, \ldots, \alpha_t)$ given α_{t+1} and y. Result 2.1 of McCausland et al. [17] implies that if $\tilde{\alpha} \sim N((a_{1|t+1}, \ldots, a_{t|t+1}), \bar{\bar{\Omega}}^{-1}_{1:t|t+1})$, then $\tilde{\alpha}_t | \tilde{\alpha}_{t+1} \sim N(a_{t|t+1}, \Sigma_{t|t+1})$, where $\Sigma_{t|t+1}$ is the final value in the following forward recursion :

$$\Sigma_{1|t+1} \doteq \bar{\bar{\Omega}}^{-1}_{11}, \qquad \Sigma_{\tau|t+1} \doteq (\bar{\bar{\Omega}}_{\tau\tau} - \bar{\bar{\Omega}}^2_{\tau,\tau-1}\Sigma_{\tau-1|t+1})^{-1}, \quad \tau = 2, \ldots, t. \tag{1.10}$$

We also define, for $t = 1, \ldots, n-1$, $s_{t|t+1}(\alpha_{t+1}) \doteq \log \Sigma_{t|t+1}(\alpha_{t+1})$.

The forward pass consists of performing the following steps, for $t = 1, \ldots, n-1$:

1. Compute

$$\begin{aligned} a_t^{(r)} &\doteq \left. \frac{\partial^r a_{t|t+1}(\alpha_{t+1})}{\partial \alpha_{t+1}^r} \right|_{\alpha_{t+1}=a_{t+1}}, \quad r = 1, \ldots, R, \\ s_t^{(r)} &\doteq \left. \frac{\partial^r s_{t|t+1}(\alpha_{t+1})}{\partial \alpha_{t+1}^r} \right|_{\alpha_{t+1}=a_{t+1}}, \quad r = 1, \ldots, R-1. \end{aligned} \tag{1.11}$$

   The choice of R determines how closely we can approximate the functions $a_{t|t+1}(\alpha_{t+1})$ and $s_{t|t+1}(\alpha_{t+1})$ using Taylor expansions. For our empirical illustration, we use $R = 5$.

   Appendix I.2 gives details. Equation (I.8) gives $a_1^{(r)}$ and for $t > 1$, (I.18) gives $a_t^{(r)}$ as a function of $a_{t-1}^{(i)}$, $i = 1, \ldots, r$, and $a_t^{(i)}$, $i = 1, \ldots, r-1$. Equations (I.15), (I.20), (I.23), (I.26) and (I.28) give simplified expressions for $r = 1, \ldots, 5$ and

$t > 1$.

Equation (I.10) gives $s_1^{(r)}$ and equations (I.18) and (I.19) give $s_t^{(r)}$. Equations (I.22), (I.25), (I.27) and (I.29) give simplified expressions for $s_t^{(r)}$, $r = 1, \ldots, 4$ and $t = 2, \ldots, n-1$.

Appendix I.2 includes a proof that these computations are exact. The proof uses a first order necessary condition for $(a_{1|t+1}, \ldots, a_{t|t+1})$ to maximize $f(\alpha_1, \ldots, \alpha_t|\alpha_{t+1}, y)$, the identity $a_{t-1|t+1}(\alpha_{t+1}) = a_{t-1|t}(\alpha_{t+1}(\alpha_{t+1}))$ and the difference equation (1.10) defining $\Sigma_{t|t+1}(\alpha_{t+1})$.

2. Compute approximations B_t, $B_t^{(1)}$, $B_t^{(2)}$, $B_t^{(3)}$ and $B_t^{(4)}$ of the value and first four derivatives of $b_{t|t+1}(\alpha_{t+1})$ at a_{t+1}. Recall that $b_{t|t+1}(\alpha_{t+1})$ is the conditional mode of α_t given α_{t+1} and y. For $t = n$, we only compute an approximation B_n of the value b_n, the conditional mode of α_n given y. Appendix I.3.3 defines these approximations and shows how to compute them. Specifically, equation (I.42) defines $B_t^{(r)}$ as a function of the $a_t^{(i)}$ and $s_t^{(i)}$. The approximations are based on an approximation of $b_{t|t+1}(\alpha_{t+1}) - a_{t|t+1}(\alpha_{t+1})$ using a first order necessary condition for $b_{t|t+1}(\alpha_{t+1})$ to maximize $f(\alpha_t|\alpha_{t+1}, y)$.

3. Compute approximations M_t, $M_t^{(1)}$, $M_t^{(2)}$, $M_t^{(3)}$ and $M_t^{(4)}$ of the value and first four derivatives of $\mu_{t|t+1}(\alpha_{t+1})$ at a_{t+1}. Recall that $\mu_{t|t+1}(\alpha_{t+1})$ is the conditional mean of α_t given α_{t+1} and y. Appendix (I.3.4) defines these approximations. We compute $M_t^{(r)}$, from equation (I.50), as a function of the $B_t^{(i)}$, $a_t^{(i)}$ and $s_t^{(i)}$.

1.2.3 A Backward Pass

We use the backward pass to draw a random variate α^* from the distribution with density $g(\alpha|y)$ and evaluate $g(\alpha^*|y)$. One can also evaluate $g(\alpha|y)$ at an arbitrary value α^* without drawing.

To implement the backward pass, we use the following approximation of the derivative of $\log f(\alpha_t|\alpha_{t+1}, y)$, based on (1.9) and the approximations $M_{t-1|t}(\alpha_t)$ of $\mu_{t-1|t}(\alpha_t)$,

$X_{t-1|t}(\alpha_t)$ of $x_{t-1|t}(\alpha_t)$ and $\Psi_t^{(1,0)}(\alpha_t, \alpha_{t+1})$ of $\psi_t^{(1,0)}(\alpha_t, \alpha_{t+1})$.

$$\begin{aligned} H_t^{(1)}(\alpha_t; \alpha_{t+1}) \doteq \bar{c}_t - \bar{\Omega}_{t-1,t} M_{t-1|t}(\alpha_t) - \bar{\Omega}_{t,t}\alpha_t - \bar{\Omega}_{t,t+1}\alpha_{t+1} \\ + X_{t-1|t}(\alpha_t) + \Psi_t^{(1,0)}(\alpha_t, \alpha_{t+1}). \end{aligned} \tag{1.12}$$

We define the approximation $X_{t-1|t}(\alpha_t)$ and show how to compute it in Appendix I.5. $M_{t-1|t}(\alpha_t)$ is the polynomial

$$M_{t-1|t}(\alpha_t) = \sum_{r=0}^{4} \frac{M_{t-1}^{(r)}}{r!}(\alpha_t - a_t)^r, \tag{1.13}$$

We require routines to evaluate $\psi_t^{(p,q)}(\alpha_t, \alpha_{t+1})$ for several orders p, q, so in principle it is not necessary to approximate $\psi_t^{(p,q)}(\alpha_t, \alpha_{t+1})$. However, we find the computational costs high relative to the benefits. We already have $\psi_t^{(p,q)} = \psi_t^{(p,q)}(a_t, a_{t+1})$ from the forward pass, and we choose to approximate $\psi_t^{(p,q)}(\alpha_t, \alpha_{t+1})$ by

$$\Psi_t^{(p,q)}(\alpha_t, \alpha_{t+1}) \doteq \sum_{r=0}^{P-p} \sum_{s=0}^{Q-q} \psi_t^{(p+r,q+s)} \frac{(\alpha_t - a_t)^r}{r!} \frac{(\alpha_{t+1} - a_{t+1})^s}{s!}. \tag{1.14}$$

The backward pass consists of performing the following steps, for $t = n, \ldots, 1$.

1. Evaluate $B_{t|t+1}(\alpha_{t+1}^*)$, where $B_{t|t+1}(\alpha_{t+1})$ is the polynomial given by

$$B_{t|t+1}(\alpha_{t+1}) = \sum_{r=0}^{5} \frac{B_t^{(r)}}{r!}(\alpha_{t+1} - a_{t+1})^r \tag{1.15}$$

2. Compute $H_t^{(r)}(B_{t|t+1}(\alpha_{t+1}^*); \alpha_{t+1}^*), r = 2, \ldots, 5$, using (1.12).
3. Draw α_t^* and evaluate $g(\alpha_t|\alpha_{t+1}, y)$ at α_t^* and α_{t+1}^*. The density $g(\alpha_t|\alpha_{t+1}, y)$ is a member of the five-parameter perturbed Gaussian distribution described in Appendix G of McCausland [16]. The mode parameter is given by $b = B_{t|t+1}(\alpha_{t+1}^*)$, and the derivative parameters are given by $h_r = H_t^{(r)}(B_{t|t+1}(\alpha_{t+1}^*); \alpha_{t+1}^*)$, $r = 2, \ldots, 5$. These give the desired mode $B_{t|t+1}(\alpha_{t+1}^*)$ and desired derivatives of $\log g(\alpha_t|\alpha_{t+1}, y)$ at this mode.

1.3 Getting it right

In posterior simulation, analytical or coding errors can lead to reasonable but inaccurate results. Geweke [11] develops tests for the correctness of posterior simulations, based on two different methods for simulating the joint distribution of a model's observable and unobservable variables. Correctness tests take the form of tests of the hypothesis that the two samples come from the same distribution. Since the two methods have little in common, the tests have power against a wide array of conceptual and coding errors. We apply these ideas to build tests for the correctness of the independence Metropolis-Hastings update of the target distribution using the HESSIAN approximation $g(\alpha|y,\theta)$ as a proposal distribution.

We do this for the the asymmetric stochastic volatility model where the observation innovation is Student t (ASV-Student) described in the next section. We choose a fixed value of θ of the parameter vector. Then we generate a large sample from the conditional distribution of α and y given θ. We initialize with a draw $\alpha^{(0)}$ from the conditional distribution of α given θ, then draw $\{\alpha^{(m)}, y^{(m)}\}_{m=1}^{M}$ as follows. For $m = 1, \ldots, M$,

1. Draw $y^{(m)}$ from the conditional distribution of y given θ and α, with α set to $\alpha^{(m-1)}$.
2. Update from $\alpha^{(m-1)}$ to $\alpha^{(m)}$ using an independence Metropolis-Hastings step, with $g(\alpha|y,\theta)$ as a proposal distribution and $y = y^{(m)}$.

This is a Gibbs sampler for the conditional distribution of α and y given θ. The initial and stationary distributions of this chain are both equal to this distribution. By induction, so are the distributions of all the $(\alpha^{(m)}, y^{(m)})$. In particular, $\alpha^{(m)} \sim \mathrm{N}(\bar{\alpha}\imath, \bar{\Omega}^{-1})$ for all m, where $\imath$ is the n-vector with all elements equal to one. This implies that for all $m = 1, \ldots, M$ and $q \in (0,1)$, the following indicators are Bernoulli with probability parameter q :

$$I_{t,q}^{(m)} \doteq 1\left(\frac{\alpha_t^{(m)} - \bar{\alpha}}{\sigma/\sqrt{1-\phi^2}} \leq \Phi^{-1}(q)\right), t = 1, \ldots, n, \tag{1.16}$$

$$I_{t|t-1,q}^{(m)} \doteq 1\left(\frac{\alpha_t^{(m)} - (1-\phi)\bar{\alpha} - \phi\alpha_{t-1}^{(m)}}{\sigma} \leq \Phi^{-1}(q)\right), t = 1, \ldots, n, \tag{1.17}$$

where $\Phi(x)$ is the cumulative distribution function of the univariate standard Gaussian distribution.

We use sample means of the $I_{t,q}^{(m)}$ and $I_{t|t-1,q}^{(m)}$ to test the hypotheses that the corresponding population means are equal to q. We report results for the ASV-Student model. The parameter values are fixed to $\bar{\alpha} = -9.0$, $\phi = 0.97$, $\sigma = 0.15$, $\rho = -0.3$ and $\nu = 10.0$. We use a vector of length $n = 20$ and a sample size of $M = 10^7$. We use the $\mathbf{R}$ package coda to compute time series numerical standard errors and use Gaussian asymptotic approximations to construct symmetric 95% and 99% intervals. The 95% confidence interval does not include q in 7 cases out of 360 (1.94%). The 99% confidence interval does not include q in a single case (0.28%). The sample mean always lies well within the interval $[q - 0.001, q + 0.001]$. These results fail to cast doubt on the correctness of the implementation.

1.4 Empirical example

1.4.1 Models

We consider two different stochastic volatility models with asymmetric volatility. The first model, which we will call ASV-Gaussian, is the basic asymmetric volatility model given in equations (1.1), (1.2) and (1.3).

The second model, which we will call ASV-Student, replaces the observation equation in (1.2) with

$$y_t = \exp(\alpha_t/2)\frac{v_t}{\sqrt{\lambda_t/\nu}}, \tag{1.18}$$

where $\lambda_t \sim \chi^2(\nu)$ and the λ_t and (u_t, v_t) are mutually independent.

In order to allow us to draw parameters and states together in a single block, we will now integrate out λ_t to obtain the conditional distribution of y_t given α_t and α_{t+1}. This distribution is a scaled non-central Student's t. To see this, write $y_t = \exp(\alpha_t/2)\sqrt{1-\rho^2}X$, where

$$X \doteq \frac{u_t/\sqrt{1-\rho^2}}{\sqrt{\lambda_t/\nu}}.$$

Now condition on α_t and α_{t+1}. The numerator and denominator are independent ; the

numerator is Gaussian with mean

$$\mu \doteq \rho\sqrt{\frac{\omega}{1-\rho^2}}[\alpha_{t+1} - d_t - \phi_t\alpha_t]$$

and unit variance; and λ_t is chi-squared with ν degrees of freedom. Therefore X is non-central Student's t with non-centrality parameter μ and ν degrees of freedom. The density of X is given by

$$f_X(x;\nu,\mu) = \frac{\nu^{\nu/2}}{2^\nu}\frac{\Gamma(\nu+1)}{\Gamma(\nu/2)}\exp(-\mu^2/2)(\nu+x^2)^{-\nu/2} \times \left[\frac{\sqrt{2}\mu x}{\nu+x^2}\frac{M\left(\frac{\nu}{2}+1;\frac{3}{2};\frac{\mu^2x^2}{2(\nu+x^2)}\right)}{\Gamma(\frac{\nu+1}{2})} + \frac{1}{\sqrt{\nu+x^2}}\frac{M\left(\frac{\nu+1}{2};\frac{1}{2};\frac{\mu^2x^2}{2(\nu+x^2)}\right)}{\Gamma(\nu/2+1)}\right], \tag{1.19}$$

where $\Gamma(\nu)$ is the gamma function and $M(a;b;z)$ is Kummer's function of the first kind, a confluent hypergeometric function given by

$$M(a;b;z) = \sum_{k=0}^{+\infty}\frac{(a)_k}{(b)_k}\frac{z^k}{k!}, \tag{1.20}$$

where $(a)_k = a(a+1)\dots(a+k-1)$. See Scharf (1991). We obtain the conditional density $f(y_t|\alpha_t,\alpha_{t+1})$ using the change of variables $y_t = \exp(\alpha_t/2)\sqrt{1-\rho^2}X$. The log conditional density $\psi_t(\alpha_t,\alpha_{t+1}) \equiv \log f(y_t|\alpha_t,\alpha_{t+1})$ and its derivatives are given in Appendix I.4.

For both models, the state equation parameters are $\omega_t = \sigma^{-2}$, $\phi_t = \phi$ and $d_t = (1-\phi)\bar{\alpha}$ for all $t > 1$. The marginal distribution of the initial state α_1 is the stationary distribution, so that $\omega_0 = (1-\phi^2)\omega$ and $d_0 = \bar{\alpha}$.

We express our prior uncertainty about the parameters in terms of a multivariate Gaussian distribution over the transformed parameter vector

$$\theta = (\log\sigma, \tanh^{-1}\phi, \bar{\alpha}, \tanh^{-1}\rho, \log\nu).$$

The marginal distribution of $(\log\sigma, \tanh^{-1}\phi, \bar{\alpha}, \log\nu)$ is the same as the prior in McCausland [16] for a Student's t stochastic volatility model without leverage, and is based on a prior predictive analysis. The parameter $\tanh^{-1}\rho$ is Gaussian and *a priori* independent, with mean -0.4 and standard deviation 0.5. This implies prior quantiles 0.1, 0.5 and 0.9 for ρ approximately equal to -0.78, -0.38 and 0.23. The result is the following prior :

$$\theta \sim N\left(\begin{bmatrix} -1.8 \\ 2.1 \\ -11.0 \\ -0.4 \\ 2.5 \end{bmatrix}, \begin{bmatrix} 0.125 & -0.05 & 0 & 0 & 0 \\ -0.05 & 0.1 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0.25 & 0 \\ 0 & 0 & 0 & 0 & 0.25 \end{bmatrix}\right).$$

1.4.2 MCMC and IS methods for posterior simulation

To illustrate the performance of the HESSIAN approximation, we use Markov chain Monte Carlo (MCMC) and importance sampling posterior simulations and compare with Omori et al. [20]. For both posterior simulations, we draw jointly θ and α. We use as proposal density (resp. importance density) $g(\alpha, \theta|y) = g(\alpha|\theta, y)g(\theta|y)$, based on an approximation $g(\theta|y)$ of $f(\theta|y)$, described below, and the HESSIAN approximation $g(\alpha|\theta, y)$ of $f(\alpha|\theta, y)$.

We construct $g(\theta|y)$ as follows. Just as $g(\alpha|\theta, y)$ is a close approximation of $f(\alpha|\theta, y)$, $\tilde{g}(\theta|y) \doteq f(\alpha, \theta, y)/g(\alpha|\theta, y)$ is a good unnormalised approximation of $f(\theta|y)$. Let θ° be the maximiser of $\tilde{g}(\theta|y)$ and Σ° be the inverse of the negative Hessian of $\log\tilde{g}(\theta|y)$ at θ°. Also let n_θ be the dimension of θ, equal to 4 for the Gaussian model and 5 for the Student's t model.

We choose $g(\theta|y)$ to be a n_θ-variate Student's t density with location parameter θ°, scale matrix Σ°, and degrees of freedom equal to 30.

In the MCMC posterior simulation, we use an independence Metropolis-Hastings chain. The joint proposal $(\alpha^\star, \theta^\star)$ from density $g(\theta|y)g(\alpha|\theta, y)$ is accepted with proba-

bility

$$\pi(\theta^\star, \alpha^\star, \theta, \alpha) = \min\left[1, \frac{f(\theta^\star)f(\alpha^\star|\theta^\star)f(y|\theta^\star, \alpha^\star)}{f(\theta)f(\alpha|\theta)f(y|\theta, \alpha)} \frac{g(\theta|y)g(\alpha|\theta, y)}{g(\theta^\star|y)g(\alpha^\star|\theta^\star, y)}\right].$$

The fact that we can approximate the entire posterior distribution opens up the possibility of doing importance sampling. Unlike proposals in MCMC, importance draws do not need to be independent and this presents opportunities for variance reduction. We exploit this fact to do importance sampling using a combination of quasi-random and pseudo-random sequences for draws of θ. We construct M blocks of length S each, for a total of MS draws. S should be a power of two, which is convenient for Sobol quasi-random sequences.

We draw $U^{(m)}$, $m = 1, \ldots, M$, independently from the uniform distribution on the hypercube $(0, 1)^{n_\theta}$. For $s = 1, \ldots, S$, $V^{(s)}$ is the s'th element of the n_θ-dimensional Sobol sequence. For $m = 1, \ldots, M$ and $s = 1, \ldots, S$, we compute $U^{(m,s)}$, defined as the modulo 1 sum of $U^{(m)}$ and $V^{(s)}$. Thus $U^{(m,s)}$ is uniformly distributed on $(0, 1)^{n_\theta}$ and the M blocks of length S are independent. We use $U^{(m,s)}$ to draw $\theta^{(m,s)}$ from $g(\theta|y)$: use $U^{(m,s)}$ to construct a 6-vector of independent standard Gaussian variates using the inverse cdf method then construct $\theta^{(m,s)}$ by pre-multiplying by the Cholesky decomposition of the scale matrix times $\sqrt{\nu/\omega^2}$, where $\omega^2 \sim \chi^2(\nu)$.

Let $h(\theta, \alpha)$ be any function of interest. The importance sampling estimator for $E[h(\theta, \alpha)|y]$ is N/D, where

$$N \doteq \sum_{m=1}^{M}\sum_{s=1}^{S} w^{(m,s)} h(\theta^{(m,s)}, \alpha^{(m,s)}), \quad D \doteq \sum_{m=1}^{M}\sum_{s=1}^{S} w^{(m,s)},$$

and

$$w^{(m,s)} = \frac{f(\theta^{(m,s)}, \alpha^{(m,s)}, y)}{g(\theta^{(m,s)}, \alpha^{(m,s)}|y)}.$$

If the posterior mean of $h(\theta, \alpha)$ exists, then the ratio $R = N/D$ is a simulation convergent estimator of $E[h(\theta, \alpha)|y]$.

Following Geweke [10], we approximate the posterior variance of $h(\theta, \alpha)$ by

$$\hat{\sigma}_h^2 \doteq \frac{\sum_{m=1}^{M} \sum_{s=1}^{S} [w^{(m,s)}(h(\theta^{(m,s)}, \alpha^{(m,s)}) - R)]^2}{D^2}.$$

We compute a numerical standard error for R using the delta method. This gives the following approximation of the numerical variance of the ratio R :

$$\hat{\sigma}_R^2 \doteq (\hat{\sigma}_N^2 - 2R\hat{\sigma}_{ND} + R^2\hat{\sigma}_D^2)(MS/D)^2,$$

where $\hat{\sigma}_N^2$ and $\hat{\sigma}_D^2$ are estimates of the variances of N and D and $\hat{\sigma}_{ND}$ is an estimate of the covariance. Specifically, $\hat{\sigma}_N^2$ is $(1/M)$ times the sample variance of the M independent terms

$$N_m = \frac{1}{S} \sum_{s=1}^{S} w^{(m,s)} h(\theta^{(m,s)}, \alpha^{(m,s)}), \quad m = 1, \ldots, M,$$

and analogously for $\hat{\sigma}_D^2$ and $\hat{\sigma}_{ND}$. Then $\hat{\sigma}_h^2 / MS\hat{\sigma}_R^2$ is an estimate of the relative numerical efficiency.

1.4.3 Marginal likelihood approximation

Efficient posterior simulation of parameters and states using a single block enables us to compute extremely precise approximations of the marginal likelihood. Using the proposal distribution as an importance distribution for the posterior distribution, the mean of the importance weights is a simulation consistent and simulation unbiased estimator of the marginal likelihood. Our close approximation makes the variation in weights extremely small, which leads to high numerical efficiency for marginal likelihood estimation.

1.4.4 Results

For the ASV-Gaussian model, we report results of the HESSIAN independence Metropolis-Hastings and importance sampling posterior simulations. We implement the procedure of Omori et al. [20], denoted OCSN, and compare results. We apply the three methods to two real data sets. The first consists of daily returns of the S&P 500 index

from January 1980 to December 1987, for a total of 2022 observations. This matches a sample used by Yu [28]. The second data set consists of 1232 daily returns of the TOPIX index. This data set, used by Omori et al. [20], is available at Nakajima's website http ://sites.google.com/site/jnakajimaweb/sv.

In the MCMC posterior simulation, the initial 10 draws are discarded and the independence Metropolis-Hastings chain is of length 12,800. We choose this chain size to match the total draws of the importance sampling chain where we use $M = 100$ and $S = 128$. In our replication of the OCSN chain, the initial 500 values are discarded and we retain the 12,800 subsequent values. Table 1.II gives the computational time by dataset and estimation procedure. For all three methods, the code is written in C++. We used a Windows PC with an Intel Core i5 2.90GHz processor.

Table 1.III summarizes estimation results of the ASV-Gaussian model. The labels *HIS*, *HIM* and *OCSN* indicate the HESSIAN importance sample, the HESSIAN independence Metropolis-Hastings chain, and the chain obtained using the Omori et al. [20] procedure. The first two columns show numerical estimates of the posterior mean and standard deviation, for the various parameters.

The third and fourth columns give the numerical standard error (NSE) and the relative numerical efficiency (RNE) of the numerical approximations of the posterior mean. The RNE measures numerical efficiency relative to that of the mean of a random sample from the posterior. We use the results of Section 1.4.2 to compute the NSE and RNE of the importance sampling chain and the OCSN chain. We use the contributed coda library of the $\mathbf{R}$ software to compute those of the HESSIAN independence Metropolis-Hastings method. This uses a time series method based on the estimated spectral density at frequency zero.

The HIS and HIM methods produce numerical estimates of the same posterior mean. We implement the procedure of Omori et al. [20] using the prior described in their article, which is different from our own. As a result, reported values are different not only because of numerical sample variance but also because the posterior mean is slightly different.

The HESSIAN importance sampler outperforms the OCSN method in all cases. Its

numerical efficiency is higher compared to OCSN, and apart from the unconditional mean $\bar{\alpha}$ of log volatility, at least four times higher. The efficiency of the importance sample means are sometimes greater than 1. This is possible because of the variance reduction achieved by using quasi-random numbers. In addition, the HIS procedure has a lower execution time and thus higher numerical precision per unit time, measured by (1/(Time$\times$ NSE^2)). Except for the unconditional mean of the log volatility, the HESSIAN independence Metropolis Hastings methods outperforms the OCSN procedure, with regard to the relative numerical efficiency and precision per unit time.

The reported posterior means of the parameters ϕ, σ and ρ are similar to the values reported by Omori et al. [20] for the TOPIX index. The difference in the posterior means $\bar{\alpha}$ is due to the fact that these authors measure daily returns in percentages. The same is true for Yu [28] in the case of the S&P500.

For the ASV-Student model, we only report results for the HESSIAN procedures. Table 1.IV summarizes the results of both datasets. The estimates of the parameters $\bar{\alpha}$, ϕ, σ and ρ, for the real data, are close to those obtained with the ASV-Gaussian. The numerical efficiency is also substantially higher.

Nakajima and Omori [18] proposed an extension of the procedure in Omori et al. [20] for ASV-Student and other models. They illustrate the procedure using S&P500 (nominally January 1, 1970 to December 31, 2003) and Topix (Janury 6, 1992 to December 30, 2004) data. Table 4 and Table 5 in Nakajima and Omori [18] report results for S&P500 and Topix data, respectively. Numerical efficiency for the ASV-Student model (SVLt in their paper) ranges from 0.006 (ν) to 0.291 (μ) for the S&P500 dataset. For the Topix data, the highest value of efficiency reported is 0.0893. To compare efficiency, we measured the numerical efficiency of the HESSIAN method, with randomised pseudo-Monte Carlo importance sampling, on S&P500 data from January 1, 1970 to December 31, 2003. Our sample size is 8586, rather than 8869 reported in Nakajima and Omori [18]. We obtain numerical efficiency ranging from 0.91 (ϕ) to 1.01 (μ).

We use the Metropolis-Hastings output to approximate marginal likelihoods. For ASV-Gaussian, we obtain a log marginal likelihood of 6595.91, with a numerical standard error of 0.043 ; for ASV-Student, 6609.67, with a numerical standard error of 0.055.

The Bayes factor of $\exp(13.76)$ decisively favours the ASV-Student model.

1.5 Conclusion

We have derived an approximation $g(\alpha|\theta, y)$ of the target density $f(\alpha|\theta, y)$ that can be used as a proposal density for MCMC or as an importance density for importance sampling. We have tested the correctness of the HESSIAN posterior simulators.

Simulations on artificial and real data suggest that the HESSIAN method, which is not model specific, is more numerically efficient than the model specific method of Omori et al. [20], which is in turn more efficient than the methods of Jacquier et al. [13] and Omori and Watanabe [19]. The high numerical efficiency relies on $g(\alpha|\theta, y)$ being extremely close to the target density $f(\alpha|\theta, y)$. Constructing a joint proposal of θ and α not only solves the problem of numerical inefficiencies due to posterior autocorrelation of α but also those due to posterior dependence between θ and α.

The scope of applications goes beyond the ASV-Gaussian and ASV-Student models. Application to a new model of the form (1.4) only requires routines to compute partial derivatives of the log conditional densities $\log f(y_t|\alpha_t, \alpha_{t+1})$ with respect to α_t and α_{t+1}. This requirement is not as demanding as it might first appear, for two reasons. First, we can use numerical derivatives or other approximations. Second, we do not require analytic expressions of these derivatives. If $\log f(y_t|\alpha_t, \alpha_{t+1})$ is a composition of primitive functions, we can combine evaluations of the derivatives of the primitive functions using routines applying Fàa Di Bruno's rule for multiple derivatives of compound functions. We have already coded these routines, which do not depend on the particular functions involved.

We now require the state vector, α, to be Gaussian. We are currently trying to extend the HESSIAN method to models where the state vector is Markov, but not necessary Gaussian. We are also working on approximations to filtering densities, useful for sequential learning.

Notation	Description
$\psi_t(\alpha_t, \alpha_{t+1})$	$\log f(y_t\|\alpha_t, \alpha_{t+1})$
$\psi_t^{p,q}(\alpha_t, \alpha_{t+1})$	derivative of $\psi_t^{p,q}(\alpha_t, \alpha_{t+1})$ with respect to α_t and α_{t+1} of orders p and q.
$\psi_n(\alpha_n)$	$\log f(y_n\|\alpha_n)$
$\psi_n^p(\alpha_n)$	p'th derivative of $\psi_n(\alpha_n)$ with respect to α_n
$a = (a_1, \ldots, a_n)$	mode of $\log f(\alpha\|y)$
Σ_t	Var$(\alpha_t\|\alpha_{t+1}, y)$ for the 1st reference distribution
$(a_{1\|t+1}(\alpha_{t+1}), \ldots, a_{t\|t+1}(\alpha_{t+1}))$	mode of the conditional density $f(\alpha_1, \ldots, \alpha_t\|\alpha_{t+1}, y)$
$\Sigma_{t\|t+1}(\alpha_{t+1})$	Var$(\alpha_t\|\alpha_{t+1}, y)$ for the 2nd reference distribution
$A_{t\|t+1}(\alpha_{t+1})$	polynomial approximation of $a_{t\|t+1}(\alpha_{t+1})$
$s_{t\|t+1}(\alpha_{t+1})$	$\log \Sigma_{t\|t+1}(\alpha_{t+1})$
$a_t^{(r)}, r = 1, \ldots, R$	r'th derivative of $a_{t\|t+1}(\alpha_{t+1})$ at $\alpha_{t+1} = a_{t+1}$
$s_t^{(r)}, r = 1, \ldots, R-1$	r'th derivatives of $s_{t\|t+1}(\alpha_{t+1})$ at $\alpha_{t+1} = a_{t+1}$.
$b_{t\|t+1}(\alpha_{t+1})$	mode of the conditional density $f(\alpha_t\|\alpha_{t+1}, y)$
$b_t, b_t^{(r)}, r = 1, \ldots, R$	value and derivatives of $b_{t\|t+1}(\alpha_{t+1})$ at $\alpha_{t+1} = a_{t+1}$
b_n	mode of the conditional density $f(\alpha_n\|y)$
$B_{t\|t+1}(\alpha_{t+1})$	polynomial approximation of $b_{t\|t+1}(\alpha_{t+1})$
$B_t, B_t^{(r)}, r = 1, \ldots, R$	value and derivatives of $B_{t\|t+1}(\alpha_{t+1})$ at $\alpha_{t+1} = a_{t+1}$
$\mu_{t\|t+1}(\alpha_{t+1})$	$E[\alpha_t\|\alpha_{t+1}, y]$
$\mu_t, \mu_t^{(r)}, r = 1, 2$	value and two derivatives of $\mu_{t\|t+1}(\alpha_{t+1})$ at $\alpha_{t+1} = a_{t+1}$
$M_{t\|t+1}(\alpha_{t+1})$	polynomial approximation of $\mu_{t\|t+1}(\alpha_{t+1})$
$M_t, M_t^{(r)}, r = 1, 2$	value and two derivatives $M_{t\|t+1}(\alpha_{t+1})$ at $\alpha_{t+1} = a_{t+1}$
$h_t(\alpha_t; \alpha_{t+1})$	first derivative of $\log f(\alpha_t\|\alpha_{t+1}, y)$ with respect to α_t
$H_t^{(p)}(\alpha_t; \alpha_{t+1}), p \geq 1$	approximation of $h_t^{(p)}(\alpha_t; \alpha_{t+1})$, p'th derivatives of $h_t(\alpha_t; \alpha_{t+1})$ with respect to α_t
$h_t(\alpha_n)$	first derivative of $\log f(\alpha_n\|y)$ with respect to α_n
$H_n^{(p)}(\alpha_n), p \geq 1$	approximation of the p'th derivatives of $h_t(\alpha_n)$ with respect to α_n

TABLE 1.I – Main notation used in the paper

	SP500	TOPIX
HIS	70	43
HIM	85	45
OCSN	108	67

TABLE 1.II – Computational time in seconds by dataset and estimation procedure for the ASV-Gaussian model. For all procedures, we draw a chain of size 12800. The S&P500 dataset is of size 2022 and the Topix dataset is of size 1232.

Parameters	Mean	Std	NSE	RNE
		S& P500		
$\bar{\alpha}$:HIS	-9.5167	0.1573	2.0113e-3	0.9082
$\bar{\alpha}$:HIM	-9.5181	0.1583	3.1266e-3	0.2002
$\bar{\alpha}$:OCSN	-9.5029	0.3378	3.4767e-3	0.7428
ϕ :HIS	0.9751	0.0080	8.9356e-5	0.9000
ϕ :HIM	0.9752	0.0081	1.3592e-4	0.2765
ϕ :OCSN	0.9776	0.0083	1.8947e-4	0.1506
σ :HIS	0.1524	0.0200	1.9681e-4	0.9871
σ :HIM	0.1521	0.0201	3.2814e-4	0.2919
σ :OCSN	0.1394	0.0203	5.8443e-4	0.0945
ρ :HIS	-0.2032	0.0957	9.2493e-4	1.0647
ρ :HIM	-0.2044	0.0950	1.3265e-3	0.4005
ρ :OCSN	-0.2007	0.1005	1.8453e-3	0.2374
		TOPIX		
$\bar{\alpha}$:HIS	-8.8545	0.1080	1.1533e-3	1.2014
$\bar{\alpha}$:HIM	-8.8545	0.1083	1.5951e-3	0.4609
$\bar{\alpha}$:OCSN	-8.8426	0.2172	2.0867e-3	0.8574
ϕ :HIS	0.9574	0.0156	1.5893e-4	0.9537
ϕ :HIM	0.9576	0.0160	2.0428e-4	0.4769
ϕ :OCSN	0.9520	0.0185	3.9992e-4	0.1664
σ :HIS	0.1408	0.0254	2.5871e-4	0.8657
σ :HIM	0.1414	0.0258	2.8818e-4	0.6277
σ :OCSN	0.1387	0.0266	5.9850e-4	0.1556
ρ :HIS	-0.3833	0.1188	1.2561e-3	0.8503
ρ :HIM	-0.3833	0.1195	1.7136e-3	0.3801
ρ :OCSN	-0.3715	0.1231	2.6536e-3	0.1792

TABLE 1.III – ASV-Gaussian parameter estimation using the HESSIAN method and the OCSN procedure on S&P500 and TOPIX data.

Parameters	Mean	Std	NSE	RNE
		S& P 500		
$\bar{\alpha}$:HIS	-9.7230	0.1865	2.8719e-3	1.0496
$\bar{\alpha}$:HIM	-9.7224	0.1806	3.1769e-3	0.2525
ϕ :HIS	0.9851	0.0054	6.8752e-5	0.9663
ϕ :HIM	0.9850	0.0053	7.9290e-5	0.3513
σ :HIS	0.1061	0.0164	1.7719e-4	1.1002
σ :HIM	0.1065	0.0164	3.0925e-4	0.2204
ρ :HIS	-0.2440	0.1224	1.6006e-4	0.8261
ρ :HIM	-0.2493	0.1222	2.2437e-3	0.2318
ν :HIS	9.8647	2.1622	2.4734e-2	0.9722
ν :HIM	9.9128	2.1828	3.6789e-2	0.2750
		TOPIX		
$\bar{\alpha}$:HIS	-8.9488	0.1156	1.5983e-3	0.9672
$\bar{\alpha}$:HIM	-8.9506	0.1115	1.9474e-3	0.2560
ϕ :HIS	0.9624	0.0142	1.7252e-4	0.8727
ϕ :HIM	0.9621	0.0144	2.2029e-4	0.3336
σ :HIS	0.1261	0.0242	2.6775e-4	0.9570
σ :HIM	0.1266	0.0240	3.7636e-4	0.3188
ρ :HIS	-0.4194	0.1285	1.3790e-4	1.1266
ρ :HIM	-0.4191	0.1236	2.2023e-3	0.2461
ν :HIS	20.6041	7.6904	8.6997e-2	0.9573
ν :HIM	20.4777	7.7394	1.4048e-1	0.2371

TABLE 1.IV – ASV-Student parameter estimation using the HESSIAN method, Independence Metropolis-Hastings and Importance Sampling, on S&P500 and TOPIX data.

CHAPITRE 2

GENERALIZED HESSIAN FOR NON-LINEAR AND NON-GAUSSIAN STATE SPACE MODELS

Abstract

We develop a method for simulation smoothing and parameter estimation in non-linear state space models where not only observations, but also states, may be non-Gaussian. Given the current state, the current observation vector and the contemporaneous state innovation need not be independent. States are univariate, but observations may be multivariate.

Our method is an extension of the HESSIAN method described in Djegnene and McCausland [4], for models where states are linear and Gaussian and where the conditional density of the current state innovation and observation, given the current state value, decomposes analytically into a marginal density for the state innovation and a conditional density for the observation given the state innovation. Like that method, ours is based on a close approximation of the conditional density of all states given all observations and parameters. We can use this approximation to construct a joint proposal density of states and parameters, for MCMC posterior simulation, or a joint importance density of states and parameters for importance sampling. Applications include the approximation of likelihood function values and the marginal likelihood, and Bayesian posterior simulation.

The procedure used to construct the approximation of the conditional density of the state vector given the observed vector and parameters is not model specific. For a given state space model, we only require routines to compute partial derivatives of the log conditional density of states, with respect to states.

We illustrate using the stochastic conditional duration model with 'leverage effect' described in Feng et al. [6]. For the numerical approximation of posterior means of parameters, our generic procedure is more numerically efficient than other posterior simulation methods for similar models. Using randomized pseudo Monte Carlo importance

sampling, we obtain relative numerical efficiencies close to 100% for all parameters while numerical efficiencies of the posterior simulator described in Strickland et al. [18] do not exceed 5%.

Keywords : State space models, Nonlinear, Non-Gaussian, MCMC, Numerical Efficiency, Stochastic Volatility

JEL Classification : C11, C15, C58, C63

2.1 Introduction

We design inferential methods for state space models where not only observations, but also states can be non-linear and non-Gaussian. These models are described by the joint density function

$$f(\alpha, y|\theta) = f(\alpha_1|\theta)\left[\prod_{t=1}^{n-1} f(\alpha_{t+1}, y_t|\alpha_t, \theta)\right] f(y_n|\alpha_n, \theta). \tag{2.1}$$

In formulation (2.1), $\alpha = (\alpha_1, \ldots, \alpha_n)^\top$ is a vector of latent univariate states, $y = (y_1^\top, \ldots, y_n^\top)^\top$ is a vector of observations, and θ is a vector of parameters.

The densities $f(\alpha_{t+1}, y_t|\alpha_t, \theta)$ are quite flexible. The state α_{t+1} and observation y_t may be conditionally dependent given α_t, and neither needs to be conditionally Gaussian. This allows models such as stochastic volatility models with leverage and asymmetric stochastic conditional duration (ASCD) models. Unlike many formulations of state space models, we do not require an analytic decomposition $f(\alpha_{t+1}, y_t|\alpha_t, \theta) = f(\alpha_{t+1}|\alpha_t, \theta)f(y_t|\alpha_t, \alpha_{t+1}, \theta)$.

Simulation smoothing methods have proven useful for approximating likelihood function values and Bayesian posterior simulation. They involve simulating the conditional distribution of states given data and parameters. We will call this distribution the *target* distribution and $f(\alpha|y, \theta)$, the *target* density. Simulation typically entails importance sampling (IS) or Markov chain Monte Carlo (MCMC). We show examples of both in Section 2.3.

Designing inferential methods for nonlinear and non-Gaussian state space models has proven difficult. Existing methods are generally model specific. They are often designed for linear and Gaussian states or use linear transformations.

We extend the HESSIAN method in Djegnene and McCausland [4] which does simulation smoothing for models with leverage-like effects. This method uses multiple partial derivatives of $\log f(y_t|\theta, \alpha_t, \alpha_{t+1})$ with respect to α_t and α_{t+1} to construct a close approximation of the target density. Here, we construct the approximation of the target density using multiple partial derivatives of $\log f(\alpha_{t+1}, y_t|\alpha_t, \theta)$ with respect to α_t and α_{t+1}.

In contrast to Djegnene and McCausland [4], we do not need the density $f(\alpha_{t+1}|\alpha_t, \theta)$ to be available in closed form, or to be Gaussian.

The HESSIAN method originates from McCausland [10], who constructs a close approximation of the target density for models with conditionally independent states and observations, and Gaussian states. Our method inherits the following features of the original method :

1. It involves direct simulation of states from their posterior distribution using a proposal or importance distribution approximating the target distribution. This is unlike auxiliary mixture model approaches, in which a model is first linearized, and then any non-Gaussian distributions in the transformed model are approximated by finite Gaussian mixtures. Xu et al. [21] use this auxiliary mixture model approach for stochastic conditional duration (SCD) models. Omori et al. [13] use it for estimating asymmetric stochastic volatility models based on the well established method of Kim et al. [9]. Using the direct approach, we avoid model-specific transformations, data augmentation, and the need to re-weight or apply additional accept-reject steps to correct for approximation error. See Djegnene and McCausland [4] for additional discussion.
2. It involves drawing the entire state sequence as a single MCMC block. This leads to efficiency improvements when there is posterior serial dependence. The entire state sequence is drawn with a proposal or importance sampling density that is not necessary Gaussian. Many articles have used multivariate Gaussian proposal distributions to update the state vector, but usually only for about 10–50 observations at a time, not the entire sample. These include Shephard and Pitt [17], Watanabe and Omori [20], Strickland et al. [18], Jungbacker and Koopman [8] and Omori and Watanabe [12]. The Efficient Importance Sampling (EIS) method of Richard and Zhang [14] features draws of the entire state sequence as a block, but since their approximate target distribution is constructed using the random numbers used to draw variates from it, EIS estimators of likelihood function values do not have the simulation consistency or lack of simulation bias that true importance sampling estimators do. See the discussion in McCausland [10] for more details.

3. Since the approximation is so close, we can draw parameters and states together as a single block. We do this using a joint proposal distribution combining our approximation of the conditional posterior distribution of states given parameters with an approximation of the marginal posterior distribution of parameters. Drawing states and parameters in a single block leads to further efficiency improvements because of posterior dependence between states and parameters. In this way, we achieve numerical efficiencies comparable to model-specific auxiliary mixture model approaches, which also often feature joint draws of parameter and states. The examples of Section 2.3 suggest that our method is even more efficient than these approaches, partly because we avoid data augmentation and the need to correct for approximation error. Being able to draw all parameters and states jointly in an untransformed model also opens up new opportunities — it allows for importance sampling, variance reduction using randomised pseudo Monte Carlo, and very efficient approximations of the marginal likelihood, as we see in Section 2.3.
4. We construct our approximation of the target distribution in a generic way. The only model-specific computation is the evaluation of partial derivatives of the log joint density of states and observations. We can easily compute analytic expressions for the derivatives of the log density of state and observations for the ASCD-models used for illustration. However, exact evaluation of derivatives does not require finding analytic expressions. Although we do not do it here, we can use generic routines to combine derivative values according to Leibniz' rule for multiple derivatives of products and Faà di Bruno rule for derivatives of composite functions. See Djegnene and McCausland [4] and particularly Appendix F of that article. Also, we could also resort to numerical derivatives — there would a cost in numerical efficiency, but simulation consistency would not be compromised. A generic approach allows for a wide scope of application of the Generalized HESSIAN method introduced here.
5. It is based on operations using the sparse Hessian matrix of the log target density, rather than on the Kalman filter. Articles using the former approach include Rue [15], for linear Gaussian Markov random fields, Chan and Jeliazkov [2] and

McCausland et al. [11], for linear Gaussian state space models, and Rue et al. [16] for non-linear non-Gaussian Markov random fields. The Integrated Nested Laplace Approximation (INLA) method described in the last article has spawned a large applied literature. Articles using the Kalman filter include Feng et al. [6], Strickland et al. [18], for SCD-models, and Carter and Kohn [1], Frühwirth-Schnatter [7], de Jong and Shephard [3] and Durbin and Koopman [5] for linear Gaussian state space models. Auxiliary mixture model methods for non-linear or non-Gaussian models tend to use the Kalman filter, but this is not an essential feature of auxiliary mixture model methods.

We use the stochastic conditional duration model with the leverage effect (ASCD) introduced by Feng et al. [6] to illustrate the Generalized HESSIAN method. Like them, we consider ASCD models with Exponential, Weibull and Gamma innovations for the observation variable. Unlike them, we do not transform the model into a linear state space model.

Following common practice, Feng et al. [6] fit ASCD models for non-zero durations only. They used the fitted parameters to simulate a large sample of durations. Comparison of model-simulated and observed durations show that model-simulated durations do not capture well the left tail of observed durations. We will see in Section 2.3.4 that including data for durations of zero lead to a better fit for the left tail of the observed durations.

The rest of the article is organized as follow. In Section 2.2 we describe our approximation $g(\alpha|y,\theta)$ of the target density $f(\alpha|y,\theta)$. We show how to evaluate it and how to draw from the distribution with density $g(\alpha|y,\theta)$. Section 2.3 describes the asymmetric stochastic conditional duration model. It also demonstrates the correctness of the posterior simulators and discusses the empirical results. Section 2.4 concludes.

2.2 An approximation of the target density

We define here our approximation $g(\alpha|y,\theta)$ of the target density $f(\alpha|y,\theta)$. The approximation is a proper and fully normalized density. We do not provide a closed form expression of this approximation, but instead show how to sample from and evaluate it

in $O(n)$ operations.

We will now condition on the vector of parameters, θ, and suppress it from the notation in this section and related appendices. Later in the empirical illustration, Section 2.3, we will be explicit again about this conditioning.

The procedure used to construct the approximation $g(\alpha|y)$ is not model specific. For a particular state space model, we construct the approximation $g(\alpha|y)$ by supplying routines to compute the value of $\log f(\alpha|y)$, up to normalization factor, and various partial derivatives with respect to elements of α. Using Equation (2.1), this amounts to computing, for $t = 1, \ldots, n-1$, the functions

$$\psi_0(\alpha_1) \doteq \log f(\alpha_1), \quad \psi_t(\alpha_t, \alpha_{t+1}) \doteq \log f(y_t, \alpha_{t+1}|\alpha_t), \quad \psi_n(\alpha_n) \doteq \log f(y_n|\alpha_n), \tag{2.2}$$

and the partial derivatives

$$\psi_0^{(q)}(\alpha_1) = \frac{\partial^q \psi_0(\alpha_1)}{\partial \alpha_1^q}, \quad \psi_t^{(p,q)}(\alpha_t, \alpha_{t+1}) \doteq \frac{\partial^{p+q} \psi_t(\alpha_t, \alpha_{t+1})}{\partial \alpha_t^p \partial \alpha_{t+1}^q}, \quad \psi_n^{(p)}(\alpha_n) \doteq \frac{\partial^p \psi(\alpha_n)}{\partial \alpha_n^p}, \tag{2.3}$$

for orders p and q up to certain values P and Q. For convenience, Table 2.I summarizes this and other important notation.

The routines to compute $\psi_0(\alpha_1)$, the $\psi_t(\alpha_t, \alpha_{t+1})$ and $\psi_n(\alpha_n)$ must give exact results, as they are used to evaluate the target density $f(\alpha|y)$ up to a normalization factor. The partial derivatives, however, may be numerical derivatives or other approximations. The approximation errors may make $g(\alpha|y)$ a cruder approximation of $f(\alpha|y)$ and thus diminish the numerical precision of IS or MCMC. But we will still be able to evaluate $g(\alpha|y)$ and sample from it without error, and so it does not compromise the simulation consistency of IS or MCMC.

Like the target density, the approximation $g(\alpha|y)$ has the Markov property that allows to decompose it as

$$g(\alpha|y) = g(a_n|y) \prod_{t=n-1}^{1} g(\alpha_t|\alpha_{t+1}, y), \tag{2.4}$$

where each factor is a proper and fully normalized density function closely approxima-

ting the corresponding factor of $f(\alpha|y)$. Whether we need to evaluate $g(\alpha|y)$, simulate it or both, the decomposition allows us to do so sequentially, for t descending from n to 1.

The densities $g(\alpha_t|\alpha_{t+1}, y)$ are members of the five-parameter perturbed Gaussian distribution described in Appendix G of McCausland [10]. The parameters give a mode of the distribution and the second through fifth derivatives of $\log g(\alpha_t|\alpha_{t+1}, y)$ at that mode. Choosing parameters amounts to approximating $b_{t|t+1}(\alpha_{t+1})$, defined as the mode of $f(\alpha_t|\alpha_{t+1}, y)$, and the second through fifth derivatives of $\log f(\alpha_t|\alpha_{t+1}, y)$ at that approximate mode.

Approximations rely on Taylor series expansions, some exact and some approximate, of various functions, including $b_{t|t+1}(\alpha_{t+1})$ and the conditional mean of $\psi_t(\alpha_t, \alpha_{t+1})$ given α_{t+1} and y, denoted $x_{t|t+1}(\alpha_{t+1})$. Some expansions are computed during a forward pass, around the mode $(a_1, \ldots, a_n)$ of the target distribution, a static point of expansion. So for example, we compute $B_{t|t+1}(\alpha_{t+1})$ and $X_{t|t+1}(\alpha_{t+1})$ as approximate Taylor series expansions of $b_{t|t+1}(\alpha_{t+1})$ and $x_{t|t+1}(\alpha_{t+1})$ around a_{t+1}.

During the backward pass, we compute approximate Taylor series expansions of $h_n(\alpha_n) \doteq \log f(\alpha_n|y)$ and $h_t(\alpha_t; \alpha_{t+1}) \doteq \log f(\alpha_t|\alpha_{t+1}, y)$, $t = n-1, \ldots, 1$, which we will treat as univariate functions of α_t with parameter α_{t+1}. Here, the point of expansion is a moving target, depending on α_{t+1}. The expansion is fifth order, allowing a much better than Gaussian (second order) approximation. This expansion is based on the following exact result for the first derivative of $h_t(\alpha_t; \alpha_{t+1})$:

$$h_t^{(1)}(\alpha_t; \alpha_{t+1}) = x_{t-1|t}(\alpha_t) + \psi_t^{(1,0)}(\alpha_t, \alpha_{t+1}), t = 2, \ldots, n-1, \tag{2.5}$$

where $x_{t-1|t}(\alpha_t) \doteq E[\psi_{t-1}^{(0,1)}(\alpha_{t-1}, \alpha_t)|\alpha_t, y]$. Equation (2.5), and analogous results for the cases $t = 1$ and $t = n$, are derived in Appendix II.3.1.

We cannot evaluate the $x_{t-1|t}(\alpha_t)$ and their derivatives exactly. Nor can we evaluate the mode $b_{t|t+1}(\alpha_{t+1})$ exactly. Instead, we provide polynomial approximations $X_{t-1|t}(\alpha_t)$ of $x_{t-1|t}(\alpha_t)$ and $B_{t|t+1}(\alpha_{t+1})$ of $b_{t|t+1}(\alpha_{t+1})$. We use these to approximate the value $b_{t|t+1}(\alpha_{t+1})$ and the derivatives $h_t^{(r)}(\alpha_t; \alpha_{t+1})$, $r = 1, \ldots, 5$.

The polynomial $X_{t-1|t}(\alpha_t)$ approximates the Taylor expansion $x_{t-1|t}(\alpha_t)$ around a_t.

Similarly, the polynomial $B_{t|t+1}(\alpha_{t+1})$ is an approximation of a Taylor expansion of $b_{t|t+1}(\alpha_{t+1})$ around a_{t+1}.

We draw α, evaluate $g(\alpha|y)$, or both using the following steps. We first compute the mode $a = (a_1, \ldots, a_n)$ of the target distribution using the method described in Appendix B of McCausland [10]. In a forward pass we compute the coefficients of the polynomials $B_{t|t+1}(\alpha_{t+1})$, and $X_{t|t+1}(\alpha_{t+1})$, for $t = 1, \ldots, n-1$. Finally, we compute, for $t = n, \ldots, 1$, $B_{t|t+1}(\alpha_{t+1})$ and $H_t^{(r)}(B_{t|t+1}(\alpha_{t+1}); \alpha_{t+1})$, using these values as the parameters of the perturbed Gaussian distribution. With these values set, we can draw α_t, evaluate $g(\alpha_t|\alpha_{t+1}, y)$ or both. In the rest of this section, we describe these steps in more detail. Full detail is left to various appendices.

2.2.1 Precomputation

We compute the mode a of the target distribution. This gives, as bi-products, several quantities used later. This includes the precision $\bar{\bar{\Omega}}$ and covector $\bar{\bar{c}}$ of a Gaussian approximation $\mathrm{N}(\bar{\bar{\Omega}}^{-1}\bar{\bar{c}}, \bar{\bar{\Omega}}^{-1})$ of the target density. It also gives the conditional variances $\Sigma_t \doteq \mathrm{Var}[\alpha_t|\alpha_{t+1}]$, $t = 1, \ldots, n-1$, and $\Sigma_n \doteq \mathrm{Var}[\alpha_n]$ implied by this Gaussian approximation.

This precomputation is similar to that described in Appendix B of McCausland [10]. Little modification is required, and we give details in Appendix II.1.

2.2.2 A Forward Pass

In order to describe the forward pass, it will be helpful to introduce a sequence of multivariate Gaussian conditional distributions. We define, for $t = 1, \ldots, n-1$, $(a_{1|t+1}(\alpha_{t+1}), \ldots, a_{t|t+1}(\alpha_{t+1}))$ as the conditional mode of $(\alpha_1, \ldots, \alpha_t)$ given α_{t+1} and y, and $\bar{\bar{\Omega}}_{1:t|t+1}$ as the negative Hessian matrix of $\log f(\alpha_1, \ldots, \alpha_t|\alpha_{t+1}, y)$ with respect to $(\alpha_1, \ldots, \alpha_t)$, evaluated at $(a_{1|t+1}, \ldots, a_{t|t+1})$. Thus we can view the distribution $N((a_{1|t+1}, \ldots, a_{t|t+1}), \bar{\bar{\Omega}}^{-1}_{1:t|t+1})$ as an approximation of the conditional distribution of $(\alpha_1, \ldots, \alpha_t)$ given α_{t+1} and y. Result 2.1 of McCausland et al. [11] implies that if $x \sim N((a_{1|t+1}, \ldots, a_{t|t+1}), \bar{\bar{\Omega}}^{-1}_{1:t|t+1})$, then $x_t|x_{t+1} \sim N(a_{t|t+1}, \Sigma_{t|t+1})$, where $\Sigma_{t|t+1}$ is

the final value in the following forward recursion :

$$\Sigma_{1|t+1} \doteq \bar{\bar{\Omega}}_{11}^{-1}, \qquad \Sigma_{\tau|t+1} \doteq (\bar{\bar{\Omega}}_{\tau\tau} - \bar{\bar{\Omega}}_{\tau,\tau-1}^{2}\Sigma_{\tau-1|t+1})^{-1}, \quad \tau = 2, \ldots, t. \tag{2.6}$$

We also define, for $t = 1, \ldots, n-1$, $s_{t|t+1}(\alpha_{t+1}) \doteq \log \Sigma_{t|t+1}(\alpha_{t+1})$.

Coefficients of the polynomial approximations $B_{t|t+1}(\alpha_{t+1})$ and $X_{t|t+1}(\alpha_{t+1})$ rely on those of $a_{t|t+1}(\alpha_{t+1})$ and $\Sigma_{t|t+1}(\alpha_{t+1})$ as demonstrated in Appendix II.3. Thus, the forward pass consists of performing the following steps, for $t = 1, \ldots, n-1$:

1. Compute

$$\begin{aligned} a_t^{(r)} &\doteq \left.\frac{\partial^r a_{t|t+1}(\alpha_{t+1})}{\partial \alpha_{t+1}^r}\right|_{\alpha_{t+1}=a_{t+1}}, r = 1, \ldots, R, \\ s_t^{(r)} &\doteq \left.\frac{\partial^r s_{t|t+1}(a_{t+1})}{\partial \alpha_{t+1}^r}\right|_{\alpha_{t+1}=a_{t+1}}, r = 1, \ldots, R-1. \end{aligned} \tag{2.7}$$

   The choice of R determines how closely we can approximate the functions $a_{t|t+1}(\alpha_{t+1})$ and $s_{t|t+1}(\alpha_{t+1})$ using Taylor expansions. For our empirical illustration, we use $R = 5$.

   Appendix II.2 gives details and prove that these computations are exact. Equation (II.16) gives explicit simplified expressions for $a_t^{(r)}$ and $s_t^{(r)}$, for $r \leq 5$.

2. Compute approximations B_t, $B_t^{(1)}$, $B_t^{(2)}$ and $B_t^{(3)}$ of the value and first three derivatives of the conditional mode $b_{t|t+1}(\alpha_{t+1})$ at a_{t+1}. For the special case $t = n$, we only compute an approximation B_n of the value b_n. Appendix II.3.3 defines these approximations and shows how to compute them. Specifically, Equation (II.23) defines $B_t^{(r)}$ as a function of the $a_t^{(i)}$ and $s_t^{(i)}$. These computations are based on an approximation of the difference $b_{t|t+1}(\alpha_{t+1}) - a_{t|t+1}(\alpha_{t+1})$ using a first order necessary condition for $b_{t|t+1}(\alpha_{t+1})$ to maximize $f(\alpha_t|\alpha_{t+1}, y)$.

3. Compute approximations $X_t^{(r)}$, $r = 0, \ldots, 4$, of the value and first four derivatives of the conditional mean $x_{t|t+1}(\alpha_{t+1})$ at a_{t+1}. Appendices (C.2) to (C.4) of Djegnene and McCausland [4] give details of these computations.

2.2.3 A Backward Pass

We use the backward pass to draw a random variate α^* from the distribution with density $g(\alpha|y)$ and evaluate $g(\alpha^*|y)$. One can also evaluate $g(\alpha|y)$ at an arbitrary value α^* without drawing.

To implement the backward pass, we use the following approximation of the derivative of $\log f(\alpha_t|\alpha_{t+1}, y)$, based on (2.5),

$$H_t^{(1)}(\alpha_t; \alpha_{t+1}) \doteq \psi_t^{(1,0)}(\alpha_t, \alpha_{t+1}) + X_{t-1|t}(\alpha_t), r = 2, \ldots, 5, \tag{2.8}$$

where X is the approximate Taylor series expansion of x constructed using the coefficients computed in step (3) of the forward pass.

For some models, there may be computational costs to evaluate $\psi_t^{(p,q)}(\alpha_t, \alpha_{t+1})$ directly. In this case, we use an approximation $\Psi_t^{(p,q)}(\alpha_t, \alpha_{t+1})$ of $\psi_t^{(p,q)}(\alpha_t, \alpha_{t+1})$. We already have $\psi_t^{(p,q)} = \psi_t^{(p,q)}(a_t, a_{t+1})$ from the forward pass, and based on a multivariate Taylor expansion, we choose to approximate $\psi_t^{(p,q)}(\alpha_t, \alpha_{t+1})$ by

$$\Psi_t^{(p,q)}(\alpha_t, \alpha_{t+1}) \doteq \sum_{r=0}^{P-p} \sum_{s=0}^{Q-q} \psi_t^{(p+r,q+s)} \frac{(\alpha_t - a_t)^r}{r!} \frac{(\alpha_{t+1} - a_{t+1})^s}{s!}. \tag{2.9}$$

The backward pass consists of performing the following steps, for $t = n, \ldots, 1$.

1. Evaluate $B_{t|t+1}(\alpha_{t+1}^*)$, where $B_{t|t+1}(\alpha_{t+1})$ is the polynomial given by [1]

$$B_{t|t+1}(\alpha_{t+1}) = \sum_{r=0}^{R} \frac{B_t^{(r)}}{r!} (\alpha_{t+1} - a_{t+1})^r \tag{2.10}$$

2. Evaluate the first through fourth derivatives of $H_t^{(1)}(\alpha_t; \alpha_{t+1}^*)$ with respect to α_t at $B_{t|t+1}(\alpha_{t+1}^*)$ using (2.8).

3. The density $g(\alpha_t|\alpha_{t+1}, y)$ is a member of the five-parameter perturbed Gaussian distribution described in Appendix F of McCausland [10]. The parameters are given by $\bar{z} = B_{t|t+1}(\alpha_{t+1}^*)$, and $h_r = H_t^{(r)}(B_{t|t+1}(\alpha_{t+1}^*); \alpha_{t+1}^*)$, $r = 2, \ldots, 5$. These

1. For $r \geq 4$, we set $B_t^{(r)} = a_t^{(r)}$.

give the desired mode $B_{t|t+1}(\alpha^*_{t+1})$ and desired derivatives of $\log g(\alpha_t|\alpha_{t+1}, y)$ at this mode.

4. Draw α^*_t from this distribution and evaluate $g(\alpha_t|\alpha_{t+1}, y)$ at α^*_t and α^*_{t+1}.

2.3 Empirical example

2.3.1 Models

We consider the asymmetric stochastic conditional duration (ASCD) model introduced by Feng et al. [6], described by equation (2.11). In contrast to these authors, we do not linearize the model.

$$\begin{aligned} y_t &= \exp(\alpha_t)\varepsilon_t,\ t = 1, \ldots, n \\ \alpha_{t+1} &= (1-\phi)\bar{\alpha} + \phi\alpha_t + \rho\log(\varepsilon_t) + \eta_t,\ t = 1, \ldots, n-1. \end{aligned} \tag{2.11}$$

The innovation vectors (ε_t, η_t) are independent and identically distributed. Also, ε_t and η_t are independent. The state innovation η_t is Gaussian with mean zero and variance σ^2. The observation innovation ε_t is scale-normalized and is either Weibull, Gamma or Exponential. The Weibull and Gamma distributions have a shape parameter ν. We assume stationarity of the state, so that $\alpha_1 \sim N\left(\bar{\alpha}, \sigma^2/(1-\phi^2)\right)$. We will call the three models ASCD-Exponential, ASCD-Weibull and ASCD-Gamma.

The latent process α represents expected duration while y is actually the observed duration. Trade durations may have local asymmetric changes. In period of high trade intensity, trade durations are lower than in period of low trade intensity. This is equivalent to a positive correlation between expected duration and observed duration. The ASCD model described in (2.11) is intended to capture this stylized facts. The introduction of the observation innovation in the state equation clearly introduce a correlation between the contemporous state innovation and the contemporous observation.

Section 1.2 of Feng et al. [6] describes the statistical properties of the ASCD models. As noted by these authors, the presence of the *leverage term*, γ, inflates the variance and fourth moment of the duration. The skewness and kurtosis vary in terms of both the sign

and magnitude of γ, which offer more flexibility to model the dynamic of the duration.

We compute the log conditional density $\psi_t(\alpha_t, \alpha_{t+1}) = \log f(\alpha_{t+1}, y_t|\alpha_t)$ in Appendix II.4, and provide partial derivatives of $\psi_t(\alpha_t, \alpha_{t+1})$ with respect to α_t and α_{t+1}.

We express our prior uncertainty about the parameters in terms of a multivariate Gaussian distribution over the transformed parameter vector $\theta = (\bar{\alpha}, \tanh^{-1}\phi, \log\sigma, \tanh^{-1}\rho, \log\nu)$. We choose as prior distribution

$$\theta \sim N\left(\begin{bmatrix} 0.0 \\ 1.5 \\ -1.5 \\ 0.1 \\ -0.1 \end{bmatrix}, \begin{bmatrix} 25 & 0 & 0 & 0 & 0 \\ 0 & 0.625 & -0.5 & 0 & 0 \\ 0 & -0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}\right).$$

The marginal distribution of $(\bar{\alpha}, \tanh^{-1}\phi, \log\sigma)$ is from McCausland [10].

2.3.2 MCMC and IS methods for posterior simulation

To illustrate the performance of the Generalized HESSIAN method, we run Markov chain Monte Carlo (MCMC) and importance sampling posterior simulations. For both posterior simulations, we jointly draw θ and α. We use as proposal density (resp. importance density) $g(\alpha, \theta|y) = g(\alpha|\theta, y)g(\theta|y)$, based on the approximation $g(\alpha|\theta, y)$ of $f(\alpha|\theta, y)$ described in the previous section and on an approximation $g(\theta|y)$ of $f(\theta|y)$ described below.

Our approximation $g(\theta|y)$ is an n_θ-variate Student's t density with location parameter θ°, scale matrix Σ°, and degrees of freedom parameter equal to 30. We use two different methods to compute the location parameter and scale matrix. The first method involves computing these parameters by optimization. Just as $g(\alpha|\theta, y)$ is a close approximation of $f(\alpha|\theta, y)$, $\tilde{g}(\theta|y) \doteq f(\alpha, \theta, y)/g(\alpha|\theta, y)$ is a close approximation of $f(\theta|y)$. We thus take θ° as the maximizer of $\tilde{g}(\theta|y)$ and Σ° as the inverse of the negative Hessian of $\log\tilde{g}(\theta|y)$ at θ°. The second method involves generating a sample from $f(\theta|y)$ using a Random-Walk Metropolis posterior simulation. Then we take θ° as the sample mean and

Σ° as the sample variance.

We use the second method when the optimization method fails to work. The choice of method is model dependent. For some models, the optimization method works properly (See Djegnene and McCausland [4]), and for other it does not. In the case of ASCD-Models, we use the Random-Walk Metropolis posterior simulation to compute θ° and Σ°.

For the MCMC posterior simulation, we use an independence Metropolis-Hastings chain. The joint proposal $(\alpha^\star, \theta^\star)$ is accepted with probability

$$\pi(\theta^\star, \alpha^\star, \theta, \alpha) = \min\left[1, \frac{f(\theta^\star)f(\alpha^\star|\theta^\star)f(y|\theta^\star, \alpha^\star)}{f(\theta)f(\alpha|\theta)f(y|\theta, \alpha)} \frac{g(\theta)g(\alpha|\theta)}{g(\theta^\star)g(\alpha^\star|\theta^\star)}\right],$$

where (θ, α) is the current state and $(\theta^\star, \alpha^\star)$ is the proposal.

In the importance sampling posterior simulation, we reduce variance using a combination of quasi-random and pseudo random numbers. We construct M blocks of length S each, for a total of MS draws. S should be a power of two, which is convenient for Sobol quasi-random sequences. See Djegnene and McCausland [4] for details on how to draw θ.

2.3.3 On the correctness of posterior simulators

Our posterior simulators consist of core generic routines and model specific routines. The core routines compute coefficients of the polynomial approximations in the forward pass, and evaluate polynomials and draw and evaluate α from the approximate distribution in the backward pass. The model specific routines compute the log-densities $\psi_t(\alpha_t, \alpha_{t+1})$ and their derivatives $\psi_t^{(p,q)}(\alpha_t, \alpha_{t+1})$ for orders p and q up to certain values P and Q. Both have been extensively tested. The core routines have been jointly tested with model-specific code for the asymmetric volatility model with Student's t innovations, described in Djegnene and McCausland [4]. The model specific routines described in the present paper have been tested using built-in functions of the Gnu Scientific Libraries (GSL) for C and C++, comparing analytical and numerical derivatives.

In addition, we use artificial data simulations, as is common practice. For a fixed

parameter θ°, we generate a vector y° of artificial observations using (2.1). We then use this artificial data to estimate θ. Section 2.3.4 gives estimation results for different true parameter values.

None of these stringent exercises raise doubts about the correctness of the posterior simulators.

2.3.4 Results

We use artificial and real data to illustrate the Generalized HESSIAN procedure for the ASCD models. Artificial data are generated according to equation (2.11), for given values of the parameter vector. Real data are IBM transaction data.

We simulate artificial data from the ASCD-Exponential model for two values of the parameter vector, $\theta_1^{\circ} = (0.87, 0.96, 0.13, -0.01)$ and $\theta_2^{\circ} = (0.50, 0.91, 0.11, 0.01)$.[2] We generate artificial data from the ASCD-Gamma model, with parameter value $\theta_3^{\circ} = (0.87, 0.96, 0.13, 0.01, 0.90)$, and from the ASCD-Weibull model, with parameter value $\theta_4^{\circ} = (0.50, 0.91, 0.13, 0.02, 0.95)$. These choices are based on empirical results reported in Feng et al. [6] and McCausland [10]. For each vector θ_i°, $i = 1, \ldots, 4$, we generate a vector y_i° of 30000 observations using Equation (2.11). We then report posterior means of the parameters using the two posterior simulators described in Section 2.3.2.

Table 2.II and Table 2.III report estimation results for artificial data. The labels *HIS* and *HIM* indicate the importance sampling and independence Metropolis-Hastings chains, respectively. In both tables, the first column gives the posterior sample mean for the two chains. The second column is the posterior sample standard deviation and the third, the numerical standard error (NSE), a measure of the simulation precision of the posterior mean. The last column gives the relative numerical efficiency (RNE). The RNE is a variance ratio that tells how numerically precise the posterior mean estimate is relative to an estimate obtained from a hypothetical i.i.d chain. We compute the NSE and RNE of the importance sampling chain using results in Section 4.2 of Djegnene and McCausland [4]. We use the contributed coda library of the **R** software to compute those of

2. $\theta_i^{\circ} = (\bar{\alpha}, \phi, \sigma, \rho)$ for the ASCD-Exponential model and $\theta_i^{\circ} = (\bar{\alpha}, \phi, \sigma, \rho, \nu)$ for the ASCD-Gamma and the ASCD-Weibull models.

the independence Metropolis-Hastings method. This software computes the NSE using a time series method based on the estimated spectral density at frequency zero.

The posterior means are close to the true parameter values. For all elements of the $\theta_i^\circ, i = 1, \ldots, 4$, the true values fall between the posterior quantiles 2.5% and 97.5%. As outlined in Section 2.3.3, these results do not raise doubts about the correctness of the posterior simulators.

The RNEs reported in Table 2.II and Table 2.III are much higher than those of other posterior simulation methods for similar models and numbers of observations. Strickland et al. [18], who describe Bayesian simulation methods for SCD models, report RNE values ranging from 0.0046 to 0.0400 for the parameters their models have in common with ours. [3] The authors claim that this much lower relative efficiency is due to the complexity introduced by the Weibull and Gamma innovations. Using the same duration innovation distributions, we achieve RNE vales greater than one.

IBM data cover the period from November 1, 1990 to December 21, 1990, matching the period used in Feng et al. [6]. IBM data are available at Ruey Tsay's teaching resource webpage at *http ://faculty.chicagobooth.edu/ruey.tsay/teaching/fts3/*. The data records have fields for trading time and adjusted durations. The adjusted durations are computed as the difference between two consecutive trading times, then adjusted for diurnal patterns as described in Section 5.5 of Tsay [19].

The adjusted durations contain a lot of zeros, which raises some modeling issues. First, the density at zero for the Gamma and Weibull innovations (except for the Exponential special case) is equal to zero or infinity, depending on the shape parameter. Second, zero durations cannot be used to construct observations in a log-linear transformation of the original model. A common solution is to apply duration models conditional on strictly positive durations. [4] We too use the positive adjusted durations from the IBM data. We will call this the *positive duration* series.

Feng et al. [6] use fitted parameters to simulate a large sample of artificial durations.

3. See Table 3 of Strickland et al. [18]. The RNE is the inverse of the inefficiency factor reported in this table.

4. See Example 5.4 in Tsay [19] and Xu et al. [21]. Feng et al. [6] used log durations as the observed variable and we infer that they estimated their duration model conditional on positive durations.

Comparison of simulated and observed durations show that the former do not match well the left tail of the observed durations. In addition to estimating ASCD models for positive durations, we propose an alternative data treatment to address mis-fitting near zero. We construct a *transformed duration* series as follows. We take the maximum of one (second) and the difference between consecutive trading times and then adjust the resulting series for diurnal patterns as described in Section 5.5 of Tsay [19].

Table 2.IV summarizes the results for the ASCD models and the *positive duration* series. The labels *HIS* and *HIM* and the table columns have the same meaning as before, and the NSE and RNE values are computed in the same way.

We use the IBM data set to best compare our results with those of previous studies, such as that of Feng et al. [6]. This article used a linear transformation of the ASCD model given by Equation 2.11. Thus, we cannot directly compare values of the long-run mean. As for the other parameters, we can observe close similarities between values of the persistence parameters, state innovation standard deviations and shape parameters of the ASCD models with those reported in Table 4 of Feng et al. [6]. [5] The leverage effect parameter is also similar for the ASCD Exponential and ASCD Weibull models. We do not delete observations before the nominal opening of the market at 9 :50 am or after the nominal closing at 4 :00pm. This may explain the higher posterior standard deviation of the state innovation we report compared with Feng et al. [6].

Table 2.V summarizes the results of the ASCD models fitted with the *transformed durations*. We observe two main changes in the results reported. Values of the shape parameter in the ASCD-Gamma and ASCD-Weibull models are lower than those reported for the *positive durations*, see Table 2.IV. However, values of the standard deviation of the state innovation are greater. The low values of the shape parameter, along with a higher posterior standard deviation of the state innovation, imply a higher probability of durations near zero. The other parameters seem robust to the treatment of zero durations.

As Feng et al. [6], we focus on the ASCD-Weibull model to analyze the effect of this data treatment. Using the posterior means of parameters estimated using the *transformed durations*, we draw a sample from the marginal distribution of durations using Equation

5. Our parameter ϕ is equivalent to β in Feng et al. [6]

(2.11) with Weibull innovations. We do the same for the *positive duration* sample. Figure 2.1 shows histograms of model-simulated and observed durations. The distribution generated with the *transformed durations* fit better the left tail of the observed distribution than the one obtained with the *positive durations*. The support of the distribution has been truncated at ten to better visualize the histograms.

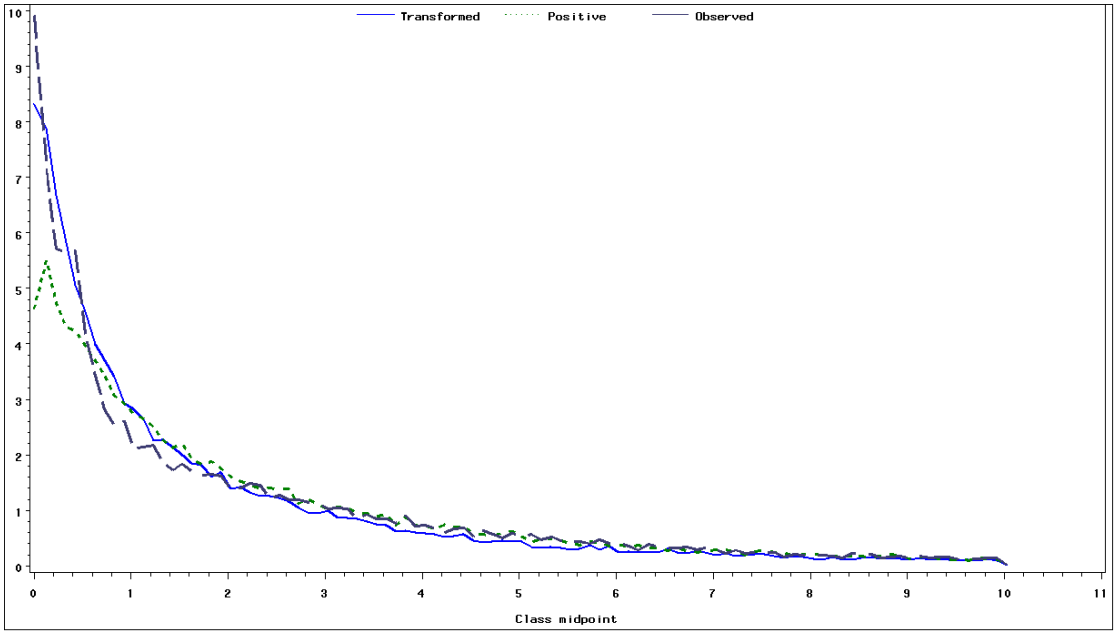

FIGURE 2.1 – Histograms of simulated duration data and observed duration data.

The exponential innovation does not have the freedom of shape of the Gamma and Weibull distribution. The small values introduced in the left tail of the observed data results in a much higher state innovation standard deviation and a lower long-run mean. Durations are also less persistent.

We focus on the relative numerical efficiency of the proposed simulators. The five main characteristics of the HESSIAN method described in Section 2.1 should produce efficient posterior simulators. The higher the RNE, the better. The independence Metropolis-Hastings chain of the ASCD-Exponential exhibits the lowest RNE, 12.68%. The independence Metropolis-Hastings chains of the ASCD-Gamma and ASCD-Weibull have greater RNEs than those of ASCD-Exponential. The lowest RNE for these two models is 24.48%, at least two times the one of ASCD-Exponential model. We compare with

Strickland et al. [18] as with the artificial data. Once again, this comparison outlined that the obtained RNEs are much higher than those produced by the posterior simulators described in this paper. Also, the importance chains allow a RNE greater than one. This is possible due to the variance reduction obtained by using randomized quasi-Monte Carlo.

Using $g(|y)g(\alpha|\theta, y)$ as an importance density, the mean of the (independent) importance weights is a simulation consistent and -unbiased estimator of the marginal likelihood. Our close approximation makes the variation in weights extremely small, which implies highly numerically efficient marginal likelihood estimation. For the *positive durations*, the log marginal likelihoods are -55351.1 for ASCD-Weibull and -55395.9 for ASCD-Gamma, with numerical standard errors of 0.0099 and 0.0089. The Bayes factor of $\exp(44.8)$ decisively favor the ASC-Weibull model. Similarly, for *transformed durations*, the marginal likelihoods are -54199.6 for ASCD-Weibull and -54386 for ASCD-Gamma, with numerical standard error of 0.0251 and 0.0150. The Bayes factor of $exp(186.4)$ favor the ASCD-Weibull model. These results are consistent with Feng et al. [6]that found that the ASCD-Weibull is better in term of marginal densities comparison and in-sample forecast performance.

2.4 Conclusion

We have provided new methods for state smoothing and parameter estimation for non-linear and non-Gaussian state space models. Posterior simulations for real and artificial data show that this procedure is highly numerical efficient. We achieve this numerical efficiency by providing a very close approximation $g(\alpha|y, \theta)$ of the target density $f(\alpha|y, \theta)$. The approximate density can be used as an importance density for importance sampling or a proposal density for Markov chain Monte Carlo posterior simulations.

The scope of application goes far beyond the ASCD-Exponential model used for illustration. We require the target density to have a unique mode and be log-differentiable, which still leaves a rich class of applicable non-linear and non-Gaussian state space models. Application to a new model only requires routines to compute partial derivatives of the log conditional densities $\log f(y_t, \alpha_{t+1}|\alpha_t)$, with respect to α_t and α_{t+1}. We do

not require analytic expressions for these derivatives. We may used numerical derivatives or use Faa-Di-Bruno's rule for combining multiple derivatives of simple functions to compute multiple derivatives of compound functions.

We now require α_t to be univariate. We are now considering state smoothing and parameter estimation for the case of multivariate α_t. Also, we are working on approximations of filtering densities, useful for sequential learning.

Notation	Description
$\psi_t(\alpha_t, \alpha_{t+1})$	$\log f(\alpha_{t+1}, y_t \vert \alpha_t)$
$\psi_t^{p,q}(\alpha_t, \alpha_{t+1})$	derivatives of $\psi_t(\alpha_t, \alpha_{t+1})$ with respect to α_t and α_{t+1} at orders p and q.
$\psi_n(\alpha_n)$	$\log f(y_n \vert \alpha_n)$
$\psi_n^p(\alpha_n)$	p'th derivative of $\psi_n(\alpha_n)$ with respect to α_n
$a = (a_1, \ldots, a_n)$	mode of $\log f(\alpha \vert y)$
Σ_t	Var$(\alpha_t \vert \alpha_{t+1}, y)$ for the 1st reference distribution
$(a_{1\vert t+1}(\alpha_{t+1}), \ldots, a_{t\vert t+1}(\alpha_{t+1}))$	mode of the conditional density $f(\alpha_1, \ldots, \alpha_t \vert \alpha_{t+1}, y)$
$a_t^{(r)}$, $r = 1, \ldots, R$	r'th derivative of $a_{t\vert t+1}(\alpha_{t+1})$ at $\alpha_{t+1} = a_{t+1}$
$\Sigma_{t\vert t+1}(\alpha_{t+1})$	Var$(\alpha_t \vert \alpha_{t+1}, y)$ for the 2nd reference distribution
$s_{t\vert t+1}(\alpha_{t+1})$	$\log \Sigma_{t\vert t+1}$
$s_t^{(r)}$, $r = 1, \ldots, R-1$	r'th derivatives of $s_{t\vert t+1}(\alpha_{t+1})$ at $\alpha_{t+1} = a_{t+1}$.
$b_{t\vert t+1}(\alpha_{t+1})$	mode of the conditional density $f(\alpha_t \vert \alpha_{t+1}, y)$
$b_t, b_t^{(1)}, b_t^{(2)}, b_t^{(3)}$	value and first three derivatives of $b_{t\vert t+1}(\alpha_{t+1})$ at $\alpha_{t+1} = a_{t+1}$
b_n	mode of the conditional density $f(\alpha_n \vert y)$
$B_t, B_t^{(1)}, B_t^{(2)}, B_t^{(3)}, B_n$	approximations of $b_t, b_t^{(1)}, b_t^{(2)}, b_t^{(3)}$ and b_n
$\mu_{t\vert t+1}(\alpha_{t+1})$	$E[\alpha_t \vert \alpha_{t+1}, y]$
$\mu_t, \mu_t^{(1)}, \mu_t^{(2)}$	value and first two derivatives of $\mu_{t\vert t+1}(\alpha_{t+1})$ at $\alpha_{t+1} = a_{t+1}$
$M_t, M_t^{(1)}, M_t^{(2)}$	approximations of μ_t, $\mu_t^{(1)}$ and $\mu_t^{(2)}$
$H_t(\alpha_t; \alpha_{t+1})$,	approximation of the derivative of $\log f(\alpha_t \vert \alpha_{t+1}, y)$ with respect to α_t
$H_n(\alpha_n)$,	approximation of the derivative of $\log f(\alpha_n \vert y)$ with respect to α_n

TABLE 2.I – Main notation used in the paper

Parameters	Mean	Std	NSE	RNE
ASCD-Exponential				
$\theta_1^\circ = (0.87, 0.96, 0.13, -0.01)$				
$\bar{\alpha}$:HIS	0.7867	0.0601	6.3373e-4	0.9320
$\bar{\alpha}$:HIM	0.7868	0.0602	8.7035e-4	0.3740
ϕ :HIS	0.9605	0.0028	3.1442e-5	0.8614
ϕ :HIM	0.9605	0.0029	4.5081e-5	0.3185
σ :HIS	0.1325	0.0049	5.49326e-5	0.8574
σ :HIM	0.1325	0.0049	6.6018e-5	0.4221
ρ :HIS	-0.01250	0.0036	3.6266e-5	1.0471
ρ :HIM	-0.01250	0.0036	5.2040e-5	0.3732
$\theta_2^\circ = (0.50, 0.91, 0.11, 0.01)$				
$\bar{\alpha}$:HIS	0.4742	0.0259	2.6803e-4	0.9089
$\bar{\alpha}$:HIM	0.4734	0.0262	3.9116e-4	0.3509
ϕ :HIS	0.9086	0.0096	1.0067e-4	0.9365
ϕ :HIM	0.9085	0.0096	1.2617e-4	0.4485
σ :HIS	0.1119	0.0080	8.5432e-5	0.8719
σ :HIM	0.1119	0.0079	1.0021e-4	0.4884
ρ :HIS	0.0088	0.0035	3.5589e-5	0.9614
ρ :HIM	0.0087	0.0036	5.9577e-5	0.2820

TABLE 2.II – ASCD model estimation results for artificial data using independence Metropolis-Hastings and importance sampling

Parameters	Mean	Std	NSE	RNE
$\theta_3^\circ = (0.87, 0.96, 0.13, 0.01, 0.90)$				
ASCD-Gamma				
$\bar{\alpha}$:HIS	0.8997	0.0894	1.1122e-3	0.9128
$\bar{\alpha}$:HIM	0.8978	0.0894	1.2638e-3	0.3906
ϕ :HIS	0.9663	0.0026	3.0271e-5	1.1299
ϕ :HIM	0.9662	0.0026	4.3124e-5	0.2781
σ :HIS	0.1239	0.0049	6.1714e-5	0.9653
σ :HIM	0.1241	0.0049	7.1299e-5	0.3687
ρ :HIS	0.0112	0.0035	4.2910e-5	0.9927
ρ :HIM	0.0112	0.0035	5.4795e-5	0.3199
ν :HIS	0.9047	0.0073	8.8113e-5	0.9509
ν :HIM	0.9049	0.0073	9.7956e-5	0.4340
ASCD-Weibull				
$\bar{\alpha}$:HIS	0.8921	0.0653	7.8648e-4	0.7929
$\bar{\alpha}$:HIM	0.8917	0.0653	1.1185e-03	0.2662
ϕ :HIS	0.9596	0.0030	3.3324e-5	0.9946
ϕ :HIM	0.9596	0.0031	4.2381e-5	0.4181
σ :HIS	0.1332	0.0057	6.3643e-5	0.9063
σ :HIM	0.1332	0.0056	8.2801e-5	0.3580
ρ :HIS	0.0096	0.0038	4.2521e-5	0.87884
ρ :HIM	0.0096	0.0037	6.1353e-5	0.2889
ν :HIS	0.9000	0.0048	5.0637e-5	0.9263
ν :HIM	0.9000	0.0048	7.3856e-5	0.3362
$\theta_4^\circ = (0.50, 0.91, 0.13, 0.02, 0.95)$				
ASCD-Gamma				
$\bar{\alpha}$:HIS	0.5128	0.0406	6.0296e-4	0.9649
$\bar{\alpha}$:HIM	0.5124	0.0402	9.7184e-4	0.1334
ϕ :HIS	0.9155	0.0070	9.4880e-5	0.9513
ϕ :HIM	0.9156	0.0069	1.4224e-4	0.1863
σ :HIS	0.1276	0.0075	9.9498e-5	1.0589
σ :HIM	0.1275	0.0076	1.7078e-4	0.1547
ρ :HIS	0.0210	0.0038	4.7404e-5	0.9182
ρ :HIM	0.0209	0.0037	6.6079e-5	0.2511
ν :HIS	0.9443	0.0079	9.0268e-5	1.2268
ν :HIM	0.9443	0.0078	1.7315e-4	0.1605
ASCD-Weibull				
$\bar{\alpha}$:HIS	0.4513	0.0335	3.6064e-4	0.9558
$\bar{\alpha}$:HIM	0.4518	0.0334	4.2367e-4	0.4870
ϕ :HIS	0.9191	0.0070	7.0012e-5	1.1029
ϕ :HIM	0.9191	0.0069	1.1159e-4	0.2960
σ :HIS	0.1262	0.0079	7.6962e-5	1.1515
σ :HIM	0.1262	0.0078	1.2169e-4	0.3213
ρ :HIS	0.0124	0.0039	4.0723e-5	0.9606
ρ :HIM	0.0125	0.0039	5.1425e-5	0.4551
ν :HIS	0.9546	0.0053	5.0605e-5	1.1667
ν :HIM	0.9546	0.0053	7.4828e-5	0.3968

	Mean	Std	NSE	RNE
ASCD-Exponential Model				
$\bar{\alpha}$:HIS	0.8056	0.0666	5.6186e-4	1.3579
$\bar{\alpha}$:HIM	0.8050	0.0669	1.1091e-3	0.2845
ϕ :HIS	0.9556	0.0040	3.5977e-5	1.1703
ϕ :HIM	0.9556	0.0040	6.0845e-5	0.3400
σ :HIS	0.1581	0.0079	6.6559e-5	1.3535
σ :HIM	0.1580	0.0078	1.1723e-4	0.3497
ρ :HIS	-0.0105	0.0046	3.8138e-5	1.4589
ρ :HIM	-0.0106	0.0047	7.9730e-5	0.2698
ASCD-Gamma				
$\bar{\alpha}$:HIS	1.0008	0.0961	1.3425e-3	0.7502
$\bar{\alpha}$:HIM	1.0001	0.0964	1.8034e-3	0.2230
ϕ :HIS	0.9636	0.0038	4.5527e-5	1.08517
ϕ :HIM	0.9636	0.0038	7.9354e-5	0.1802
σ :HIS	0.1384	0.0082	9.4210e-5	1.1154
σ :HIM	0.1383	0.0081	1.5171e-4	0.2229
ρ :HIS	-0.0000	0.0048	7.0586e-5	0.7436
ρ :HIM	-0.0001	0.0048	8.0025e-5	0.2757
ν :HIS	0.9600	0.0092	1.3037e-4	0.7673
ν :HIM	0.9601	0.0092	2.3659e-4	0.1187
ASCD-Weibull				
$\bar{\alpha}$:HIS	1.2160	0.1117	2.8008e-3	0.9445
$\bar{\alpha}$:HIM	1.2104	0.1044	1.4656e-3	0.3961
ϕ :HIS	0.9729	0.0030	6.3976e-5	0.911438
ϕ :HIM	0.9727	0.0028	4.9166e-5	0.2563
σ :HIS	0.1126	0.0067	1.2770e-4	0.9277
σ :HIM	0.1129	0.0065	9.9064e-5	0.3354
ρ :HIS	0.0123	0.0044	6.8877e-5	0.9585
ρ :HIM	0.0122	0.0043	5.4893e-5	0.4825
ν :HIS	0.9396	0.0054	7.5690e-5	1.0168
ν :HIM	0.9397	0.0053	8.0868e-5	0.3371

TABLE 2.IV – ASCD model estimation results for IBM duration data using independence Metropolis-Hastings and importance sampling

Parameters	Mean	Std	NSE	RNE
ASCD-Exponential				
$\bar{\alpha}$:HIS	0.2107	0.0365	5.3651e-4	1.1790
$\bar{\alpha}$:HIM	0.2121	0.0371	9.2005e-4	0.1268
ϕ :HIS	0.8492	0.0092	1.5727e-4	0.9202
ϕ :HIM	0.8492	0.0091	1.9864e-4	0.1643
σ :HIS	0.4030	0.0153	2.5235e-4	1.0157
σ :HIM	0.4032	0.0152	3.5252e-4	0.1455
ρ :HIS	-0.0986	0.0070	1.0864e-4	1.0344
ρ :HIM	-0.0984	0.0071	1.9061e-4	0.1070
ASCD-Gamma				
$\bar{\alpha}$:HIS	1.0431	0.1323	1.8397e-3	0.8255
$\bar{\alpha}$:HIM	1.0468	0.1305	2.1590e-3	0.2852
ϕ :HIS	0.9663	0.0034	4.5226e-5	0.8606
ϕ :HIM	0.9662	0.0033	5.8228e-5	0.2565
σ :HIS	0.1546	0.0083	1.1053e-4	0.8406
σ :HIM	0.1547	0.0082	1.3653e-4	0.2833
ρ :HIS	0.0026	0.0041	5.4987e-5	0.8785
ρ :HIM	0.0027	0.0041	6.7975e-5	0.2791
ν :HIS	0.8009	0.0070	9.0401e-5	0.8404
ν :HIM	0.8009	0.0069	1.2165e-4	0.2536
ASCD-Weibull				
$\bar{\alpha}$:HIS	1.0964	0.1391	1.4642e-3	1.1589
$\bar{\alpha}$:HIM	1.0958	0.1398	2.4383e-3	0.2568
ϕ :HIS	0.9782	0.0023	2.5198e-5	1.0146
ϕ :HIM	0.9782	0.0023	3.6880e-5	0.3044
σ :HIS	0.1155	0.0063	6.7147e-5	1.0026
σ :HIM	0.1155	0.0062	1.1144e-4	0.2448
ρ :HIS	0.0139	0.0039	3.9974e-5	1.1953
ρ :HIM	0.0139	0.0038	6.0898e-5	0.3114
ν :HIS	0.8348	0.0044	4.3241e-5	1.1054
ν :HIM	0.8348	0.0044	6.9128e-5	0.3118

TABLE 2.V – ASCD model estimation results for IBM data using independence Metropolis-Hastings and importance sampling

CHAPITRE 3

COOPERATIVES' REPUTATION AND ENDOGENOUS MEMBERSHIP IN A MIXED DUOPSONY

Abstract

This article studies competition on the price paid to farmers between a farmer-owned cooperative (FOC) and an investor-owned firm (IOF). Using Hotelling's spatial model, it is shown that a cooperative's reputation, managers' incompetency and financial market tightness, are key factors explaining the difficult growth of cooperatives. Cooperatives play an active role in raising financial resources in order to pay at least the delivery price at storage. After storage, there is uncertainty concerning the farmer's residual claim, which may be diverted. The trust relationship between the farmer and the cooperative, or the cooperative's reputation, then becomes a key factor in explaining the farmer's decision to patronize the cooperative. The delivery price, which also influences the farmer's decision, depends on how easy it is for the cooperative to raise financial resources. Financial market tightness and cooperative incompetency play an important role at this level. It is shown that a good reputation and more competent managers raise the cooperatives' market share and force the investor-owned firm to increase the price it pays to the farmers. Those conclusions fit the empirical observations better than the models of Sexton [11], Albæk and Schultz [1] and Karantininis and Zago [9] on the same subject.

Keywords : Cooperatives, Duopsony, Endogenous Membership, Reputation.

JEL Classification : Q13.

3.1 Introduction

The liberalization of agricultural markets in the 1990s in most sub-Saharan countries, especially in Côte d'Ivoire, introduced investor-owned firms (IOFs) in the local cocoa and coffee markets. Within a few years, these developed into oligopsonistic structures. This outcome contrasts with the expected result of market deregulation - enhanced poor rural farmer revenue. Given this market failure, farmer-owned cooperatives (FOCs) were regarded as an alternative way for small farmers to improve their welfare. However, since the liberalization and promotion of collective marketing, no noteworthy progress has been made by cooperatives either in farm gate and export market share or in manufacturing. In 2007, the cooperative market share at farm gate was less than 7.57%.

These national figures hide some disparities at the regional level. Figure 3.1 highlights the cooperative membership rate distribution and the social homogeneity distribution according to production areas. [1] The social homogeneity index is computed as the ratio of the natives to the total population living in the region expressed as a percentage.

At first what stands out is the deep contrast between the East and the Southwest cooperative membership rate. While most of the East's producers are cooperative members, less than a quarter of the Southwest's producers belong to cooperatives. We have intermediate situations in the Midwest and West regions. The second significant fact which emerges is the positive correlation between membership rate and social homogeneity. These observations challenge us on how social environment affects the development of cooperative membership.

Repeated interaction between the members of a small homogenous society where individual behavior can be perfectly recorded may sustain trust and reputation as noted by Berg et al. [4]. Therefore, we choose to analyze the relationship between social environment and cooperative membership rates through these two intermediary variables : reputation and trust.

James and Sykuta [7], based on a survey of Missouri corn and soybean farmers, argue that trust, perception of honesty, and competence are key factors explaining the choice

1. Data source : "Production et offre du cacao et du café en Côte d'Ivoire", *International Institute of Agriculture, 2002*.

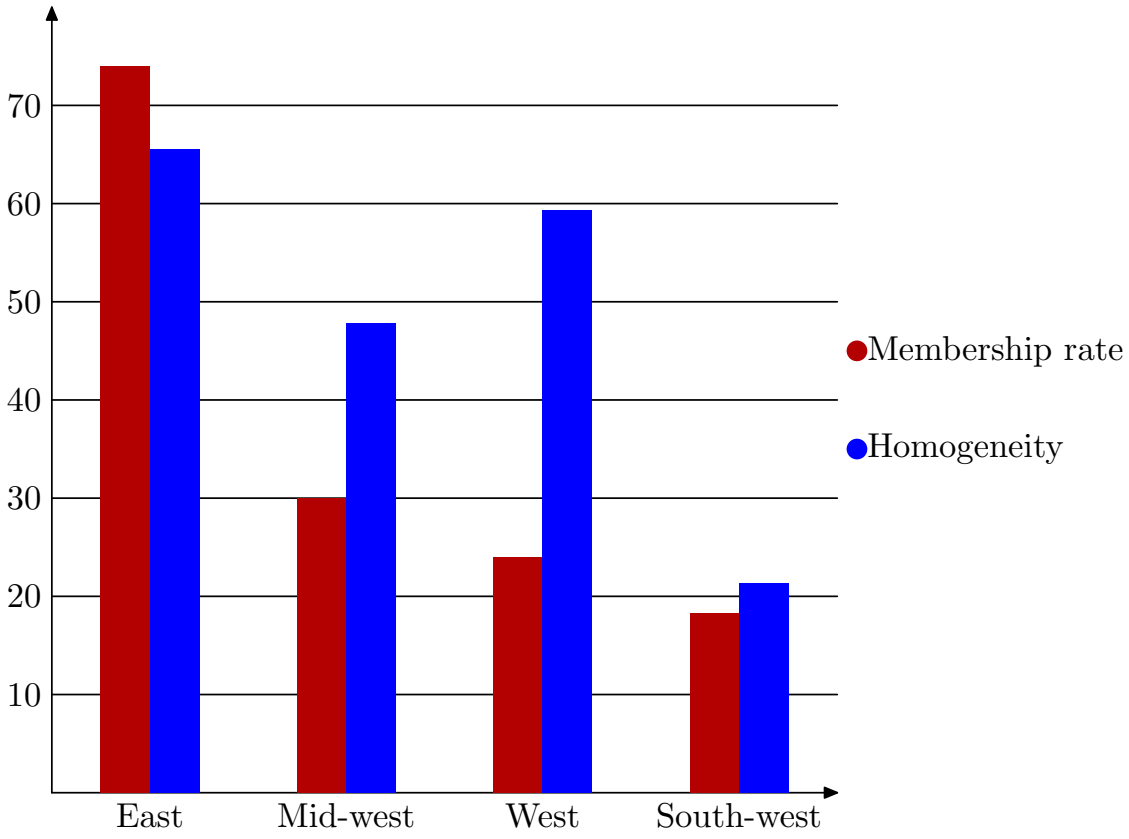

FIGURE 3.1 – **Cooperatives' membership rate and social homogeneity by production area, Ivory Coast, 2002**

of farmers to market through cooperatives rather than IOFs. James and Sykuta [8] try to find evidence on how trust emerges in agricultural cooperatives by linking the level of perceived trust in an agricultural cooperative with its organizational characteristic. They worry that there is not a well-developed theory laying out precisely how and why trust in cooperatives emerges along with the why and wherefore of its maintenance.

Trust is still an object of debate and can be conceptualized in different ways Hardin [6]. We shall consider here the approach of Gambetta [5] who defines trust as "Ěa particular level of the subjective probability with which an agent assesses that another agent or group of agents will perform a particular action, both before he can monitor such action and in a context in which it affects his own action". Hence as emphasized by Hardin [6], trust is a three component relationship : A trust B to do X. So, farmers trust

FOCs to do X, where X may be, among others, not diverting the residual claimant or be competent to recover the full output price.

Using a price-competition model with incomplete information between an IOF and a FOC, we show that reputation, competence, and financial conditions can be an impediment to cooperative growth. In our model, trust and reputation distributions are primitives that help explain cooperative membership rate distribution. While building on the economic structure of Sexton's model Sexton [11], we replace the spatial dimension with trust. Furthermore, our model differs from those of Sexton [11], Albæk and Schultz [1], Karantininis and Zago [9], among others, by the fact that the cooperative is a non-neutral intermediary. The cooperative proposes a schedule of payments that includes two parts : a delivery price and a dividend paid after the raw product is processed and sold. With incomplete information on the cooperative type, this dividend becomes risky and the cooperative's reputation is now a key factor in the farmer's decision. We show that good reputation not only forces the IOF to price high, but it also reduces its market share. Conversely, tight financial conditions push down the IOF spot price and raise the FOCs' market share.

To some extent, our results generalize those of Sexton [11], Albæk and Schultz [1] and Karantininis and Zago [9]. Assuming perfect information and a trustworthy cooperative that transfers the total net output price to farmers, these authors claim the absolute advantage of FOCs over IOFs. According to them, price competition between a FOC and an IOF will end up with the FOC being the sole buyer in the market. This result is a special case of ours.

Trust and reputation, like any economic asset, need some initial stock and a constant flow of investment, barring which they crumble and bring down the group effort. Thus, in a region where these values are not commonly shared and invested in, making a cooperative work seems to be at best difficult. Arrow [2] was aptly inspired when concluding an essay on gifts and exchanges with the following contention : "It can plausibly be argued that much of the economic backwardness in the world can be explained by the lack of mutual confidence."

This article is organized as follows : in Section 2, we introduce the model ; Sec-

tion 3 describes the equilibrium and makes some comparative static analysis ; Section 4 concludes with some final remarks.

3.2 The model

We consider an agricultural commodity market in which IOFs and FOCs purchase a raw commodity from a large number of small farmers. The processed product is sold on a competitive market. A cooperative schedule of payment generally involves two components : the first component is a delivery price to stock the physical product and the second component is a dividend paid after the processed product is sold. We can reasonably argue that the delivery price is secure while the dividend is at risk. The latter may be diverted, a behavior deeply grounded in developing countries. Banerjee et al. [3] describes how wealthy members of sugar cooperatives in Maharashtra (India) siphon off the cooperatives' retained earnings. Moreover, the FOC's incompetency may not allow it to recover the full output price of the processed product and thus it cannot pay the full dividend. Consequently, we assume that there are two types of cooperatives : trustworthy cooperatives that always pay back the dividend and untrustworthy cooperatives that may not pay back the dividend. [2]

The behavior of untrustworthy cooperatives may be shaped by the social environment in which they operate. We consider that each farmer's valuation of the residual claimant risk can be decomposed into two parts : the first is the risk involving the cooperative type and the second is the risk of the untrustworthy cooperative not paying back the dividend. The first component of the risk is common to all farmers and represents prior beliefs about the cooperative's type. The second component is idiosyncratic and can be used to assess the perception of property-rights security in the social environment.

To analyze the outcome of price competition between IOF and FOC in such context, we consider a game that borrows its economic structure from Sexton [11], with the exception that the spatial dimension is replaced by the distribution of farmers according to their trust in the property-rights security. We consider a two-stage game that involves

2. Farmers' trust is based on either honesty or competency.

two processors, an IOF and a FOC, and a continuum of farmers of mass unity.

In the first stage, the FOC and the IOF compete in price. They simultaneously make price proposals to farmers for delivery of the primary commodity. The commodity is processed and sold in a competitive market. We assume a constant unit processing cost and normalize the net output price to one. There is imperfect information on the FOC : it may be trustworthy (t) or untrustworthy (u). A FOC of type $\tau \in \{t, u\}$ commits to a certain level of effort, $e^\tau \in [0, 1]$, in order to raise the financial resources necessary to pay the delivery price. We assume that this level of effort represents the delivery price offered by the cooperative. For a given delivery price, e^τ, the difference $1 - e^\tau$ represents the implied dividend. The IOF's contract consists of a delivery price $\omega \in [0, 1]$.

At the second stage, the farmers make their decisions as to how much to supply to each of the processors. These farmers are distributed according to the cumulative $F(\alpha), \alpha \in [0, 1]$ where an $\alpha-$type farmer assigns a probability α on the event that the untrustworthy cooperative will not pay back the dividend. For tractability, we assume for the rest of the article that this distribution is an uniform distribution on the interval $[0, 1]$. Farmers cannot distinguish between the two types of FOC. They observe only the delivery price, e, of the FOC and the delivery price of the IOF, ω. So, given the processors' proposal (e, ω), an $\alpha-$type farmer chooses a production plan $\mathbf{q}_\alpha = (q^{iof}(e, \omega; \alpha), q^{foc}(e, \omega; \alpha))$, where the components are respectively delivery to the IOF and delivery to the FOC.

Farmers' payoffs are computed using their posterior beliefs on the cooperative's type. The cooperative's reputation is farmers' posterior beliefs distribution. Let β be the posterior beliefs of an $\alpha-$type farmer that the FOC is of type t, for the proposed level of effort e.[3] Assuming that an $\alpha-$type farmer is risk-neutral, he evaluates his payoff using the expected price from the cooperative. So, his expected profit is given by

$$\Pi(e, \omega, \mathbf{q}_\alpha; \alpha, \beta) = [\beta + (1 - \beta)(e + \alpha(1 - e))]\, q^{foc} + \omega q^{iof} - c(q^{foc} + q^{iof}) \quad (3.1)$$

3. In fact a farmer's posterior beliefs should depend on his type and on the cooperative's proposal. Farmers' share the same information set, so we can drop the dependence on the farmer's type. In the equilibrium considered later, this belief is a discrete-valued function, assumed constant in a neighborhood of the equilibrium proposal. So, for ease of notation, we drop the dependance on e and then view β as the anticipated equilibrium beliefs.

The compensation of the cooperative board for the efforts made to manage the cooperative and to raise the necessary financial resources may be of either a monetary or a non-monetary form. For simplicity we will assume that it takes the form a non-monetary satisfaction from helping farmers obtain a credible alternative to the IOF. Furthermore, we assume that this non-monetary compensation amounts to the farmers' cumulative profit. The effort is costly and the two types of cooperatives evaluate it differently. For the same level of effort, the resulting cost for the untrustworthy type is higher for than the trustworthy type. In addition, it is more painful for the untrustworthy type to increase marginally its level of effort :

$$c(e,u) \geq c(e,t), \qquad \frac{\partial c(e,u)}{\partial e} \geq \frac{\partial c(e,t)}{\partial e} \geq 0. \tag{3.2}$$

Hence the payoff of a cooperative of type τ is equal to the expected difference between its compensation and the cost of the effort :

$$\Pi^{\tau}(e^{\tau},\omega;\beta) = \int_{\Delta_{\tau}} \left[\Pi(e^{\tau},\omega,\mathbf{q}_{\alpha};\alpha) - c(e^{\tau},\tau)\right] dF(\alpha), \tag{3.3}$$

with $\tau \in \{t,u\}$ and where $\Delta_{\tau} \subseteq [0,1]$ is the set of farmers who deliver to the cooperative. [4]

Using the normal representation of the Bayesian game described at the first stage, the IOF's payoff depends on the action profile (e^t, e^u, ω). Let γ be the IOF's posterior belief that the cooperative is of the trustworthy type. Then its payoff is equal to its expected profit :

$$\Pi^{iof}(e^t,e^u,\omega;\gamma) = (1-\omega)\int_0^1 \left[\gamma q^{iof}(e^t,\omega;\alpha) + (1-\gamma)q^{iof}(e^u,\omega;\alpha)\right] dF(\alpha). \tag{3.4}$$

We search for an equilibrium in pure strategies and borrow from the sequential equilibrium concept to characterize the rational and beliefs consistent outcome of the game. Hence, an assessment $(\gamma,\beta,e^t,e^u,\omega,(\mathbf{q}_{\alpha})_{\alpha\in[0,1]})$ is a pure strategy sequential equili-

4. The cooperative's payoff depends conditionally on the anticipated posterior beliefs through farmers' response functions.

brium if :

1. Given the farmers' strategies $\mathbf{q}_\alpha$ for each $\alpha-$type, given the IOF contract ω, proposing e^τ maximizes the payoff of the cooperative of type τ, $\tau \in \{t, u\}$;
2. The IOF posterior beliefs satisfy Baye's rule :$\gamma = \theta$;
3. Given the farmers' production plans $\mathbf{q}_\alpha$ for each $\alpha-$type, given the cooperative's contract $\{e^t, e^u\}$, the IOF contract ω maximizes its payoff, given its beliefs γ ;
4. Farmers' posterior beliefs satisfy Baye's rule :
   (a) $\beta \in [0, 1]$ for all $e \in [0, 1]$.
   (b) If $e^t \neq e^u$, then $\beta = 1$ if $e^\tau = e^t$, and $\beta = 0$ if $e^\tau = e^u$.
   (c) If $e^t = e^u$, then $\beta = \theta$;
5. For every couple of contracts (e, ω), the $\alpha-$type farmer production plan $\mathbf{q}_\alpha$ maximizes its payoff, given its beliefs β.

If the two types of cooperatives propose different levels of effort at equilibrium, then observing the trustworthy (resp. untrustworthy) proposal, each farmer will infer that he faces the trustworthy (untrustworthy) cooperative. Such an equilibrium is termed a separating equilibrium. If the two types of cooperatives make the same proposal at equilibrium, then by observing this proposal, farmers cannot distinguish between them. Their beliefs remain unchanged and equal to their prior belief. Such an equilibrium is termed a pooling equilibrium. In solving this game we emphasize the pooling equilibrium.

3.3 Characterization of the equilibrium

The game is solved backwards. Therefore we consider first the farmers' problem and then the processors' programs.

3.3.1 The farmer's production and patronizing decision

An $\alpha-$type farmer chooses the production plan (q^{iof}, q^{foc}) that maximizes his expected payoff, given by (3.1). Thus his program can be written :

$$\max_{q^{iof},q^{foc},q} \left\{ (1-(1-e)(1-\beta)(1-\alpha))q^{foc} + \omega q^{iof} - \frac{b}{2}q^2 \right\}$$
$$s.t \quad q^{foc}, q^{iof}, q \geq 0, \quad q^{foc} + q^{iof} = q.$$

Let $e_\alpha = 1-(1-e)(1-\beta)(1-\alpha)$. The quantity e_α is still the price expected by an $\alpha-$type farmer from the cooperative, but expressed here as the difference between the net output price and the expected loss from the untrustworthy cooperative. The production plan of an $\alpha-$type farmer is driven by the relative position of e_α and ω. Thus, we have

$$\begin{aligned}
&q^{foc} = \frac{e_\alpha}{b} \text{ and } q^{iof} = 0 && \text{if } e_\alpha > \omega, \\
&q^{foc} \in [0, \frac{\omega}{b}] \text{ and } q^{iof} = \frac{\omega}{b} - q^{foc} && \text{if } e_\alpha = \omega, \\
&q^{foc} = 0 \text{ and } q^{iof} = \frac{\omega}{b} && \text{if } e_\alpha < \omega.
\end{aligned} \tag{3.5}$$

An $\alpha-$type farmer compares the expected price from the cooperative with the spot price of the IOF. If the cooperative's expected price is strictly higher than that of the IOF, the farmer delivers his entire production to the cooperative. Conversely, if the IOF delivery price is strictly higher than that of the cooperative, the entire production goes to the IOF. If both contracts are equivalent, the farmer shares indifferently his production between the two processors. This decision rule determines, for every pair of contracts, the set of farmers patronizing the cooperative. An $\alpha-$type farmer patronizes the cooperative if and only if $\alpha >= \alpha^\star(e,\omega)$ with :

$$\begin{aligned}
&\alpha^\star(e,\omega) = 0 && \text{if } \beta = 1 \text{ or } e = 1 \\
&\alpha^\star(e,\omega) = \frac{\omega - e}{1-e} && \text{if } \beta = 0 \text{ and } e < 1 \\
&\alpha^\star(e,\omega) = 1 - \frac{1-\omega}{(1-\beta)(1-e)} && \text{if } \beta \in (0,1) \text{ and } e < 1
\end{aligned} \tag{3.6}$$

If farmers believe that the chosen cooperative is of the trustworthy type or if the cooperative is able to raise sufficient resources to offer a delivery price equal to the local net price, $e = 1$, then it will get the entire market. When the farmers believe that the chosen cooperative is of the untrustworthy type, only those with a high level of trust in the property-rights security deliver to the cooperative. The latter are always part of the farmers who deliver to the cooperative when we cannot distinguish between the two types.

The threshold $\alpha^\star(e, \omega)$ can be viewed as the ratio of the opportunity cost of patronizing the cooperative under the hypothesis that it is trustworthy versus the hypothesis that it is untrustworthy. In order to have a mixed duopsony, the cooperative has to play a "pacemaker role"LeVay [10] for the IOF. From equation (3.6), third line, a necessary condition for $\alpha^\star(e, \omega)$ to be nonnegative is that the price of the IOF cannot be less than the price of the cooperative under the pessimistic hypothesis that the untrustworthy cooperative will never pay back the dividend : $\omega \geq e + \beta(1 - e)$.

3.3.2 The cooperative pricing decision

Cooperatives are distinguished by their cost function, which is an increasing convex function of effort. Let us assume the following functional form $c(e, \tau) = a_\tau e^2$. Given the farmers' production plans, $(\mathbf{q}_\alpha)_{\alpha \in [0,1]}$, and the IOF delivery price, ω, a FOC chooses a level of effort that maximizes its payoff. When it anticipates that it will be treated as trustworthy $(\beta = 1)$, its payoff function is equal to :

$$\Pi^\tau(e, \omega; 1) = \frac{1}{2b} - a_\tau e^2$$

This payoff is a decreasing function of e and its maximum is attained for $e^\tau(\omega; 1) = 0$.

When the FOC anticipates that it will be treated as not trustworthy ($\beta < 1$), its payoff

function becomes :

$$\Pi^{\tau}(e,\omega;\beta) = \frac{1}{2b}\int_{\alpha^{\star}}^{1}\left(1-(1-e)(1-\beta)(1-\alpha)\right)^2 d\alpha - a_{\tau}e^2(1-\alpha^{\star})$$
$$= \frac{1-\omega}{6b(1-\beta)}\left[\frac{(1+\omega+\omega^2)-(6a_t b)e^2}{1-e}\right].$$

Assuming an interior solution, the level of effort that maximizes this payoff is the solution of the following first-order condition :

$$e^{\star}(2-e^{\star}) = \frac{1+\omega+\omega^2}{6ba_{\tau}}. \tag{3.7}$$

Assuming $a_{\tau}b \geq \frac{1}{2}$, the polynomial (3.7) has two real roots, given by : [5]

$$e_1^{\star} = 1+\sqrt{1-\frac{1+\omega+\omega^2}{6ba_{\tau}}} \quad e_2^{\star} = 1-\sqrt{1-\frac{1+\omega+\omega^2}{6ba_{\tau}}},$$

In order to satisfy the constraint $e \in [0,1]$, the only admissible candidate is $e^{\star}(\omega) = e_2^{\star}$.

The second derivative with respect to e of the cooperative's payoff, evaluated at $e^{\star}(\omega)$ is :

$$\left.\frac{\partial^2\Pi^t(e,\omega)}{\partial e^2}\right|_{e=e^{\star}} = \frac{1-\omega}{6b(1-\beta)}\frac{2(6ba_{\tau})(e^{\star}-1)(1-e^{\star})}{(1-e^{\star})^3} < 0.$$

The candidate $e^{\star}$ is therefore a maximum and the optimal proposal of the cooperative of type τ, for a given delivery price ω of the investor owned firm, can be written :

$$e^{\star\tau}(\omega;\beta) = e^{\star} = 1-\sqrt{1-\frac{1+\omega+\omega^2}{6ba_{\tau}}}. \tag{3.8}$$

The FOC's best response function (3.8) is a decreasing function of a_{τ}. Hence, since $a_u > a_t$, for the same anticipated equilibrium the untrustworthy FOC makes less effort than the trustworthy FOC. We can interpret a_{τ} as an index of the tightness of the financial market. Hence an increase in a_{τ} reflects a tightening of the financial market, making it more difficult to raise the resources necessary to fund the purchase of the primary

5. Note that since $a_u > a_t$, $a_t b \geq \frac{1}{2}$ implies $a_u b \geq \frac{1}{2}$ and so it is sufficient to assume $a_t b \geq \frac{1}{2}$.

commodity at delivery.

Let $\pi^\tau = 1 - e^{\star\tau}$ denote the net price of the processed product to the cooperative of type-τ. Then the optimal profit of the the cooperative of type-τ is :

$$\Pi^\tau(\omega;\beta) = \frac{1-\omega}{(1-\beta)} 2a_\tau(1-\pi^\tau). \tag{3.9}$$

As aforementioned, we consider only the pooling equilibrium. To characterize this equilibrium it is useful to consider the FOC's optimal behavior in a perfect information world, that is a world where the farmers know for sure whether they are facing a trustworthy or an untrustworthy cooperative, and to specify the participation constraints in an imperfect information world, where the farmers are uncertain as to the type of cooperative they are dealing with.

Let $\hat{e}^t(\omega;1)$ and $\hat{e}^u(\omega;0)$ denote respectively the optimal delivery price of the trustworthy FOC and of the untrustworthy FOC in perfect information. In such a world of perfect information, the farmer knows with certainty whether he is facing a trustworthy or an untrustworthy FOC. If the farmer believes he is facing a trustworthy cooperative he is assured of receiving the full dividend at the end. We then have :

$$\hat{e}^t(\omega;1) = 0 \qquad \hat{e}^u(\omega;0) = 1 - \sqrt{1 - \frac{1+\omega+\omega^2}{6ba_u}}.$$

The payoff function $\Pi^t(e,\omega;\beta = 1)$ of the trustworthy FOC is decreasing in e, so its maximum is attained in $e = 0$. This means that in a trustworthy environment, the financial tightness index a_t does not matter. The cooperative can choose to propose a zero delivery price and get the entire market, since the farmers trust that they will receive the full dividend once the processed product is sold. In an untrustworthy environment, the cooperative needs to do more to prove its goodwill, and then the financial tightness index matters.

Consider now the imperfect information world and let $e^t(\omega;\beta)$ and $e^u(\omega;\beta)$ denote the trustworthy FOC and untrustworthy FOC proposals in such a world. The worst scenario for both types of FOC, and especially for the trustworthy FOC, is that farmers

believe that they are untrustworthy ($\beta = 0$). In an imperfect information equilibrium, they will want to secure at least the optimal payoffs for this case in order to participate. This yields the following participation constraints for respectively the trustworthy and the untrustworthy FOC :

$$\begin{aligned} (a) \quad \Pi^t(e^t, \omega; \beta) \geq \quad \tilde{\Pi}^t = \max_e \Pi^t(e, \omega; 0) \\ (b) \quad \Pi^u(e^u, \omega; \beta) \geq \quad \hat{\Pi}^u = \Pi^u(\hat{e}^u, \omega; 0) \end{aligned} \tag{3.10}$$

We turn now to the characterization of FOC's optimal behavior in a pooling equilibrium. Such an equilibrium does not always exist and we need some restrictions on the set of parameters to guarantee existence.

Proposition 1 *Let* $\Delta(\omega; \theta, a_u, a_t, b) = 2\theta\pi^t(1 - \pi^u) - (\pi^t - \pi^u)^2$ *and*

$$\Lambda = \left\{ (a_u, a_t, b) \in \mathbb{R}^{+3} : ba_t \geq \frac{1}{2}, \max_\omega \Delta(\omega; 1, a_u, a_t, b) > 0 \right\}.$$

Then, for all $(a_u, a_t, b) \in \Lambda$*, there exists* $\underline{\theta} \in (0, 1)$ *such that for all* $\theta \geq \underline{\theta}$*, for all* $\omega \in [0, 1]$*, there exists proposal* $e^{\star t}(\omega; \theta)$ *representing FOCs' best response in a pooling equilibrium, with*

$$e^{\star t}(\omega; \theta) = 1 - \sqrt{1 - \frac{1 + \omega + \omega^2}{6ba_t}} \tag{3.11}$$

The proof of this proposition is given in the Appendix.

In a pooling equilibrium, the untrustworthy FOC mimics the trustworthy FOC. This results in the untrustworthy FOC making an extra effort, $e^{\star t} > e^{\star u}$, since its separating equilibrium delivery price is simply its optimal delivery price in imperfect information, $e^{\star u} = \hat{e}^u$. We can consider such behavior as an investment in reputation. The FOC's best response function is increasing in ω. When the IOF increases its proposal, the FOC responds by making a greater effort.

3.3.3 Investor owned firm pricing behavior

The pricing behavior of the investor owned firm is analyzed in the framework of a pooling equilibrium on the part of the FOC. Farmers who deliver to the IOF produce the same optimal quantity $q^{iof} = \frac{\omega}{b}$. Thus, the total purchase of the IOF is simply this quantity times the mass of farmers contracting with it. We have :

$$S^{iof}(\omega; e) = \int_0^{\alpha^\star} \frac{\omega}{b} d\alpha = \frac{\omega}{b}\left(1 - \frac{1-\omega}{(1-\theta)(1-e)}\right). \tag{3.12}$$

The price elasticity of the IOF's demand curve is :

$$\varepsilon_{S,\omega} = \left(1 + \frac{\omega}{\omega - (1-(1-\theta)(1-e))}\right). \tag{3.13}$$

Given that cooperatives are pooling, the IOF's payoff is :

$$\begin{aligned} \Pi^{iof}(e,\omega) &= (1-\omega)S^{iof}(e,\omega) \\ &= \frac{((1-\theta)(1-e)-1)\omega - ((1-\theta)(1-e)-2)\omega^2 - \omega^3}{b(1-\theta)(1-e)}. \end{aligned} \tag{3.14}$$

As derived from (3.6), a mixed market requires $1-(1-\theta)(1-e) \leq \omega$. So, the IOF maximizes its payoff, given by (3.14), with respect to ω and subject to the constraints $1-(1-\theta)(1-e) \leq \omega \leq 1$.

Assume an interior solution. The optimality condition can then be expressed as :

$$\frac{1-\omega}{\omega} = \frac{1}{\varepsilon_{S,\omega}} = \left(1 + \frac{\omega}{\omega - (1-(1-\theta)(1-e))}\right)^{-1}. \tag{3.15}$$

In other words, the price spread equals the inverse of the price elasticity of the demand curve. The cooperative plays a pro-competitive role if it induces the IOF to set a lower price spread, $((1-\omega)/\omega)$, which is equivalent to having a higher price for the primary commodity. We can now derive the following comparative static statements from the first-order condition given by equation (3.15).

First, the IOF price spread decreases with the cooperative's reputation. A quick proof

of this assertion is the following : if θ increases, the right-hand side (RHS) of equation (3.15) decreases. Thus, in order to maintain equality, the left-hand side (LHS) must also decrease. Using the same argument, we deduce that if the level of effort increases, the IOF price spread decreases.

This and the sensitivity analysis on the other parameters can be formally derived by the use of implicit derivation. Let

$$K(\omega, e, \theta) = 3\omega^2 - 2(2 - (1-\theta)(1-e)))\omega + (1 - (1-\theta)(1-e)).$$

Then the first-order condition in (3.15) can alternatively be expressed as

$$K(\omega, e, \theta) = 0. \tag{3.16}$$

Using implicit differentiation, the derivative of the IOF's price with respect to any parameter x is equal to

$$\frac{\partial \omega}{\partial x} = -\frac{\partial K/\partial x}{\partial K/\partial \omega}. \tag{3.17}$$

Then, if $\omega > \frac{1}{2}$, we have :

$$\begin{aligned} \frac{\partial K}{\partial \omega} =& 2(\omega - (1 - (1-\theta)(1-e)) + 2(2\omega - 1) > 0 \\ \frac{\partial K}{\partial \theta} =& (1-e)(1-2\omega) < 0 \\ \frac{\partial K}{\partial e} =& (1-\theta)(1-2\omega) < 0. \end{aligned} \tag{3.18}$$

The condition is not restrictive, since in a mixed market the IOF cannot set a price inferior to its monopsony price, $\omega = \frac{1}{2}$. It follows that the IOF's price is an increasing function of the cooperative spot price, e, as well as of the reputation index θ.

Consider now the expression for the IOF's best response function. Given that $1 - (1-\theta)(1-e) + (1-\theta)^2(1-e)^2 \geq 0$, the two stationary points obtained from solving

There are two solutions to (3.16), namely :

$$\omega_1^\star = \frac{(2-A) - \sqrt{1-A+A^2}}{3} \qquad \omega_2^\star = \frac{(2-A) + \sqrt{1-A+A^2}}{3},$$

where $A = (1-\theta)(1-e)$. Those two roots are real, since $A < 1$. The second derivative of $\Pi^{iof}(e,\omega)$ with respect to ω is

$$\frac{\partial^2 \Pi^f(e,\omega)}{\partial \omega^2} = -2\frac{(A-2)+3\omega}{bA}.$$

It is positive when evaluated at $\omega_1^\star$, making $\omega_1^\star$ a local minimum, and negative when evaluated at $\omega_2^\star$, making $\omega_2^\star$ a local maximum. Hence the IOF's best response function is :

$$\omega^\star(e;\theta) = \frac{(2-(1-\theta)(1-e)) + \sqrt{1-(1-\theta)(1-e)+(1-\theta)^2(1-e)^2}}{3}. \quad (3.19)$$

Having derived the response functions of the different players, we can now characterize the equilibrium of the game.

3.3.4 The equilibrium spot price, effort and market shares

Equations (3.5), (3.8) and (3.19) define the equilibrium of the game, $\left(e^\star, \omega^\star, (\mathbf{q}_\alpha(e^\star,\omega^\star)_{\alpha\in[0,1]})\right)$ with the resulting equilibrium threshold $\alpha^\star(e^\star,\omega^\star)$. Although the model seems simple, a closed-form analytical characterization of the equilibrium is not possible, so we make use of a numerical resolution.

As previously demonstrated, the best response functions of the two processors are upward sloping, so their delivery prices are strategic complements. Figure 3.2 illustrates the players' responses and the corresponding equilibria for given levels of the cooperative's reputation index, θ, and the financial tightness index, a_t.

We highlight four possible equilibria, each of them corresponding to a couple $(\theta, a_t) \in \{0.1, 0.7\} \times \{0.52, 1.0\}$ for the parameters.[6] For each equilibrium the "*pace-maker*

6. We set $b = 1$ in each case. If we consider the couple of parameters $(a_t, a_u) = (0.52, 0.55)$, we have $\underline{\theta} = 0.0405$. For the couple $(a_u, a_t) = (1, 1.5)$ we have $\underline{\theta} = 0.0858$.

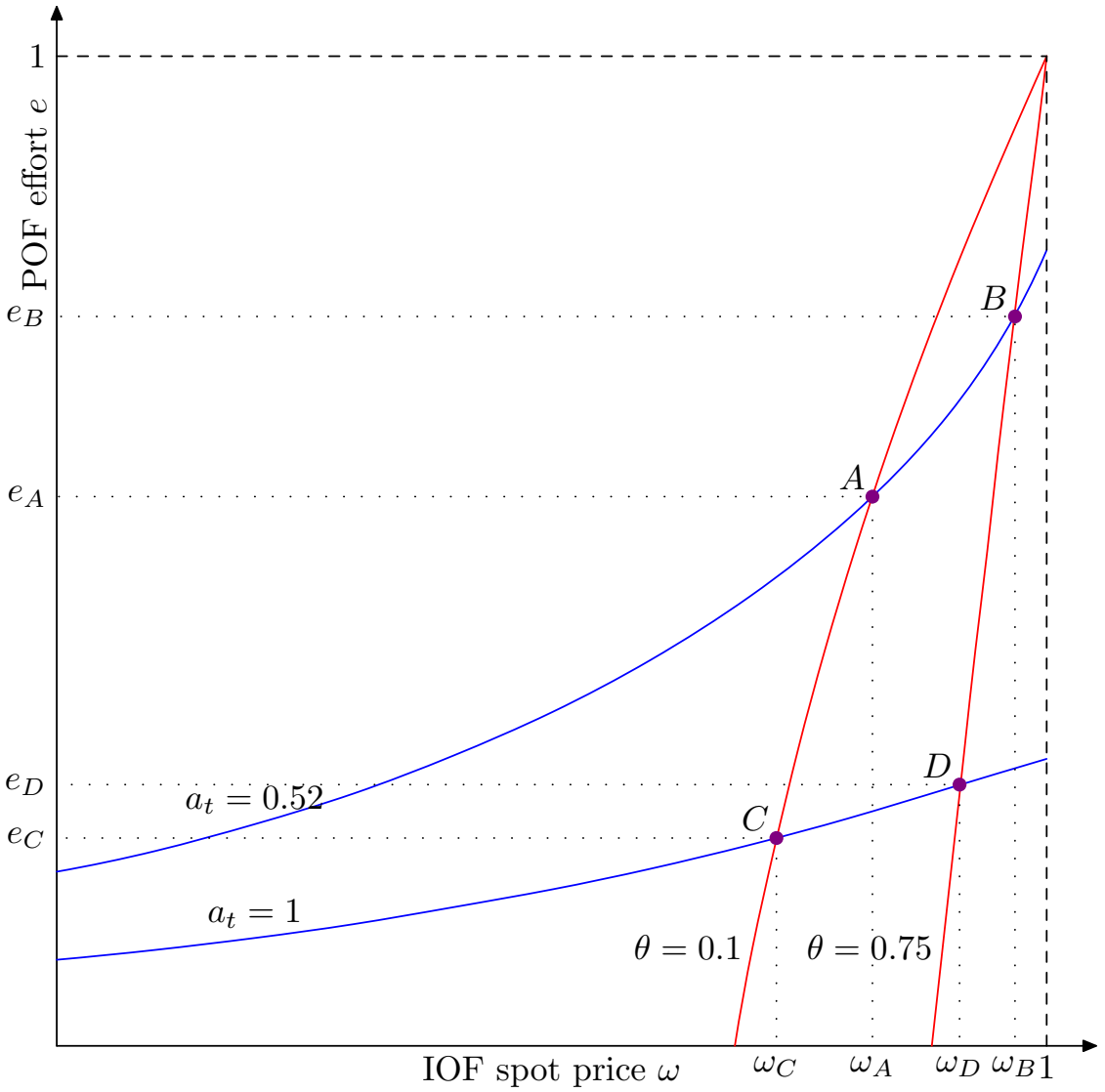

FIGURE 3.2 – **Players' best responses and corresponding equilibria given reputation index,θ, and financial market tightness index,** a_τ

condition" $e^\star + \theta(1-e^\star) < \omega^\star$ is met, in order to have a mixed duopsony. The cooperative's level of effort, which is also its delivery price, is always inferior to the delivery price of the IOF. The equilibrium labeled A is obtained for a couple $(\theta, a_t) = (0.1, 0.52)$. If, from this position, we increase the reputation index to 0.7, the new equilibrium moves to B where the prices of both processors are strictly higher. Players move simultaneously, so we cannot perform a rigorous dynamic analysis, although we may suggest the follo-

wing adjustment process. All other things equal, an improvement in the reputation index raises the expected secured price of the cooperative and the IOF responds with a higher price ; this in turn induces the cooperative to make more effort as a response to the move of the IOF.

If the financial tightness index is increased, we move from A to C, and both processors price strictly lower. With a tighter financial market, the cooperative's effort is not appropriately rewarded, so it chooses to make less effort. Hence, the IOF finds it possible to increase its profits by pricing lower.

The impact of the reputation and financial tightness can be better appreciated if we take into account the other component of the equilibrium : the market share threshold, or the level of trust $\alpha^\star$ that splits farmers in two sets, those dealing with the IOF and those participating in the cooperative. Figure 3.3 illustrates the evolution of the market-share threshold with respect to the reputation index and for different levels of financial tighteness index. The four equilibria highlighted in figure (3.3) are the same as those of figure 3.2. A good reputation compels the IOF not only to price higher, but also to content itself with a lower market share, as illustrated by equilibria (A, B) and (C, D). Conversely, tighter financial conditions are an impediment for cooperative growth. It compels the IOF not only to price lower, but to have a higher market share, as illustrated by equilibria (A, C) and (B, D).

To summarize, good reputation and good financial conditons foster the cooperative's pro-competitive role in a mixed duopsony. In a perfect market, both processors would price at the net output price and make no profit. Hence, farmers would get the maximal price possible. These results seem to be at the basis of the purpose of agricultural market liberalization. The fact is that decision makers do not take into account certain rigidities rooted in developing countries, such as those of reputation and financial conditions.

It is interesting to look at the behavior of overall production. Without perfect reputation and with financial tightness, the total output generated by the two processors as intermediaries is less than the optimal output which equals $\frac{1}{b}$. To see this, let $S(\theta, a_t) = S^{iof}(\theta, a_t) + S^{foc}(\theta, a_t)$ denote the overall production, where $S^{iof}(\theta, a_t)$ is the IOF's purchase and $S^{foc}(\theta, a_t)$ is the cooperative's purchase. From the market-share

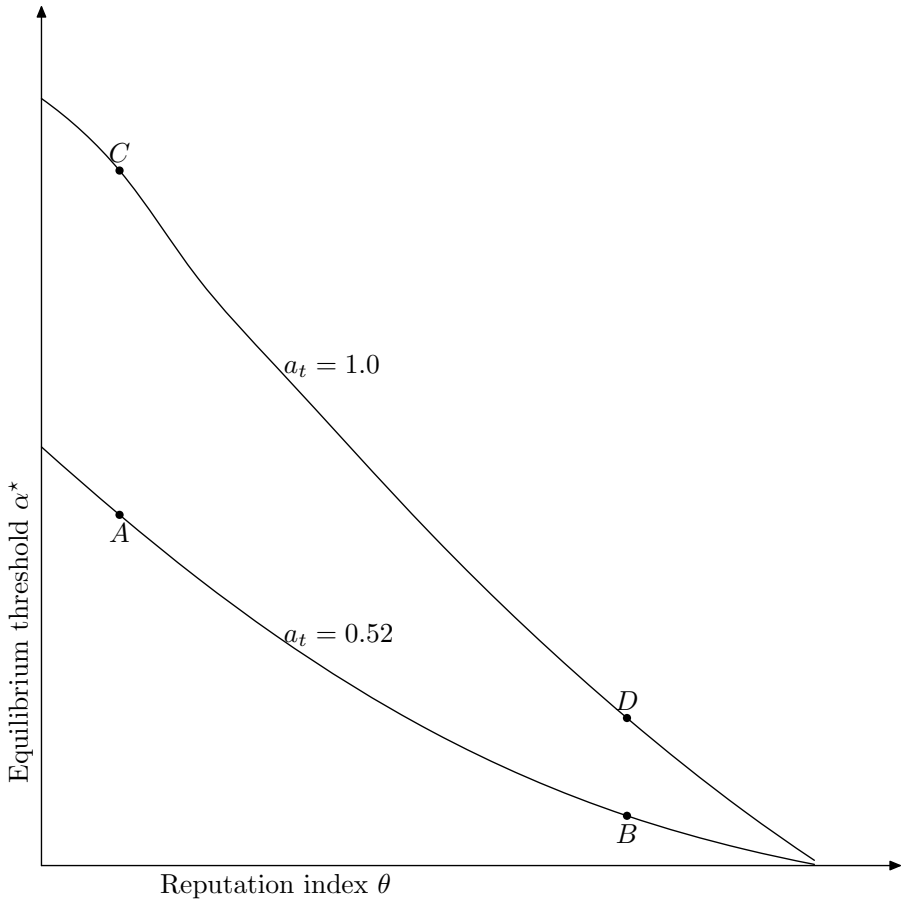

FIGURE 3.3 – **Market share threshold with respect to reputation index, θ, and financial market tightness index a_τ**

threshold, we have $\omega = 1 - (1-\theta)(1-e)(1-\alpha^\star)$ and then, the IOF purchase becomes

$$S^{iof}(\theta, a_t) = \frac{\alpha^\star}{b} - \frac{(1-\theta)(1-e)}{b}\alpha^\star(1-\alpha^\star).$$

If we integrate over the production of the farmers participating in the cooperatives, we have

$$\begin{aligned} S^{foc} &= \int_{\alpha^\star}^{1} \frac{[1-(1-\theta)(1-e)(1-\alpha)]}{b} d\alpha \\ &= \frac{1-\alpha^\star}{b} - \frac{(1-\theta)(1-e)}{2b}(1-\alpha^\star)^2 \end{aligned}$$

Summing up $S^{iof}(\theta, a_t)$ and $S^{foc}(\theta, a_t)$ and a little algebra yields

$$\begin{aligned} S(\theta, a_t) &= \frac{1}{b} - (1-\theta)(1-e^\star)(1-\alpha^\star)\frac{1+\alpha^\star}{2b} \\ &= \frac{1}{b} - (1-\omega)\frac{1+\alpha^\star}{2b} \end{aligned} \tag{3.20}$$

The overall production will be less than the vertically-integrated production $\frac{1}{b}$ unless one of the following conditions is fulfilled :

- $e = 1$, that is, no financial constraint prevents the cooperative from paying as delivery price the net output price ;
- $\theta = 1$, we are in a social environment with perfect reputation ;
- $\omega = 1$, the delivery price paid by the IOF equals the full output price.

3.4 Conclusion

In competition with an IOF, a FOC has not an absolute advantage as pointed out by Sexton [11] and Albæk and Schultz [1]. The result of this competition depends on the cooperative's reputation, based on its moral and technical skills, and a perfect financial market. The difficulties of growth of agricultural cooperatives in most sub-Saharan countries, after market liberalization, can be explained in good part by the lack of investment in reputation and an imperfect financial market. It would be useful to carry out an investigative survey similar to that of James and Sykuta [7] to test the predictions of this model.

The ethnic mix resulting from migration towards economic development areas implies that we can no longer count on social homogeneity to foster trust and reputation in those areas. Policymakers then need to establish an environment that secures property rights in order to sustain the growth of FOCs. Some suggestions can be made towards this. First of all, it is necessary to reinforce the legal framework that governs the creation of agricultural cooperatives and to harden the penal sanctions for anyone who diverts small farmers savings. In addition, one could create a fund to compensate farmers whose dividends would be diverted. An expected result is to reduce opportunistic behavior, increase the security of property rights, foster trust and increase the cooperatives' membership rates. A regulatory agency that frequently audits farmers' and cooperatives' books and management could also be helpful.

The local financial market is not only an impediment to the growth of the agricultural sector, but to the growth of the whole economy. Policymakers should take vigourous actions to make the financing of economic activities more easily available. The training of farmers in modern management skills is an additional factor that could help sustain cooperative growth.

Agricultural market liberalization has shown its limits. Policymakers need to be more imaginative in developing policies that support this liberalization, taking account the local realities.

CONCLUSION

Les modèles espaces états sont utiles pour représenter les relations dynamiques avec variables latentes. L'analyse des modèles espaces-états non linéaires et non Gaussiens se fait généralement sous l'hypothèse de l'indépendance conditionnelle entre observation courante et innovation courante de la variable d'état. Cette hypothèse d'indépendance conditionnelle apparaît beaucoup plus comme un artifice de calcul qu'une représentation de la réalité des liens entre variables latentes et variables observées. Les travaux de Jacquier et al[2004], Omori et al[2007] pour le modèle de volatilité stochastique, et ceux de Feng et al[2004] pour le modèle de durée conditionnelle stochastique, ont montré que la dépendance conditionnelle est plus réaliste pour modéliser les relations dynamiques avec variables latentes.

Les modèles espace-états avec dépendance conditionnelle sont peu utilisés à cause de la difficulté à trouver une procédure générique d'estimation des paramètres. Les propositions à date dans la littérature sont spécifiques aux modèles de volatilité stochastique et au modèle de durée conditionnelle stochastique.

Les deux premiers articles de ma thèse ont proposé des procédures génériques pour analyser les modèles espace-états non linéaires et non Gaussiens avec variable d'état univariée. Ces procédures sont basées sur des simulations du vecteur d'état et du vecteur de paramètres de leur distribution à posteriori. Les simulateurs élaborés sont numériquement efficaces en ce sens qu'ils permettent de construire des chaînes de Markov du vecteur d'état et du vecteur de paramètres avec une très faible dépendance linéaire. L'utilisation par un tiers modélisateur, de ces procédures, ne requière que le calcul des dérivées partielles de la densité conditionnelle du vecteur d'observation ou de la densité jointe des vecteurs d'état et d'observation décrivant le modèle espace état. Il existe des routines de calcul amplement testées qui rendent le calcul analytique de ces dérivées partielles non nécessaire.

Nos travaux ouvrent ainsi de nouvelles perspectives pour l'analyse des modèles espaces états non linéaires et non Gaussiens. Les procédures d'estimation proposées concernent les modèles espaces-états avec variable d'état unidimensionnelle. Ces tra-

vaux peuvent être étendus pour proposer des estimateurs numériquement efficaces pour les modèles espaces états avec une variable d'état multidimensionnelle. La simulation des vecteurs d'état et de paramètre se fait avec réalisation fixée du vecteur d'observation. Étendre ces travaux à l'apprentissage séquentiel peut aussi être envisagé.

Notre troisième article fut un essai dans le domaine de l'analyse des marchés de commodité. Essayant de comprendre les difficultés de croissance des coopératives agricoles dans les pays d'Afrique subsaharienne, après la libéralisation des filières agricoles, nous avons fait le constat que la confiance et la réputation jouaient un rôle important. Ce constat n'est pas spécifique aux pays africains. James et Sykuta [2004] arrivent à une conclusion similaire après une enquête auprès des fermiers de mais et soja dans le Missouri, aux États Unis. Pourtant, les modèles théoriques n'incluent pas de façon spécifique la confiance et la réputation comme facteur déterminant de l'équilibre entre une coopérative de fermiers et une firme privée en compétition dans un marché agricole. Nous avons proposé un modèle théorique qui fait de la confiance et de la réputation un facteur déterminant de l'issue de la compétition prix entre une firme privée et une coopérative agricole. La réputation accroît l'incidence positive de la coopérative. Une coopérative avec une très bonne réputation oblige la firme privée à pratiquer un prix à la livraison plus élevée tout en réduisant ses parts de marché.

Annexe I

Appendix to Chapter 1

I.1 Precomputation

Here we compute the precision $\bar{\Omega}$ and covector $\bar{c}$ of the marginal distribution of α, and the mode $a = (a_1, \ldots, a_n)$ of the target distribution. Bi-products of the computation of a include several quantities used elsewhere, including $\bar{\bar{\Omega}}$ and $\bar{\bar{c}}$, the precision and covector of a Gaussian approximation $\mathrm{N}(\bar{\bar{\Omega}}^{-1}\bar{\bar{c}}, \bar{\bar{\Omega}}^{-1})$ of the target distribution, and the conditional variances $\Sigma_1, \ldots, \Sigma_t, \ldots, \Sigma_n$.

As the state dynamics are no different, we compute $\bar{\Omega}$ and $\bar{c}$ exactly as in McCausland (2010) :

$$\bar{\Omega}_{t,t} = \omega_{t-1} + \omega_t \phi_t^2, \quad \bar{\Omega}_{t,t+1} = -\omega_t \phi_t, \quad t = 1, \ldots, n-1,$$

$$\bar{\Omega}_{n,n} = \omega_{n-1},$$

$$\bar{c}_t = \begin{cases} \omega_{t-1} d_{t-1} - \omega_t \phi_t d_t & t = 1, \ldots, n-1, \\ \omega_{n-1} d_{n-1} & t = n. \end{cases} \tag{I.1}$$

As in McCausland (2010), we use a Newton-Raphson method to find the mode of the target distribution. At each iteration, we compute a precision $\bar{\bar{\Omega}}(\alpha)$ and covector $\bar{\bar{c}}(\alpha)$ of a Gaussian approximation to the target distribution based on a second order Taylor series expansion of the log target density around the current value of α. Specifically, $\bar{\bar{\Omega}}(\alpha)$ is the negative Hessian matrix of $\log f(\alpha|y)$ with respect to α at the current value of α. It is a symmetric tri-diagonal matrix, with non-zero upper triangular elements given by

$$\bar{\bar{\Omega}}_{t,t}(\alpha) = \bar{\Omega}_{t,t} - \left(\psi_t^{(2,0)}(\alpha_t, \alpha_{t+1}) + \psi_{t-1}^{(0,2)}(\alpha_{t-1}, \alpha_t)\right), \quad t = 2, \ldots, n-1,$$

$$\bar{\bar{\Omega}}_{1,1}(\alpha) = \bar{\Omega}_{1,1} - \psi_t^{(2,0)}(\alpha_t, \alpha_{t+1}), \quad \bar{\bar{\Omega}}_{nn}(\alpha) = \bar{\Omega}_{n,n} - \left(\psi_n^{(2)}(\alpha_n) + \psi_{n-1}^{(0,2)}(\alpha_{n-1}, \alpha_n)\right),$$

$$\bar{\bar{\Omega}}_{t,t+1}(\alpha) = \bar{\Omega}_{t,t+1} - \psi_t^{(1,1)}(\alpha_t, \alpha_{t+1}), \quad t = 1, \ldots, n-1.$$

The covector $\bar{\bar{c}}(\alpha)$ is

$$\bar{\bar{c}}(\alpha) \doteq \bar{\bar{\Omega}}(\alpha)\alpha + \frac{\partial \log f(y|\alpha)}{\partial \alpha^\top},$$

and its elements are

$$\bar{\bar{c}}_t(\alpha) = \begin{cases} \bar{c}_t + \bar{\bar{\Omega}}_{t,t}\alpha_t + \bar{\bar{\Omega}}_{t,t+1}\alpha_{t+1} + \psi_t^{(1,0)}(\alpha_t, \alpha_{t+1}) & t = 1 \\ \bar{c}_t + \bar{\bar{\Omega}}_{t,t-1}\alpha_{t-1} + \bar{\bar{\Omega}}_{t,t}\alpha_t + \bar{\bar{\Omega}}_{t,t+1}\alpha_{t+1} + \psi_t^{(1,0)}(\alpha_t, \alpha_{t+1}) + \psi_{t-1}^{(0,1)}(\alpha_{t-1}, \alpha_t) & t = 2, \ldots, n-1 \\ \bar{c}_n + \bar{\bar{\Omega}}_{n,n-1}\alpha_{n-1} + \bar{\bar{\Omega}}_{nn}\alpha_n + \psi_n^{(1)}(\alpha_n) + \psi_{n-1}^{(0,1)}(\alpha_{n-1}, \alpha_n)(\alpha_{n-1}, \alpha_n) & t = n. \end{cases} \tag{I.2}$$

Let $\bar{\bar{\Omega}} \doteq \bar{\bar{\Omega}}(a)$ and $\bar{\bar{c}} \doteq \bar{\bar{c}}(a)$. Then the mean (and mode) of the Gaussian approximation $\mathrm{N}(\bar{\bar{\Omega}}^{-1}\bar{\bar{c}}, \bar{\bar{\Omega}}^{-1})$ is a, the mode of the target distribution, and its log density has the same Hessian matrix as the log target density at a.

While these expressions for $\bar{\bar{\Omega}}$ and $\bar{\bar{c}}$ are more complicated than those in McCausland (2010), once we have them, we compute the mode a in the same way. Roughly speaking, we iterate the computation $\alpha' = \bar{\bar{\Omega}}(\alpha)^{-1}\bar{\bar{c}}(\alpha)$ until numerical convergence. We use two modifications to this procedure, one to accelerate convergence using higher order derivatives and the other to resort to line searches in the rare cases of non-convergence.

I.2 Polynomial approximations of $a_{t|t+1}$ and $s_{t|t+1}$

Here we compute coefficients of polynomial approximations of $a_{t|t+1}(\alpha_{t+1})$ and $s_{t|t+1}(\alpha_{t+1})$. Recall that these are the conditional mean and log variance of α_t given α_{t+1} according to a Gaussian approximation of the conditional distribution of $\alpha_1, \ldots, \alpha_t$ given α_{t+1} and y. The approximations are exact Taylor series expansions around a_{t+1} and so the coefficients are based on the derivatives of these functions at a_{t+1}.

We derive recursive expressions for these derivatives that are correct for any order r. In practice, the computational cost rises quickly and the benefits diminish quickly in r. We provide simplified expressions for $a_t^{(r)} \doteq a_{t|t+1}^{(r)}(a_{t+1})$ up to order $r = 5$ and $s_t^{(r)} \doteq s_{t|t+1}^{(r)}(a_{t+1})$ up to order $r = 4$.

The basic strategy involves taking derivatives of two identities. The first is a first order necessary condition on $a_{t-1|t+1}(\alpha_{t+1})$ and $a_{t|t+1}(\alpha_{t+1})$ for $(a_{1|t+1}(\alpha_{t+1}), \ldots, a_{t|t+1}(\alpha_{t+1}))$

to be the conditional mode of $(\alpha_1, \ldots, \alpha_t)$ given α_{t+1} and y. The second is the identity $a_{t-1|t+1}(\alpha_{t+1}) = a_{t-1|t}(a_{t|t+1}(\alpha_{t+1}))$.

I.2.1 General Formula

We begin with the case $t = 1$. Since $f(\alpha_1|\alpha_2, y) \propto f(\alpha_1, \alpha_2)f(y_1|\alpha_1, \alpha_2)$, we can write

$$\log f(\alpha_1|\alpha_2, y) = -\frac{1}{2}\bar{\Omega}_{1,1}\alpha_1^2 - \bar{\Omega}_{1,2}\alpha_1\alpha_2 + \bar{c}_1\alpha_1 + \log f(y_1|\alpha_1, \alpha_2) + k. \tag{I.3}$$

where k does not depend on α_1. The conditional mode $a_{1|2}(\alpha_2)$ maximizes $\log f(\alpha_1|\alpha_2, y)$ and must therefore satisfy

$$-\bar{\Omega}_{1,1}a_{1|2}(\alpha_2) - \bar{\Omega}_{1,1}\alpha_2 + \bar{c}_1 + \psi_1^{(1,0)}(a_{1|2}(\alpha_2), \alpha_2) = 0. \tag{I.4}$$

Taking the derivative of (I.4) with respect to α_2, and using the definitions $\bar{\bar{\Omega}}_{1,1|2}(\alpha_2) = (\bar{\Omega}_{1,1} - \psi_1^{(2,0)}(a_{1|2}(\alpha_2), \alpha_2))$ and $\bar{\bar{\Omega}}_{1,2|2}(\alpha_2) = \bar{\Omega}_{1,2} - \psi_1^{(1,1)}(a_{1|2}(\alpha_2), \alpha_2)$ gives

$$\bar{\bar{\Omega}}_{1,1|2}(\alpha_2)a_{1|2}^{(1)}(\alpha_2) = -\bar{\bar{\Omega}}_{1,2|2}(\alpha_2). \tag{I.5}$$

Solving for $a_{1|2}^{(1)}(\alpha_2)$, we obtain

$$a_{1|2}^{(1)}(\alpha_2) = -\Sigma_{1|2}(\alpha_2)\bar{\bar{\Omega}}_{1,2|2}(\alpha_2), \tag{I.6}$$

where $\Sigma_{1|2}(\alpha_2) = \bar{\bar{\Omega}}_{1,1|2}^{-1}(\alpha_2)$ from equation (1.10). Setting $\alpha_2 = a_2$ gives $a_1^{(1)} = -\Sigma_1\bar{\bar{\Omega}}_{1,2}$.

We now derive an expression allowing us to compute $a_1^{(r)}$ in terms of $a_1^{(i)}$, $i < r$. First, differentiate (I.5) $(r-1)$ times with respect to α_2. Using Leibniz's rule, we obtain

$$\sum_{i=0}^{r-1}\binom{r-1}{i}\bar{\bar{\Omega}}_{1,1|2}^{(r-1-i)}(\alpha_2)a_{1|2}^{(i+1)}(\alpha_2) = -\bar{\bar{\Omega}}_{1,2|2}^{(r-1)}(\alpha_2).$$

Then solving for $a^{(r)}_{1|2}(\alpha_2)$ gives

$$a^{(r)}_{1|2}(\alpha_2) = -\Sigma_{1|2}(\alpha_2)\left[\sum_{i=0}^{r-2}\binom{r-1}{i}\bar{\bar{\Omega}}^{(r-1-i)}_{1,1|2}(\alpha_2)a^{(i+1)}_{1|2}(\alpha_2) + \bar{\bar{\Omega}}^{(r-1)}_{1,2|2}(\alpha_2)\right]. \quad \text{(I.7)}$$

Finally, we evaluate (I.7) at $\alpha_2 = a_2$ to obtain

$$a^{(r)}_1 = -\Sigma_1\left[\sum_{i=0}^{r-2}\binom{r-1}{i}\bar{\bar{\Omega}}^{(r-1-i)}_{1,1}a^{(i+1)}_1 + \bar{\bar{\Omega}}^{(r-1)}_{1,2}\right]. \quad \text{(I.8)}$$

We now derive an expression relating the $a^{(r)}_1$ and the $s^{(r)}_1$, which we will use to obtain the latter from the former. First recall the definition $\Sigma_{1|2}(\alpha_2) = \exp(s_{1|2}(\alpha_2))$. Using Faà Di Bruno's formula for derivatives of compound functions, we obtain, for $i \geq 1$,

$$\begin{aligned}\Sigma^{(i)}_{1|2}(\alpha_2) &= \sum_{j=1}^{i}\exp(s_{1|2}(\alpha_2))B_{i,j}(s^{(1)}_{1|2}(\alpha_2), \ldots, s^{(i-j+1)}_{1|2}(\alpha_2)) \\ &= \Sigma_{1|2}(\alpha_2)B_i(s^{(1)}_{1|2}(\alpha_2), \ldots, s^{(i)}_{1|2}(\alpha_2)), \end{aligned} \quad \text{(I.9)}$$

where the $B_{i,j}$ are Bell polynomials and B_i is the i'th complete Bell polynomial. Appendix I.5 shows how to compute these polynomials. We now differentiate (I.6) $(r-1)$ times with respect to α_2, to obtain

$$\begin{aligned}a^{(r)}_{1|2}(\alpha_2) &= -\sum_{i=0}^{r-1}\binom{r-1}{i}\Sigma^{(i)}_{1|2}(\alpha_2)\bar{\bar{\Omega}}^{(r-1-i)}_{1,2|2}(\alpha_2) \\ &= -\Sigma_{1|2}(\alpha_2)\sum_{i=0}^{r-1}\binom{r-1}{i}B_i(s^{(1)}_{1|2}(\alpha_2), \ldots, s^{(i)}_{1|2}(\alpha_2))\bar{\bar{\Omega}}^{(r-1-i)}_{1,2|2}(\alpha_2).\end{aligned}$$

Evaluating at $\alpha_2 = a_2$ gives us the desired expression :

$$a^{(r)}_1 = -\Sigma_1\sum_{i=0}^{r-1}\binom{r-1}{i}B_i(s^{(1)}_1, \ldots, s^{(i)}_1)\bar{\bar{\Omega}}^{(r-1-i)}_{1,2}. \quad \text{(I.10)}$$

We now move on to the case $1 < t < n$. The conditional mode $a_{1:t|t+1}(\alpha_{t+1}) =$

$(a_{1|t+1}(\alpha_{t+1}), \ldots, a_{t|t+1}(\alpha_{t+1}))$ must satisfy the first order necessary condition

$$\begin{aligned} 0 =& \bar{c}_t - \bar{\Omega}_{t-1,t} a_{t-1|t+1}(\alpha_{t+1}) - \bar{\Omega}_{t,t} a_{t|t+1}(\alpha_{t+1}) - \bar{\Omega}_{t,t+1}\alpha_{t+1} \\ &+ \psi_{t-1}^{(0,1)}(a_{t-1|t}(a_{t|t+1}), a_{t|t+1}) + \psi_t^{(1,0)}(a_{t|t+1}, \alpha_{t+1}). \end{aligned} \tag{I.11}$$

Taking the derivative of (I.11) with respect to α_{t+1} gives

$$\bar{\bar{\Omega}}_{t,t-1}(\alpha_{t+1}) a^{(1)}_{t-1|t+1}(\alpha_{t+1}) + \bar{\bar{\Omega}}_{t,t}(\alpha_{t+1}) a^{(1)}_{t|t+1}(\alpha_{t+1}) + \bar{\bar{\Omega}}_{t,t+1}(\alpha_{t+1}) = 0. \tag{I.12}$$

Using the identity $a_{t-1|t+1}(\alpha_{t+1}) = a_{t-1|t}\left(a_{t|t+1}(\alpha_{t+1})\right)$ and the chain rule gives

$$a^{(1)}_{t-1|t+1}(\alpha_{t+1}) = a^{(1)}_{t-1|t}(a_{t|t+1}(\alpha_{t+1})) a^{(1)}_{t|t+1}(\alpha_{t+1}). \tag{I.13}$$

Substituting (I.13) in (I.12), we obtain

$$\left(\bar{\bar{\Omega}}_{t,t-1}(\alpha_{t+1}) a^{(1)}_{t-1|t}(a_{t|t+1}(\alpha_{t+1})) + \bar{\bar{\Omega}}_{t,t}(\alpha_{t+1})\right) a^{(1)}_{t|t+1}(\alpha_{t+1}) = -\bar{\bar{\Omega}}_{t,t+1}(\alpha_{t+1}).$$

Then, following an analogous development in **?**], we can show by induction that

$$a^{(1)}_{t|t+1}(\alpha_{t+1}) = -\Sigma_{t|t+1}(\alpha_{t+1}) \bar{\bar{\Omega}}_{t,t+1}(\alpha_{t+1}), \quad t = 2, \ldots, n-1, \tag{I.14}$$

where $\left[\Sigma_{t|t+1}(\alpha_{t+1})\right]^{-1} = \bar{\bar{\Omega}}_{t,t-1}(\alpha_{t+1}) a^{(1)}_{t-1|t}(a_{t|t+1}(\alpha_{t+1})) + \bar{\bar{\Omega}}_{t,t}(\alpha_{t+1})$. Taking $\alpha_{t+1} = a_{t+1}$ in (I.14) gives

$$a^{(1)}_t = -\Sigma_t \bar{\bar{\Omega}}_{t,t+1}. \tag{I.15}$$

For $r \geq 2$, we use Leibniz's rule to differentiate (I.12) $(r-1)$ times with respect to α_{t+1} and obtain

$$\sum_{i=0}^{r-1} \binom{r-1}{i} \left(\bar{\bar{\Omega}}^{(i)}_{t,t-1}(\alpha_{t+1}) a^{(r-i)}_{t-1|t+1}(\alpha_{t+1}) + \bar{\bar{\Omega}}^{(i)}_{t,t}(\alpha_{t+1}) a^{(r-i)}_{t|t+1}(\alpha_{t+1})\right) = -\bar{\bar{\Omega}}^{(r-1)}_{t,t+1}(\alpha_{t+1}). \tag{I.16}$$

Using Faà di Bruno's formula for arbitrary order derivatives of compound functions, we

compute the i'th derivative of $a_{t-1|t+1}(\alpha_{t+1})$ with respect to α_{t+1} as

$$a^{(i)}_{t-1|t+1}(\alpha_{t+1}) = \sum_{j=1}^{i} a^{(j)}_{t-1|t}(a_{t|t+1}) B_{i,j}(a^{(1)}_{t|t+1}(\alpha_{t+1}), \ldots, a^{(i-j+1)}_{t|t+1}(\alpha_{t+1})). \tag{I.17}$$

If we substitute $a^{(i)}_{t-1|t+1}(\alpha_{t+1})$ of (I.17) in (I.16) and set $\alpha_{t+1} = a_{t+1}$, we obtain

$$\sum_{i=0}^{r-1} \binom{r-1}{i} \left\{ \bar{\bar{\Omega}}^{(i)}_{t,t-1} \left[\sum_{j=1}^{r-i} a^{(j)}_{t-1} B_{r-i,j}(a^{(1)}_t, \ldots, a^{(r-i-j+1)}_t) \right] + \bar{\bar{\Omega}}^{(i)}_{t,t} a^{(r-i)}_t \right\} = -\bar{\bar{\Omega}}^{(r-1)}_{t,t+1}. \tag{I.18}$$

This gives an expression for $a^{(r)}_t$ in terms of $a^{(i)}_t$, $i = 0, \ldots, r-1$; $a^{(i)}_{t-1}$, $i = 0, \ldots, r$; $\bar{\bar{\Omega}}^{(i)}_{t,t-1}$ and $\bar{\bar{\Omega}}^{(i)}_{t,t}$, $i = 1, \ldots, r-1$; and $\bar{\bar{\Omega}}^{(r-1)}_{t,t+1}$.

We now derive a result that will give us $s^{(r)}_t$ in terms of $a^{(i)}_t$ and $s^{(i)}_t$, $i = 1, \ldots, r-1$ and $a^{(i)}_{t-1}$, $i = 1, \ldots, r+1$. Analogously with equation (I.9), we have

$$\Sigma^{(r)}_{t|t+1}(\alpha_{t+1}) = \Sigma_{t|t+1}(\alpha_{t+1}) B_r(s^{(1)}_{t|t+1}(\alpha_{t+1}), \ldots, s^{(r)}_{t|t+1}(\alpha_{t+1})).$$

Using Leibniz's rule to take derivatives of (I.14) with respect to α_{t+1}, and evaluating at $\alpha_{t+1} = a_{t+1}$, we obtain

$$a^{(r)}_t = \sum_{i=0}^{r-1} \binom{r-1}{i} B_i(s^{(1)}_t, \ldots, s^{(i)}_t) \Sigma_t \bar{\bar{\Omega}}^{(r-1-i)}_{t,t+1}. \tag{I.19}$$

The quantities $\bar{\bar{\Omega}}^{(r)}_{t,s}$ involved in the computation of $a^{(r)}_t$ and $s^{(r)}_t$ are functions of derivatives of $\psi^{(p,q)}_t(a_{t|t+1}, \alpha_{t+1})$ with respect to α_{t+1}, evaluated at a_{t+1}. Equations (I.63) and (I.64) of Appendix I.5 show how to compute these derivatives as functions of derivatives of $\psi^{(p,q)}_t(\alpha_t, \alpha_{t+1})$, supplied as part of the model specification.

I.2.2 Explicit Formula for $R = 5$

We now derive simplified expressions for $a^{(r)}_t$, $r = 1, \ldots, 5$ and $s^{(r)}_t$, $r = 1, \ldots, 4$, for $t = 1, \ldots, n-1$. We give details of the computation for $t = 2, \ldots, n-1$. For the special case $t = 1$, we can obtain analogous results simply by setting any terms with a

time index of zero to zero.

We have already have an expression for $a_t^{(1)}$, $t = 1, \ldots, n-1$, in (I.15). Taking $r = 2$ in (I.18) gives

$$\bar{\bar{\Omega}}_{t,t-1}\left(a_{t-1}^{(1)}a_t^{(2)} + a_{t-1}^{(2)}\left(a_t^{(1)}\right)^2\right) + \bar{\bar{\Omega}}_{t,t}a_t^{(2)} + \bar{\bar{\Omega}}_{t,t-1}^{(1)}a_{t-1}^{(1)}a_t^{(1)} + \bar{\bar{\Omega}}_{t,t}^{(1)}a_t^{(1)} = \bar{\bar{\Omega}}_{t,t+1}^{(1)},$$

which simplifies to

$$a_t^{(2)} = \left(\gamma_t a_t^{(1)}a_{t-1}^{(2)} - \Sigma_t\bar{\bar{\mathbf{\Omega}}}_t^{(1)}\right)a_t^{(1)} - \Sigma_t\bar{\bar{\Omega}}_{t,t+1}^{(1)}, \tag{I.20}$$

where $\gamma_t = -\Sigma_t\bar{\bar{\Omega}}_{t,t-1}$ and $\bar{\bar{\mathbf{\Omega}}}_t^{(i)} = \bar{\bar{\Omega}}_{t,t-1}^{(i)}a_t^{(1)} + \bar{\bar{\Omega}}_{t,t}^{(i)}$. Setting $r = 2$ in (I.19) gives

$$a_t^{(2)} = s_t^{(1)}a_t^{(1)} - \Sigma_t\bar{\bar{\Omega}}_{t,t+1}^{(1)}. \tag{I.21}$$

Equating the right hand sides of (I.20) and (I.21) and solving for $s_t^{(1)}$ gives

$$s_t^{(1)} = \gamma_t a_t^{(1)}a_{t-1}^{(2)} - \Sigma_t\bar{\bar{\mathbf{\Omega}}}_t^{(1)}. \tag{I.22}$$

Setting $r = 3$ in (I.18) gives

$$\begin{aligned}-\bar{\bar{\Omega}}_{t,t+1}^{(2)} =&\bar{\bar{\Omega}}_{t,t-1}\left(a_{t-1}^{(1)}a_t^{(3)} + 3a_{t-1}^{(2)}a_t^{(1)}a_t^{(2)} + a_{t-1}^{(3)}\left(a_t^{(1)}\right)^3\right) + \bar{\bar{\Omega}}_{t,t}a_t^{(3)}\\ &+ 2\left(\bar{\bar{\Omega}}_{t,t-1}^{(1)}\left(a_{t-1}^{(1)}a_t^{(2)} + a_{t-1}^{(2)}\left(a_t^{(1)}\right)^2\right) + \bar{\bar{\Omega}}_{t,t}^{(1)}a_t^{(1)}\right) + \bar{\bar{\Omega}}_{t,t-1}^{(2)}a_{t-1}^{(1)}a_t^{(1)} + \bar{\bar{\Omega}}_{t,t}^{(2)}a_t^{(1)}.\end{aligned}$$

Solving for $a_t^{(3)}$, we obtain

$$\begin{aligned}a_t^{(3)} =&\gamma_t\left(3a_t^{(1)}a_t^{(2)}a_{t-1}^{(2)} + \left(a_t^{(1)}\right)^3a_{t-1}^{(3)}\right) - 2\Sigma_t\left(\bar{\bar{\Omega}}_{t,t-1}^{(1)}\left(a_t^{(1)}\right)^2a_{t-1}^{(2)} + \bar{\bar{\mathbf{\Omega}}}_t^{(1)}a_t^{(2)}\right)\\ &- \Sigma_t\bar{\bar{\mathbf{\Omega}}}_t^{(2)}a_t^{(1)} - \Sigma_t\bar{\bar{\Omega}}_{t,t+1}^{(2)}\\ =&2\left(\gamma_t a_t^{(1)}a_{t-1}^{(2)} - \Sigma_t\bar{\bar{\mathbf{\Omega}}}_t^{(1)}\right)a_t^{(2)} + \left(\gamma_t a_t^{(1)}a_{t-1}^{(3)} - 2\Sigma_t\bar{\bar{\Omega}}_{t,t-1}^{(1)}a_{t-1}^{(2)}\right)\left(a_t^{(1)}\right)^2\\ &+ \left(\gamma_t a_t^{(2)}a_{t-1}^{(2)} - \Sigma_t\bar{\bar{\mathbf{\Omega}}}_t^{(2)}\right)a_t^{(1)} - \Sigma_t\bar{\bar{\Omega}}_{t,t+1}^{(2)}.\end{aligned}$$

We use (I.22) to simplify this to

$$\begin{aligned} a_t^{(3)} =& 2s_t^{(1)}a_t^{(2)} + \left(\gamma_t a_t^{(1)}a_{t-1}^{(3)} - 2\Sigma_t\bar{\bar{\Omega}}_{t,t-1}^{(1)}a_{t-1}^{(2)}\right)\left(a_t^{(1)}\right)^2 \\ &+ \left(\gamma_t a_t^{(2)}a_{t-1}^{(2)} - \Sigma_t\bar{\bar{\mathbf{\Omega}}}_t^{(2)}\right)a_t^{(1)} - \Sigma_t\bar{\bar{\Omega}}_{t,t+1}^{(2)}. \end{aligned} \tag{I.23}$$

Setting $r = 3$ in (I.19) gives an alternative expression for $a_t^{(3)}$:

$$\begin{aligned} a_t^{(3)} &= \left(s_t^{(2)} + \left(s_t^{(1)}\right)^2\right)a_t^{(1)} - \Sigma_t\bar{\bar{\Omega}}_{t,t+1}^{(2)} - 2s_t^{(1)}\Sigma_t\bar{\bar{\Omega}}_{t,t+1}^{(1)} \\ &= \left(s_t^{(2)} + \left(s_t^{(1)}\right)^2\right)a_t^{(1)} - \Sigma_t\bar{\bar{\Omega}}_{t,t+1}^{(2)} + 2s_t^{(1)}\left(a_t^{(2)} - s_t^{(1)}a_t^{(1)}\right) \\ &= \left(s_t^{(2)} - \left(s_t^{(1)}\right)^2\right)a_t^{(1)} + 2s_t^{(1)}a_t^{(2)} - \Sigma_t\bar{\bar{\Omega}}_{t,t+1}^{(2)}. \end{aligned} \tag{I.24}$$

Equating the right hand sides of (I.23) and (I.24) and solving for $s_t^{(2)}$ gives

$$s_t^{(2)} = \left(s_t^{(1)}\right)^2 + \left(\gamma_t a_t^{(1)}a_{t-1}^{(3)} - 2\Sigma_t\bar{\bar{\Omega}}_{t,t-1}^{(1)}a_{t-1}^{(2)}\right)a_t^{(1)} + \left(\gamma_t a_t^{(2)}a_{t-1}^{(2)} - \Sigma_t\bar{\bar{\mathbf{\Omega}}}_t^{(2)}\right). \tag{I.25}$$

We follow a similar procedure to compute the following formulas for $a_t^{(4)}, s_t^{(3)}$, and $a_t^{(5)}, s_t^{(4)}$:

$$\begin{aligned} a_t^{(4)} =& \left(\gamma_t a_t^{(1)}a_{t-1}^{(4)} - 3\Sigma_t\bar{\bar{\Omega}}_{t,t-1}^{(1)}a_{t-1}^{(3)}\right)\left(a_t^{(1)}\right)^3 + 3\left(\gamma_t a_t^{(2)}a_{t-1}^{(3)} - \Sigma_t\bar{\bar{\Omega}}_{t,t-1}^{(2)}a_{t-1}^{(2)}\right)\left(a_t^{(1)}\right)^2 \\ &+ \left(\gamma_t a_t^{(3)}a_{t-1}^{(2)} - 3\Sigma_t\bar{\bar{\Omega}}_{t,t-1}^{(1)}a_t^{(2)}a_{t-1}^{(2)} - \Sigma_t\bar{\bar{\mathbf{\Omega}}}_t^{(3)}\right)a_t^{(1)} - \Sigma_t\bar{\bar{\Omega}}_{t,t+1}^{(3)} \\ &+ 3\left(s_t^{(2)} - \left(s_t^{(1)}\right)^2\right)a_t^{(2)} + 3s_t^{(1)}a_t^{(3)}, \end{aligned} \tag{I.26}$$

$$\begin{aligned} s_t^{(3)} =& -\left(s_t^{(1)}\right)^3 + 3s_t^{(1)}s_t^{(2)} + \left(\gamma_t a_t^{(1)}a_{t-1}^{(4)} - 3\Sigma_t\bar{\bar{\Omega}}_{t,t-1}^{(1)}a_{t-1}^{(3)}\right)\left(a_t^{(1)}\right)^2 \\ &+ 3\left(\gamma_t a_t^{(2)}a_{t-1}^{(3)} - \Sigma_t\bar{\bar{\Omega}}_{t,t-1}^{(2)}a_{t-1}^{(2)}\right)a_t^{(1)} + \left(\gamma_t a_t^{(3)} - 3\Sigma_t\bar{\bar{\Omega}}_{t,t-1}^{(1)}a_t^{(2)}\right)a_{t-1}^{(2)} - \Sigma_t\bar{\bar{\mathbf{\Omega}}}_t^{(3)} \end{aligned} \tag{I.27}$$

$$
\begin{aligned}
a_t^{(5)} = &- \Sigma_t \bar{\bar{\Omega}}_{t,t+1}^{(4)} + \left(\gamma_t a_{t-1}^{(5)} a_t^{(1)} - 4\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_{t-1}^{(4)}\right)\left(a_t^{(1)}\right)^4 \\
&+ 6\left(\gamma_t a_{t-1}^{(4)} a_t^{(2)} - \Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(2)} a_{t-1}^{(3)}\right)\left(a_t^{(1)}\right)^3 \\
&+ 4\left(\gamma_t a_{t-1}^{(3)} a_t^{(3)} - \Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(3)} a_{t-1}^{(2)} - 2\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_{t-1}^{(3)} a_t^{(2)}\right)\left(a_t^{(1)}\right)^2 \\
&+ \left(\gamma_t\left(a_{t-1}^{(2)} a_t^{(4)} + 3a_{t-1}^{(3)}\left(a_t^{(2)}\right)^2\right) - \Sigma_t \bar{\bar{\mathbf{\Omega}}}_t^{(4)} - 6\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(2)} a_{t-1}^{(2)} a_t^{(2)} - 4\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_{t-1}^{(2)} a_t^{(3)}\right) a_t^{(1)} \\
&+ 4s_t^{(1)} a_t^{(4)} + 6\left(s_t^{(2)} - \left(s_t^{(1)}\right)^2\right) a_t^{(3)} + 4\left(s_t^{(3)} + \left(s_t^{(1)}\right)^3 - 3s_t^{(1)} s_t^{(2)}\right) a_t^{(2)},
\end{aligned}
\tag{I.28}
$$

$$
\begin{aligned}
s_t^{(4)} = &\left(\gamma_t a_{t-1}^{(5)} a_t^{(1)} - 4\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_{t-1}^{(4)}\right)\left(a_t^{(1)}\right)^3 \\
&+ 6\left(\gamma_t a_{t-1}^{(4)} a_t^{(2)} - \Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(2)} a_{t-1}^{(3)}\right)\left(a_t^{(1)}\right)^2 \\
&+ 4\left(\gamma_t a_{t-1}^{(3)} a_t^{(3)} - \Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(3)} a_{t-1}^{(2)} - 2\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_{t-1}^{(3)} a_t^{(2)}\right) a_t^{(1)} \\
&+ \left(\gamma_t\left(a_{t-1}^{(2)} a_t^{(4)} + 3a_{t-1}^{(3)}\left(a_t^{(2)}\right)^2\right) - \Sigma_t \bar{\bar{\mathbf{\Omega}}}_t^{(4)} - 6\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(2)} a_{t-1}^{(2)} a_t^{(2)} - 4\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_{t-1}^{(2)} a_t^{(3)}\right) \\
&+ \left(s_t^{(1)}\right)^4 + 4s_t^{(1)} s_t^{(3)} + 3\left(s_t^{(2)} - 2\left(s_t^{(1)}\right)^2\right) s_t^{(2)}.
\end{aligned}
\tag{I.29}
$$

I.3 Polynomial approximations of $b_t^{(r)}$ and $\mu_t^{(r)}$

I.3.1 First derivative of $\log f(\alpha_t|\alpha_{t+1}, y)$

In this subsection, we derive an exact expression for $h_t^{(1)}(\alpha_t; \alpha_{t+1})$, the first derivative of $\log f(\alpha_t|\alpha_{t+1}, y)$ with respect to α_t.

The case $t = 1$ is straightforward using Bayes' rule. We have

$$
\frac{\partial \log f(\alpha_1|\alpha_2, y)}{\partial \alpha_1} = \frac{\partial \log f(y_1|\alpha_1, \alpha_2)}{\partial \alpha_1} + \frac{\partial \log f(\alpha_2, \alpha_1)}{\partial \alpha_1}
$$

Recalling the definition of $\psi_t^{(p,q)}(\alpha_t, \alpha_{t+1})$ in (1.7), the first derivative of $h_1(\alpha_1; \alpha_2) =$

$\log f(y_1|\alpha_1, \alpha_2)$ is written

$$h_1^{(1)}(\alpha_1; \alpha_2) = \psi_1^{(1,0)}(\alpha_1, \alpha_2) + \bar{c}_1 - \bar{\Omega}_{1,2}\alpha_2 - \bar{\Omega}_{1,1}\alpha_1. \tag{I.30}$$

For $t = 2, \ldots, n-1$, we compute $f(\alpha_t|\alpha_{t+1}, y)$ as a marginal density of $f(\alpha_{1:t}|\alpha_{t+1}, y)$. Thus, we have

$$\begin{aligned} f(\alpha_t|\alpha_{t+1}, y) &= \int f(\alpha_{1:t-1}, \alpha_t|\alpha_{t+1}, y)\, d\alpha_{1:t-1} \\ &\propto f(\alpha_{t+1}|\alpha_t) f(y_t|\alpha_t, \alpha_{t+1}) c(\alpha_t), \end{aligned} \tag{I.31}$$

where

$$c(\alpha_t) = \int f(\alpha_t|\alpha_{t-1}) f(y_{t-1}|\alpha_{t-1}, \alpha_t) f(y_{1:t-2}, \alpha_{1:t-1})\, d\alpha_{1:t-1}.$$

Taking the logarithm of (I.31) and differentiating with respect to α_t gives

$$h_t^{(1)}(\alpha_t; \alpha_{t+1}) = \frac{\partial \log c(\alpha_t)}{\partial \alpha_t} + \frac{\partial \log f(\alpha_{t+1}|\alpha_t)}{\partial \alpha_t} + \frac{\partial \log f(y_t|\alpha_t, \alpha_{t+1})}{\partial \alpha_t}. \tag{I.32}$$

We use a development similar to Appendix C of **?**] to show that

$$\frac{\partial \log c(\alpha_t)}{\partial \alpha_t} = E\left[\frac{\partial \log f(\alpha_t|\alpha_{t-1})}{\partial \alpha_t} + \frac{\partial \log f(y_{t-1}|\alpha_{t-1}, \alpha_t)}{\partial \alpha_t} \middle| \alpha_t, y \right].$$

The first derivatives $h_t(\alpha_t; \alpha_{t+1})$ then becomes

$$\begin{aligned} h_t^{(1)}(\alpha_t; \alpha_{t+1}) =& E\left[\frac{\log f(\alpha_t|\alpha_{t-1})}{\partial \alpha_t} + \frac{\log f(y_{t-1}|\alpha_{t-1}, \alpha_t)}{\partial \alpha_t} \middle| \alpha_t, \alpha_{t+1}, y \right] \\ &+ \frac{\partial \log f(\alpha_{t+1}|\alpha_t)}{\partial \alpha_t} + \frac{\partial \log f(y_t|\alpha_t, \alpha_{t+1})}{\partial \alpha_t} \\ =& E\left[\frac{\log f(\alpha_t|\alpha_{t-1})}{\partial \alpha_t} + \frac{\log f(\alpha_{t+1}|\alpha_t)}{\partial \alpha_t} \middle| \alpha_t, \alpha_{t+1}, y \right] \\ &+ E\left[\frac{\log f(y_{t-1}|\alpha_{t-1}, \alpha_t)}{\partial \alpha_t} \middle| \alpha_t, \alpha_{t+1}, y \right] + \frac{\partial \log f(y_t|\alpha_t, \alpha_{t+1})}{\partial \alpha_t}. \end{aligned}$$

The first term above simplifies as in Appendix C of McCausland (2010). We use (1.7) to

finally derive

$$\begin{aligned} h_t^{(1)}(\alpha_t;\alpha_{t+1}) =& \bar{c}_t - \bar{\Omega}_{t,t}\alpha_t - \bar{\Omega}_{t,t+1}\alpha_{t+1} + \psi_t^{(1,0)}(\alpha_t,\alpha_{t+1}) \\ & - \bar{\Omega}_{t-1,t}\mu_{t-1|t}(\alpha_t) + x_{t-1|t}(\alpha_t), \end{aligned} \tag{I.33}$$

where $\mu_{t-1|t}(\alpha_t) = E[\alpha_{t-1}|\alpha_t, y]$ and $x_{t-1}(\alpha_t) = E\left[\psi_{t-1}^{(0,1)}(\alpha_{t-1},\alpha_t)\,|\alpha_t, y\right]$. The case $t = n$ is similar, and we obtain

$$h_n^{(1)}(\alpha_n) = \bar{c}_n - \bar{\Omega}_{n,n}\alpha_n + \psi_n^{(1)}(\alpha_n) - \bar{\Omega}_{n-1,n}\mu_{n-1|n}(\alpha_n) + x_{n-1|n}(\alpha_n). \tag{I.34}$$

I.3.2 Approximation of $h_t^{(1)}(\alpha_t;\alpha_{t+1})$

Since we do not know the conditional expectations $\mu_{t-1|t}(\alpha_t)$ and $x_{t-1|t}(\alpha_t)$, we cannot compute $h_t(\alpha_t;\alpha_{t+1})$ exactly. We propose an approximation $H_t^{(1)}(\alpha_t;\alpha_{t+1})$ of $h_t^{(1)}(\alpha_t;\alpha_{t+1})$. For $t = 2,\ldots,n-1$, we have

$$H_t^{(1)}(\alpha_t;\alpha_{t+1}) \doteq \bar{c}_t - \bar{\Omega}_{t,t}\alpha_t - \bar{\Omega}_{t,t+1}\alpha_{t+1} + \Psi_t^{(1,0)}(\alpha_t,\alpha_{t+1}) - \bar{\Omega}_{t-1,t}M_{t-1|t}(\alpha_t) + X_{t-1|t}(\alpha_t) \tag{I.35}$$

where $M_{t-1|t}(\alpha_t)$ is an approximation of $\mu_{t-1|t}(\alpha_t)$, $X_{t-1|t}(\alpha_t)$ is an approximation of $x_{t-1|t}(\alpha_t)$ and $\Psi_t^{(1,0)}(\alpha_t,\alpha_{t+1})$ is an approximation of $\psi_t^{(1,0)}(\alpha_t,\alpha_{t+1})$.[1] The polynomials $M_{t-1|t}(\alpha_t)$ and $\Psi_t^{(p,q)}(\alpha_t,\alpha_{t+1})$ are defined in (1.13) and (1.14).

We construct $X_{t-1|t}(\alpha_t)$ in two steps. First, we approximate $\psi_{t-1}^{(0,1)}(\alpha_{t-1},\alpha_t)$, as a function of α_{t-1}, by its second order Taylor series expansion around $a_{t-1|t}(\alpha_t)$:

$$\begin{aligned} \psi_{t-1}^{(0,1)}(\alpha_{t-1},\alpha_t) \approx & \psi_{t-1}^{(0,1)}(a_{t-1|t}(\alpha_t),\alpha_t) + \psi_{t-1}^{(1,1)}(a_{t-1|t}(\alpha_t),\alpha_t)(\alpha_{t-1} - a_{t-1|t}(\alpha_t)) \\ & + \tfrac{1}{2}\psi_{t-1}^{(2,1)}(a_{t-1|t}(\alpha_t),\alpha_t)(\alpha_{t-1} - a_{t-1|t}(\alpha_t))^2. \end{aligned} \tag{I.36}$$

Taking the conditional expectation of both sides of (I.36), given α_t and y, and using $\Sigma_{t-1|t}(\alpha_t)$ as an approximation of $E\left[(\alpha_{t-1} - a_{t-1|t}(\alpha_t))^2|\alpha_t, y\right]$ gives the approxima-

1. For $t = n$, we need just to replace $\Psi_t^{(1,0)}(\alpha_t,\alpha_{t+1})$ by $\Psi_n^{(1)}(\alpha_n)$ in (I.35) to obtain $H_n^{(1)}(\alpha_n)$, the approximation of $h_n^{(1)}(\alpha_n)$.

tion

$$x_{t-1|t}(\alpha_t) \approx \psi_{t-1}^{(0,1)}(a_{t-1|t}(\alpha_t), \alpha_t) + \psi_{t-1}^{(1,1)}(a_{t-1|t}(\alpha_t), \alpha_t)(\mu_{t-1|t}(\alpha_t) - a_{t-1|t}(\alpha_t)) + \frac{1}{2}\psi_{t-1}^{(2,1)}(a_{t-1|t}(\alpha_t), \alpha_t)\Sigma_{t-1|t}(\alpha_t). \tag{I.37}$$

Now we define the polynomial $X_{t-1|t}(\alpha_t)$ as the R'th order Taylor series expansion of the right hand side of (I.37) :

$$X_{t-1|t}(\alpha_t) \doteq \sum_{r=0}^{R} \frac{X_{t-1}^{(r)}}{r!}(\alpha_t - a_t)^r, \tag{I.38}$$

where $X_{t-1}^{(r)}$ is the r'th derivative of the RHS of (I.37) with respect to α_t, evaluated at a_t. We evaluate these derivatives bottom up using Faà Di Bruno's formula, equations (I.61) and (I.62), and Leibniz's rule, equation (I.57).

I.3.3 Approximation of the conditional mode $b_{t|t+1}(\alpha_{t+1})$

Recall that $b_{t|t+1}(\alpha_{t+1})$ is the conditional mode of α_t given α_{t+1} and y. We provide an approximation $B_{t|t+1}(\alpha_{t+1})$ of the Taylor expansion of $b_{t|t+1}(\alpha_{t+1})$ around $\alpha_{t+1} = a_{t+1}$. We show in this subsection how to compute the coefficients of the resulting polynomial. The degree of this polynomial is $R - 1 = 4$.

By definition, $b_{t|t+1}(\alpha_{t+1})$ is the root of $h_t^{(1)}(\alpha_t; \alpha_{t+1}) = 0$. We can approximate this root, as a function of α_{t+1}, using one iteration of the Newton-Raphson algorithm for root finding, from the starting point $a_{t|t+1}(\alpha_{t+1})$:

$$b_{t|t+1}(\alpha_{t+1}) \approx a_{t|t+1}(\alpha_{t+1}) - \frac{h_t^{(1)}(a_{t|t+1}(\alpha_{t+1}); \alpha_{t+1})}{h_t^{(2)}(a_{t|t+1}(\alpha_{t+1}); \alpha_{t+1})}. \tag{I.39}$$

We want to approximate the function $b_{t|t+1}(\alpha_{t+1})$, not just perform the Newton-Raphson step for a particular value of $a_{t|t+1}$. Our strategy will be to find an approximate Taylor expansion of the second term of the right hand side around $\alpha_{t+1} = a_{t+1}$.

Our approximations of numerator and denominator are, using (I.35) and its deriva-

tive, both evaluated at $\alpha_t = a_{t|t+1}(\alpha_{t+1})$, are

$$H_t^{(1)}(a_{t|t+1};\alpha_{t+1}) = \bar{c}_t - \bar{\Omega}_{t,t}a_{t|t+1} - \bar{\Omega}_{t,t+1}\alpha_{t+1} + \Psi_t^{(1,0)}(a_{t|t+1},\alpha_{t+1}) \\ - \bar{\Omega}_{t-1,t}M_{t-1|t}(a_{t|t+1}) + X_{t-1|t}(a_{t|t+1}) \tag{I.40}$$

$$H_t^{(2)}(a_{t|t+1};\alpha_{t+1}) = -\bar{\Omega}_{t,t} + \Psi_t^{(2,0)}(a_{t|t+1},\alpha_{t+1}) \\ - \bar{\Omega}_{t-1,t}M_{t-1|t}^{(1)}(a_{t|t+1}) + X_{t-1|t}^{(1)}(a_{t|t+1}), \tag{I.41}$$

where we suppress the argument of $a_{t|t+1}(\alpha_{t+1})$ to write $a_{t|t+1}$.

We compute total derivatives of $H_t^{(1)}(a_{t|t+1}(\alpha_{t+1});\alpha_{t+1})$ and $H_t^{(2)}(a_{t|t+1}(\alpha_{t+1});\alpha_{t+1})$ at $\alpha_{t+1} = a_{t+1}$ using Faà di Bruno's formula to compute the derivatives of $M_{t-1|t}(a_{t|t+1}(\alpha_{t+1}))$, $a_{t-1|t}(a_{t|t+1}(\alpha_{t+1}))$ and $X_{t-1|t}(a_{t|t+1}(\alpha_{t+1}))$ with respect to α_{t+1}, at $\alpha_{t+1} = a_{t+1}$.

Based on equation (I.39), we define the following approximations $B_t^{(r)}$ of $b_t^{(r)}$, $r = 0,1,2,3$:

$$B_t^{(r)} \doteq a_t^{(r)} - \frac{\partial^r}{\partial \alpha_{t+1}^r}\left(\frac{H_t^{(1)}(a_{t|t+1}(\alpha_{t+1});\alpha_{t+1})}{H_t^{(2)}(a_{t|t+1}(\alpha_{t+1});\alpha_{t+1})}\right)\Bigg|_{\alpha_{t+1}=a_{t+1}}. \tag{I.42}$$

The second term on the right hand side of (I.42) is the r'th order derivative of a quotient, which we compute using the quotient rule for derivatives, equation (I.58) in Appendix I.5.

In practice, we find that going beyond a third order approximation of $b_{t|t+1}(\alpha_{t+1}) - a_{t|t+1}(\alpha_{t+1})$ does not justify the computational cost and so we set $B_t^{(4)} = a_t^{(4)}$.

For $t = n$, we approximate a value b_n, not a function. We define, analogously, the following approximation of b_n :

$$B_n \doteq a_n - \frac{H_n^{(1)}(a_n)}{H_n^{(2)}(a_n)}. \tag{I.43}$$

I.3.4 Coefficients of the polynomial approximation of $\mu_{t|t+1}(\alpha_{t+1})$

Recall that $\mu_{t|t+1}(\alpha_{t+1}) = E[\alpha_t|\alpha_{t+1}, y]$. We provide an approximation $M_{t|t+1}(\alpha_{t+1})$ of a Taylor expansion of $\mu_{t|t+1}(\alpha_{t+1})$ around $\alpha_{t+1} = a_{t+1}$. We show in this subsection how to compute the coefficients of the resulting fourth order polynomial.

McCausland(2011) suggests the following approximation for $\mu_{t|t+1} - b_{t|t+1}$:

$$\mu_{t|t+1} - b_{t|t+1} \approx \frac{1}{2} h_t^{(3)}(b_{t|t+1}; \alpha_{t+1}) \left[h_t^{(2)}(b_{t|t+1}; \alpha_{t+1}) \right]^{-2} \tag{I.44}$$

As the mode $b_{t|t+1}$ is the root of $h_t^{(1)}(\alpha_t; \alpha_{t+1})$, we have

$$h_t^{(1)}(b_{t|t+1}; \alpha_{t+1}) = 0 \tag{I.45}$$

Taking the derivative of (I.45) two times with respect to α_{t+1} gives

$$h_t^{(2)}(b_{t|t+1}; \alpha_{t+1}) b_{t|t+1}^{(1)} = \bar{\Omega}_{t,t+1} - \psi_t^{(1,1)}(b_{t|t+1}, \alpha_{t+1}) \tag{I.46}$$

and

$$\begin{aligned} h_t^{(3)}(b_{t|t+1}; \alpha_{t+1}) \left(b_{t|t+1}^{(1)} \right)^2 + h_t^{(2)}(b_{t|t+1}; \alpha_{t+1}) b_{t|t+1}^{(2)} = & -2 \frac{d\psi_t^{(1,1)}(b_{t|t+1}, \alpha_{t+1})}{d\alpha_{t+1}} \\ & + \psi_t^{(1,1)}(b_{t|t+1}, \alpha_{t+1}) \end{aligned} \tag{I.47}$$

Solve for $h_t^{(3)}(b_{t|t+1}; \alpha_{t+1})$ in equation (I.47) and divide by the square of $h_t^{(2)}(b_{t|t+1}; \alpha_{t+1})$ to obtain

$$\begin{aligned} \frac{h_t^{(3)}(b_{t|t+1}; \alpha_{t+1})}{\left(h_t^{(2)}(b_{t|t+1}; \alpha_{t+1}) \right)^2} = & - \frac{b_{t|t+1}^{(2)} / b_{t|t+1}^{(1)}}{h_t^{(2)}(b_{t|t+1}; \alpha_{t+1}) b_{t|t+1}^{(1)}} \\ & - \frac{2 d\psi_t^{(1,1)}(b_{t|t+1}, \alpha_{t+1}) / d\alpha_{t+1} - \psi_t^{(1,1)}(b_{t|t+1}, \alpha_{t+1})}{\left(h_t^{(2)}(b_{t|t+1}; \alpha_{t+1}) b_{t|t+1}^{(1)} \right)^2} \end{aligned} \tag{I.48}$$

Substitute the right hand side of equation (I.46) in (I.48) to obtain

$$\begin{aligned}\mu_{t|t+1} - b_{t|t+1} \approx & -\frac{1}{2}\frac{b^{(2)}_{t|t+1}/b^{(1)}_{t|t+1}}{\bar{\Omega}_{t,t+1} - \psi^{(1,1)}_t(b_{t|t+1}, \alpha_{t+1})} \\ & -\frac{1}{2}\frac{2d\psi^{(1,1)}_t(b_{t|t+1}, \alpha_{t+1})/d\alpha_{t+1} - \psi^{(1,1)}_t(b_{t|t+1}, \alpha_{t+1})}{\left(\bar{\Omega}_{t,t+1} - \psi^{(1,1)}_t(b_{t|t+1}, \alpha_{t+1})\right)^2}\end{aligned} \tag{I.49}$$

Based on equation (I.49), we define our approximation $M_{t|t+1}$ of $\mu_{t|t+1}$ as the Taylor series expansion of :

$$-\frac{1}{2}\frac{B^{(2)}_{t|t+1}/B^{(1)}_{t|t+1}}{\bar{\Omega}_{t,t+1} - \Psi^{(1,1)}_t(B_{t|t+1}, \alpha_{t+1})} - \frac{1}{2}\frac{2d\Psi^{(1,1)}_t(B_{t|t+1}, \alpha_{t+1})/d\alpha_{t+1} - \Psi^{(1,1)}_t(B_{t|t+1}, \alpha_{t+1})}{\left(\bar{\Omega}_{t,t+1} - \Psi^{(1,1)}_t(B_{t|t+1}, \alpha_{t+1})\right)^2} \tag{I.50}$$

The derivatives of $B^{(2)}_{t|t+1}/B^{(1)}_{t|t+1}$ with respect to α_{t+1} are computed using the quotient rule for derivatives, equation (I.58). Those of $\Psi^{(1,1)}_t(B_{t|t+1}, \alpha_{t+1})$ and $d\Psi^{(1,1)}_t(B_{t|t+1}, \alpha_{t+1})/d\alpha_{t+1}$ are computed using the Faà-Di-Bruno formula, equations (I.61) and (I.62). Derivatives of the two main ratios in (I.50) are computed using the quotient rule in equation (I.58). We compute $M^{(r)}_t = M^{(r)}_{t|t+1}(a_{t+1})$, $r = 0, 1, 2$ using (I.50).

In practice, we find that going beyond a second order approximation of $\mu_{t|t+1}(\alpha_{t+1}) - b_{t|t+1}(\alpha_{t+1})$ does not justify the computational cost and so we set $M^{(3)}_t = B^{(3)}_t$ and $M^{(4)}_t = a^{(4)}_t$.

I.4 Model derivatives

Here we show how to compute partial derivatives of $\psi_t(\alpha_t, \alpha_{t+1})$ and derivatives $\psi_n(\alpha_n)$, for the ASV-Gaussian and ASV-Student models. In our empirical applications, we compute $\psi^{(p,q)}_t(\alpha_t, \alpha_{t+1})$ up to orders $P = 7$ and $Q = 7$ and $\psi^{(p)}_n(\alpha_n)$ up to order $P = 7$.

I.4.1 ASV-Gaussian

Using (1.5), we can write

$$\psi_t(\alpha_t, \alpha_{t+1}) = -\frac{1}{2}\left[\log(2\pi/\beta) + \alpha_t + \beta(\varphi_t - \theta u_t)^2\right], \qquad t = 1, \ldots, n-1, \quad \text{(I.51)}$$

$$\psi_n(\alpha_n) = -\frac{1}{2}\left[\log(2\pi) + \alpha_n + \varphi_n^2\right], \quad \text{(I.52)}$$

where $\beta \doteq (1-\rho^2)^{-1}$, $\theta \doteq \rho/\sigma$, $u_t \doteq \alpha_{t+1} - d_t - \phi\alpha_t$ and $\varphi_t \doteq y_t \exp(-\alpha_t/2)$.

For $t = 1, \ldots, n-1$ and $(p, q) \neq (0, 0)$ we have

$$\psi_t^{(p,q)}(\alpha_t, \alpha_{t+1}) = \begin{cases} -\frac{1}{2} - \frac{\beta}{2}\left(\tilde{\varphi}_{t,p} - 2\theta^2\phi u_t\right) & q = 0,\ p = 1 \\ -\frac{\beta}{2}\left(\tilde{\varphi}_{t,p} + 2\theta^2\phi^2\right) & q = 0,\ p = 2 \\ -\frac{\beta}{2}\tilde{\varphi}_{t,p} & q = 0, p \geq 3 \\ \beta\theta\left(\varphi_t - \theta u_t\right) & q = 1,\ p = 0 \\ \beta\theta\left(-\frac{1}{2}\varphi_t + \theta\phi\right) & q = 1,\ p = 1 \\ \beta\theta\left(-\frac{1}{2}\right)^p \varphi_t & q = 1,\ p \geq 2 \\ -\beta\theta^2 & q = 2,\ p = 0 \\ 0 & \text{otherwise,} \end{cases} \quad \text{(I.53)}$$

where

$$\tilde{\varphi}_{t,p} \doteq (-1)^p\varphi_t^2 - \left(-\frac{1}{2}\right)^{p-2}\theta\varphi_t\left(p\phi + \frac{1}{2}u_t\right), \qquad t = 1, \ldots, n-1. \quad \text{(I.54)}$$

For $t = n$,

$$\psi_n^{(p)}(\alpha_n)(\alpha_n) = \begin{cases} -\frac{1}{2} - \frac{1}{2}\tilde{\varphi}_{n,p} & p = 1 \\ -\frac{1}{2}\tilde{\varphi}_{n,p} & p \geq 2, \end{cases} \quad \text{(I.55)}$$

where

$$\tilde{\varphi}_{n,p} = (-1)^p\varphi_n^2.$$

I.4.2 ASV-Student

We use the definitions of β, θ, u_t and φ_t from I.4.1. Using (1.19) we can write $\psi_t(\alpha_t, \alpha_{t+1})$, for $t = 1, \ldots, n-1$, as

$$\psi_t(\alpha_t, \alpha_{t+1}) = k + \psi_{1,t}(\alpha_t, \alpha_{t+1}) + \psi_{2,t}(\alpha_t) + \psi_{3,t}(\alpha_t, \alpha_{t+1}), \tag{I.56}$$

where k does not depend on α_t and α_{t+1},

$$\begin{aligned}
\psi_{1,t}(\alpha_t, \alpha_{t+1}) &\doteq -\frac{1}{2}(\theta^2 \beta u_t^2 + \alpha_t), & \psi_{2,t}(\alpha_t) &\doteq -(\nu+1)\log d(\alpha_t), \\
\psi_{3,t}(\alpha_t, \alpha_{t+1}) &\doteq \log m(z(\alpha_t, \alpha_{t+1})), & m(z) &= 2\frac{\Gamma\left(\frac{\nu}{2}+1\right)}{\Gamma\left(\frac{\nu+1}{2}\right)} z m_1(z) + m_2(z), \\
m_1(z) &= M\left(\frac{\nu}{2}+1; \frac{3}{2}; z^2\right), & m_2(z) &= M\left(\frac{\nu+1}{2}; \frac{1}{2}; z^2\right), \\
z(\alpha_t, \alpha_{t+1}) &= \frac{n(\alpha_t, \alpha_{t+1})}{d(\alpha_t)}, & n(\alpha_t, \alpha_{t+1}) &= \frac{\theta\beta}{\sqrt{2\nu}} u_t \varphi_t,\ d(\alpha_t) = \sqrt{1 + \frac{\beta}{\nu}\varphi_t^2}.
\end{aligned}$$

Computing analytical expressions for high order partial derivatives of $\psi_t(\alpha_t, \alpha_{t+1})$ is daunting, but fortunately we can avoid it. All we need to do is evaluate the derivatives at a given point (α_t, α_{t+1}), and for this, we can use general purpose routines to combine derivatives of products, quotients and composite functions.

We first compute the derivatives of the third component $\psi_{3,t}(\alpha_t, \alpha_{t+1})$ of the log-density of the ASV-Student model. We do it bottom up using the following steps :

1. Evaluate $n(\alpha_t, \alpha_{t+1})$ and its derivatives with respect to α_t and α_{t+1} up to orders P and Q :

$$n^{(p,q)}(\alpha_t, \alpha_{t+1}) = \begin{cases} \frac{\beta\theta}{\sqrt{2\nu}} \left(-\frac{1}{2}\right)^p (2p\phi + u_t)\, \varphi_t & p \geq 0, q = 0 \\ \frac{\beta\theta}{\sqrt{2\nu}} \left(-\frac{1}{2}\right)^p \varphi_t & p \geq 0, q = 1 \\ 0 & p \geq 0, q \geq 2. \end{cases}$$

2. Evaluate derivatives of $(1 + \beta/\nu\varphi_t^2(\alpha_t))$ with respect to α_t up to order P :

$$\frac{d^p}{d\alpha_t}\left(1 + \frac{\beta}{\nu}\varphi_t^2(\alpha_t)\right) = (-1)^p\frac{\beta}{\nu}\varphi_t^2(\alpha_t), \quad p = 0, \ldots, P.$$

3. Evaluate $d(\alpha_t)$ and its derivatives with respect to α_t, up to order P. Use derivatives of the square root function, evaluated at $(1 + \beta/\nu\varphi_t^2(\alpha_t))$ and the derivatives evaluated in step 2, combining them using Faà Di Bruno's formula, equations (I.61) and (I.62).

4. Evaluate $z = n/d$ and partial derivatives $z^{(p,q)}(\alpha_t, \alpha_{t+1})$ up to order P and Q. Use the value n and partial derivatives $n^{(p,q)}(\alpha_t, \alpha_{t+1})$ computed at step (1), as well as the value d and derivatives $d^{(p)}(\alpha_t)$ computed at step (3). For each $p = 1, \ldots, P$, compute $z^{(p,q)}(\alpha_t, \alpha_{t+1})$ using the quotient rule, equation (I.58).

5. Evaluate $M(\nu/2 + 1, 3/2, x)$ and partial derivatives $M^{(0,0,p)}(\nu/2, 3/2, x)$ up to order P. We use the property $M^{(0,0,p)}(a, b, x) = (a)_k/(b)_k M(a + k, b + k, x)$ and compute values of $M(a, b, x)$ using the routine `gsl_sf_hyperg_1F1` in the GNU scientific library. Similarly, compute $M((\nu + 1)/2, 1/2, x)$ and partial derivatives $M^{(0,0,p)}((\nu + 1)/2; 1/2; x)$ up to order P.

6. Set $m_1(z) = M(\nu/2+1, 3/2, z^2)$ and compute P derivatives of $m_1(z)$ with respect to z. Use P derivatives of $M(\nu/2 + 1, 3/2, x)$ with respect to x, computed in step 5 and P derivatives (only 2 are non-zero) of $x = z^2$ with respect to z, evaluated at z, combining them using the Faà Di Bruno's rule, equations (I.61) and (I.62). Similarly, set $m_2(z) = M((\nu+1)/2, 1/2, z^2)$ and evaluate P derivatives of $m_2(z)$ with respect to z.

7. Evaluate P derivatives of $m(z)$ with respect to z using the derivatives evaluated at step 6, combining them according to

$$m^{(p)}(z) = 2\frac{\Gamma\left(\frac{\nu}{2} + 1\right)}{\Gamma\left(\frac{\nu+1}{2}\right)}\left(zm_1^{(p)}(z) + rm_1^{(p-1)}(z)\right) + m_2^{(p)}(z), \quad p = 1, \ldots, P.$$

8. Evaluate P derivatives of $\log m(z)$ with respect to z using the derivatives evaluated at step 7, and the logarithm rule, equations (I.59) and (I.60).

9. Evaluate partial derivatives of $\psi_{3,t}(\alpha_t, \alpha_{t+1})$ up to orders P and Q. Use derivatives of $\log m(z)$ with respect to z computed in step 8 and partial derivatives of $z(\alpha_t, \alpha_{t+1})$ computed in step 4, combining them according to the multivariate Faa-Di-Bruno rule defined in equations (I.65) and (I.66).

The first component, $\psi_{1,t}(\alpha_t, \alpha_{t+1})$, is a quadratic function of α_t and α_{t+1}. Its derivatives, for $(p, q) \neq (0, 0)$ are

$$\psi_{1,t}^{(p,q)}(\alpha_t, \alpha_{t+1}) = \begin{cases} -\frac{1}{2}\theta^2\beta u_t & p = 0,\ q = 1, \\ -\frac{1}{2}\theta^2\beta & p = 0,\ q = 2, \\ -\frac{1}{2}(-\phi\theta^2\beta u_t + 1) & p = 1,\ q = 0, \\ \frac{1}{2}\phi\theta^2\beta & p = 1,\ q = 1, \\ -\frac{1}{2}\phi^2\theta^2\beta & p = 2,\ q = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Recall that $\psi_{2,t}(\alpha_t) = -(\nu+1)\log d(\alpha_t)$. We compute derivatives of $\log d(\alpha_t)$ using the log rule in equations (I.59) and (I.60). Derivatives of $\psi_{2,t}(\alpha_t)$ are simply $-(\nu+1)$ times the derivatives of $\log d(\alpha_t)$.

The special case of $t = n$ is easily handled. We have

$$\psi_n(\alpha_n) = \log \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi}} - \frac{1}{2}\left[\alpha_n + (\nu+1)\log\left(1 + \frac{\varphi_n^2}{\nu}\right)\right],$$

whose derivatives are the same as those of $\psi_{2,t}$ except for β replaced by 1.

I.5 Rules for derivatives of compound functions

In this paper, we make extensive use of automatic rules for evaluating multiple derivatives of compound functions at a point. These rules combine multiple derivatives of component functions, also evaluated at points. This Appendix gathers these rules in one place.

For univariate functions f and g, we give well known rules for multiple derivatives of

the product fg, the quotient f/g, and the composition $f \circ g$. We give a rule for multiple derivatives of $\log g$, a special case where we exploit the properties of the logarithmic function to simplify computations. We also give derivatives of $f \circ g$ for $f\colon \mathbb{R} \to \mathbb{R}$ and $g\colon \mathbb{R}^2 \to \mathbb{R}$ and partial derivatives of $f \circ g$ for $f\colon \mathbb{R}^2 \to \mathbb{R}$ and $g\colon \mathbb{R} \to \mathbb{R}^2$.

We have coded all of these rules as computer routines. Values passed to these routines are vectors (or matrices) giving multiple derivatives (or partial derivatives) of f and g, evaluated at particular points. The routines return a vector (or a matrix) giving multiple derivatives (or partial derivatives) of a compound function, evaluated at a point. For example, the routine computing P derivatives of the product function fg at a point x takes as input the integer P, a P-vector with the first P derivatives of f at x and a P-vector with the first P derivatives of g at x. It returns a P-vector with the first P derivatives of fg at x.

I.5.1 Univariate functions

For the first three rules, let x be a point in $\mathbb{R}$ and let f and g be two univariate functions, continuously differentiable at x up to order P.

Leibniz rule for products

The product fg is differentiable up to order P at x and

$$(fg)^{(p)}(x) = \sum_{r=0}^{p} \binom{p}{r} f^{(r)}(x) g^{(p-r)}(x), \quad p = 1, \ldots, P. \tag{I.57}$$

We have a routine taking the first P derivatives of f at x and the first P derivatives of g at x and returning the first P derivatives of fg at x.

Quotient rule

Applying Leibniz' rule to the product of f/g and g gives the recursive rule

$$(f/g)^{(p)}(x) = \frac{1}{g(x)}\left[f^{(p)}(x) - \sum_{r=0}^{p-1}\binom{p}{r}(f/g)^{(r)}(x)g^{(p-r)}(x)\right], \quad p = 1, \ldots, P. \tag{I.58}$$

We have a routine taking the first P derivatives of f at x and the first P derivatives of g at x and returning the first P derivatives of f/g at x.

Log rule

We consider the compound function $h = \log f$ and suppose that $f(x) > 0$. Then the function h is differentiable up to order P. Applying the quotient rule to

$$h^{(1)}(x) = \frac{f^{(1)}(x)}{f(x)} \tag{I.59}$$

gives

$$h^{(p)}(x) = \frac{1}{f(x)}\left[f^{(p)}(x) - \sum_{r=1}^{p-1}\binom{p-1}{r-1}h^{(p)}(x)f^{(p-r)}(x)\right], \quad p = 2, \ldots, P. \tag{I.60}$$

Together, equations (I.59) and (I.60) give the first P derivatives of $\log(f(x))$. We have a routine taking the first P derivatives of f at x and returning the first P derivatives of $\log f$ at x.

Faà di Bruno's rule for composite functions

Now suppose that x is a point in $\mathbb{R}$, g is a univariate function, P times differentiable at x, and f is a univariate function, P times differentiable at $g(x)$. Faà di Bruno's rule gives the p'th derivative of $f \circ g$ at x as

$$(f \circ g)^{(p)}(x) = \sum_{r=1}^{p} f^{(r)}(g(x))B_{p,r}(g^{(1)}(x), \ldots, g^{(p-r+1)}(x)), \tag{I.61}$$

where the $B_{p,r}(z_1, \ldots, z_{p-r+1})$ are Bell polynomials. The Bell polynomials are a triangular array of polynomials that can be computed using the boundary conditions $B_{0,0}(z_1) = 1$ and $B_{p,0}(z_1, \ldots, z_{p+1}) = 0$, $p > 0$, and the recursion

$$B_{p,r}(z_1, \ldots, z_{p-r+1}) = \sum_{i=r-1}^{p-1} \binom{p-1}{i} z_{p-i} B_{i,r-1}(z_1, \ldots, z_{i-r}), \quad r = 1, \ldots, p. \quad \text{(I.62)}$$

For example, we have $B_{1,1}(z_1) = z_1 B_{0,0}(z_1) = z_1$, which gives $(f \circ g)^{(1)}(x) = f^{(1)}(g(x))g^{(1)}(x)$, the chain rule. For the second derivative, we compute $B_{2,1}(z_1, z_2) = z_2 B_{0,0}(z_1) + z_1 B_{1,0}(z_1, z_2) = z_2$ and $B_{2,2}(z_1) = z_1 B_{1,1}(z_1) = z_1^2$, which gives

$$(f \circ g)^{(2)}(x) = f^{(1)}(g(x))g^{(2)}(x) + f^{(2)}(g(x)) \left(g^{(1)}(x)\right)^2.$$

We have a routine taking the first P derivatives of g at x and the first P derivatives of f at $g(x)$, returning the first P derivatives of $f \circ g$ at x.

I.5.2 Multivariate functions

?] generalizes Faà di Bruno's rule to multivariate functions. Equations (3.1) and (3.5) in that paper give multiple partial derivatives of $f \circ g$, where $f \colon \mathbb{R}^m \to \mathbb{R}$ and $g \colon \mathbb{R}^d \to \mathbb{R}^m$. We are only concerned with two special cases here, and we describe below how to compute partial derivatives for these cases.

Case $d = 1$ **and** $m = 2$

Here $(f \circ g)(x) = f(g_1(x), g_2(x))$, where f is a scalar valued function with continuous partial derivatives up to orders P and P, and g_1 and g_2 are scalar-valued functions, continuously differentiable up to order P. The value of the p'th derivative of $f \circ g$ at is

$$(f \circ g)^{(p)}(x) = \sum_{r=0}^{p} \sum_{s=\max\{0,1-r\}}^{p-r} f^{(r,s)}(g_1(x), g_2(x)) v_{p,(r,s)}, \quad \text{(I.63)}$$

where the values $v_{p,(r,s)}$ are defined by the boundary conditions $v_{0,(0,0)} = 1$ and $v_{p,(0,0)} = 0$ for $p > 0$, and the recursion

$$v_{p,(r,s)} = \sum_{i=r+s-1}^{p-1} \binom{p-1}{i} \left[g_1^{(p-i)}(x) v_{i,(r-1,s)} + g_2^{(p-i)}(x) v_{i,(r,s-1)} \right]. \tag{I.64}$$

We have a routine taking as input the first P derivatives of g_1 at x, the first P derivatives of g_2 at x, and the partial derivatives $f^{(p,q)}$ at $(g_1(x), g_2(x))$ up to orders P and P, returning the first P derivatives of $f(g_1(x), g_2(x))$ at x.

Case $d = 2$**,** $m = 1$

Here $(f \circ g)(x) = f(g(x_1, x_2))$, where x_1 and x_2 are scalars, f is continuously differentiable up to order $P + Q$, and g is a scalar-valued function with continuous partial derivatives up to orders P and Q. The values of the derivatives of $f \circ g$ at (x_1, x_2) are computed using

$$(f \circ g)^{(p,q)}(x_1, x_2) = \sum_{r=1}^{p+q} f^{(r)}(g(x_1, x_2)) v_{(p,q),r}, \tag{I.65}$$

where the values $v_{(p,q),r}$ are defined by the conditions $v_{(0,0),0} = 1$ and $v_{(p,q),0}(x_1, x_2) = 0$ for $(p, q) \neq (0, 0)$, $v_{(p,q),r} = 0$ for $r < 0$ or $p + q < r$ and the recursion

$$v_{(p,q),r} = \begin{cases} \sum_{i=r-1}^{p-1} \binom{p-1}{i} g^{(p-i,0)}(x_1, x_2) v_{(i,0),r-1} & q = 0,\ p \geq 1 \\ \sum_{i=0}^{p} \sum_{j=0}^{q-1} \binom{p}{i} \binom{q-1}{j} g^{(p-i,q-j)}(x_1, x_2) v_{(i,j),r-1} & q \geq 1,\ p \geq 0. \end{cases} \tag{I.66}$$

We have a routine taking as input the partial derivatives $g^{(p,q)}$ at (x_1, x_2), up to orders P and Q and the first $P + Q$ derivatives of f at $g(x_1, x_2)$, returning the partial derivatives $(f \circ g)^{(p,q)}$ at (x_1, x_2), up to orders P and Q.

Annexe II

Appendix to Chapter 2

II.1 Mode of the target density

We describe in this section the computation of the unique mode $a = (a_1, \ldots, a_n)$ of the target density. The second order Taylor expansion of $\log f(\alpha|y)$ around this unique mode, a, can be expressed as :

$$\log f(\alpha|y) \approx -\frac{1}{2}\left[\alpha^\top \bar{\bar{\Omega}}\alpha - 2\bar{\bar{c}}^\top \alpha\right], \tag{II.1}$$

where $\bar{\bar{\Omega}}$ is the negative Hessian matrix of $\log f(\alpha|y)$ at a and $\bar{\bar{c}}$ equals

$$\bar{\bar{c}} = \bar{\bar{\Omega}}a + \left.\frac{\partial \log f(\alpha|y)}{\partial \alpha}\right|_{\alpha=a}. \tag{II.2}$$

Let $\psi_t^{(p,q)}$, $t = 1, \ldots, n-1$, denote the partial derivative of $\psi_t(\alpha_t, \alpha_{t+1})$ with respect to α_t and α_{t+1}, at orders p and q, respectively. Similarly, we define $\psi_0^{(q)}(\alpha_1)$ as the q'th derivative of $\psi_0(\alpha_1)$ with respect to α_1 and $\psi_n^{(p)}(\alpha_n)$ as the p'th derivative of $\psi_n(\alpha_n)$ with respect to α_n. The Hessian matrix is a symmetric tri-diagonal matrix. Its upper triangular elements are given by :

$$\begin{aligned}
\bar{\bar{\Omega}}_{t,t+1} &= -\psi_t^{(1,1)}(a_t, a_{t+1}), \quad t = 1, \ldots, n-1, \\
\bar{\bar{\Omega}}_{1,1} &= -\psi_0^{(2)}(a_1) - \psi_1^{(2,0)}(a_1, a_2), \bar{\bar{\Omega}}_{nn} = -\psi_{n-1}^{(0,2)}(a_{n-1}, a_n) - \psi_n^{(2)}(a_n), \\
\bar{\bar{\Omega}}_{t,t} &= -\psi_{t-1}^{(0,2)}(a_{t-1}, a_t) - \psi_t^{(2,0)}(a_t, a_{t+1}), \quad t = 2, \ldots, n-1.
\end{aligned} \tag{II.3}$$

The covector is given by

$$\bar{\bar{c}}_t = \begin{cases} \bar{\bar{\Omega}}_{t,t} a_t + \bar{\bar{\Omega}}_{t,t+1} a_{t+1} + \psi_0^{(1)}(a_1) + \psi_1^{(1,0)}(a_1, a_2) & t = 1 \\ \bar{\bar{\Omega}}_{t,t-1} a_{t-1} + \bar{\bar{\Omega}}_{t,t} a_t + \bar{\bar{\Omega}}_{t,t+1} a_{t+1} + \psi_{t-1}^{(0,1)}(a_{t-1}, a_t) + \psi_t^{(1,0)}(a_t, a_{t+1}) & t = 2, \ldots, n-1 \\ \bar{\bar{\Omega}}_{n,n-1} a_{n-1} + \bar{\bar{\Omega}}_{nn} a_n + \psi_{n-1}^{(0,1)}(a_{n-1}, a_n) + \psi_n^{(1)}(a_n) & t = n \end{cases} \tag{II.4}$$

The mean (and mode) of the Gaussian approximation $\mathrm{N}(\bar{\bar{\Omega}}^{-1} c, \bar{\bar{\Omega}}^{-1})$ is the mode of the target distribution and its log density has the same Hessian matrix as the log target density at this mode.

While the expressions for $\bar{\bar{\Omega}}$ and $\bar{\bar{c}}$ are more complicated than those in McCausland (2010), once we have them, we compute the mode a in the same way. Roughly speaking, we iterate the computation $\alpha' = \bar{\bar{\Omega}}(\alpha)^{-1} \bar{\bar{c}}(\alpha)$ until numerical convergence. We use two modifications to this procedure, one to accelerate convergence using higher order derivatives and the other to resort to one-at-a-time updates of the α_t in the rare cases of non-convergence.

II.2 Polynomial approximations of $a_{t|t+1}$ and $s_{t|t+1}$

Here we compute the coefficients of the Taylor series expansions of $a_{t|t+1}(\alpha_{t+1})$ and $s_{t|t+1}(\alpha_{t+1})$. These are the conditional mean and log variance of α_t given α_{t+1} according to a Gaussian approximation of the distribution of $(\alpha_1, \ldots, \alpha_t)$ given α_{t+1} and y. The point of expansion is a_{t+1} and so we compute the derivatives of $a_{t|t+1}(\alpha_{t+1})$ and $s_{t|t+1}(\alpha_{t+1})$ there.

We derive recursive expressions for these derivatives that are correct for any order r. In practice, the computational cost rises quickly and the benefits diminish quickly with r. We provide simplified expressions for $a_t^{(r)} \doteq a_{t|t+1}^{(r)}(a_{t+1})$ up to order $r = 5$ and $s_t^{(r)} \doteq s_{t|t+1}^{(r)}(a_{t+1})$ up to order $r = 4$.

We develop recursive expressions for the derivatives $a_t^{(r)}$ and $s_t^{(r)}$ for $t = 2, \ldots, n-1$. For the special case of $t = 1$, we just need to replace $\psi_{t-1}^{(0,1)}(a_{t-1|t+1}, a_{t|t+1})$ by $\psi_0^{(1)}(\alpha_1)$ in the expressions below.

The conditional mode $(a_{1|t+1}, \ldots, a_{t|t+1})$ is solution of the problem

$$(a_{1|t+1}, \ldots, a_{t|t+1}) = \arg \max_{(\alpha_1,\ldots,\alpha_t)} \log f(\alpha_1, \ldots, \alpha_t | \alpha_{t+1}, \ldots, \alpha_n, y). \tag{II.5}$$

By the conditional probability rule, we have $\log f(\alpha_1, \ldots, \alpha_t|\alpha_{t+1}, \ldots, \alpha_n, y) = \log f(\alpha|y) - \log f(\alpha_{t+1}, \ldots, \alpha_n|y)$. Using (2.3), the mode $(a_{1|t+1}, \ldots, a_{t|t+1})$ must solve

$$\psi_{t-1}^{(0,1)}(a_{t-1|t+1}, a_{t|t+1}) + \psi_t^{(1,0)}(a_{t|t+1}, \alpha_{t+1}) = 0. \tag{II.6}$$

Taking the derivatives of (II.6) with respect to α_{t+1} gives

$$\begin{aligned} 0 = &\left[\psi_{t-1}^{(0,2)}(a_{t-1|t+1}, a_{t|t+1}) + \psi_t^{(2,0)}(a_{t|t+1}, \alpha_{t+1})\right] a_{t|t+1}^{(1)} \\ &+ \psi_{t-1}^{(1,1)}(a_{t-1|t+1}, a_{t|t+1}) a_{t-1|t+1}^{(1)} + \psi_t^{(1,1)}(a_{t|t+1}, \alpha_{t+1}). \end{aligned} \tag{II.7}$$

Using the expression for the Hessian matrix in (II.3), we rewrite (II.7) as

$$\bar{\bar{\Omega}}_{t,t-1|t+1} a_{t-1|t}^{(1)}(a_{t|t+1}) + \bar{\bar{\Omega}}_{t,t|t+1} a_{t|t+1}^{(1)} = -\bar{\bar{\Omega}}_{t,t+1|t+1}. \tag{II.8}$$

?] establishes the identity $a_{t-1|t+1} = a_{t-1|t}(a_{t|t+1})$. Using the chain rule, the derivative of $a_{t-1|t+1}$ with respect to α_{t+1} is

$$a_{t-1|t+1}^{(1)} = a_{t-1|t}^{(1)}(a_{t|t+1}) a_{t|t+1}^{(1)}. \tag{II.9}$$

Substituting the right hand side (RHS) of (II.9) in (II.8), we obtain

$$\left(\bar{\bar{\Omega}}_{t,t-1|t+1} a_{t-1|t}^{(1)}(a_{t|t+1}) + \bar{\bar{\Omega}}_{t,t|t+1}\right) a_{t|t+1}^{(1)} = -\bar{\bar{\Omega}}_{t,t+1|t+1}.$$

Using similar arguments in McCausland(2010), we show by induction that for all $t = 1, \ldots, n-1$,

$$a_{t|t+1}^{(1)} = -\Sigma_{t|t+1} \bar{\bar{\Omega}}_{t,t+1|t+1}, \tag{II.10}$$

where $\Sigma_{t|t+1}^{-1} = \bar{\bar{\Omega}}_{t,t-1|t+1} a_{t-1|t}^{(1)}(a_{t|t+1}) + \bar{\bar{\Omega}}_{t,t|t+1}$. Taking $\alpha_{t+1} = a_{t+1}$ in (II.10) gives us

value of the coefficient of the monomial of degree one in the polynomial approximation of $a_{t|t+1}$ at a_{t+1} :

$$a_t^{(1)} = -\Sigma_t \bar{\bar{\Omega}}_{t,t+1}. \tag{II.11}$$

For $r \geq 2$, we use Leibniz's rule to compute $(r-1)$ derivatives of (II.8) with respect to α_{t+1} and obtain

$$\sum_{i=0}^{r-1} \binom{r-1}{i} \left(\bar{\bar{\Omega}}^{(i)}_{t,t-1|t+1} a^{(r-i)}_{t-1|t}(a_{t|t+1}) + \bar{\bar{\Omega}}^{(i)}_{t,t|t+1} a^{(r-i)}_{t|t+1} \right) = -\bar{\bar{\Omega}}^{(r-1)}_{t,t+1|t+1}. \tag{II.12}$$

Using Faà di Bruno's formula (see Appendix E of Djegnene and McCausland[2011]) for higher derivatives of a compound function, the i'th derivative of $a_{t|t+1}$ with respect to α_{t+1} is

$$a^{(i)}_{t-1|t+1} = \sum_{j=1}^{i} a^{(j)}_{t-1|t}(a_{t|t+1}) B_{i,j}\left(a^{(1)}_{t|t+1}, \ldots, a^{(i-j+1)}_{t|t+1} \right), \tag{II.13}$$

where the $B_{i,j}$ are Bell polynomials. If we replace $a^{(i)}_{t-1|t+1}$ by the RHS of (II.13) in (II.12) and set $\alpha_{t+1} = a_{t+1}$, we obtain

$$\sum_{i=0}^{r-1} \binom{r-1}{i} \left\{ \bar{\bar{\Omega}}^{(i)}_{t,t-1} \left[\sum_{j=1}^{r-i} a^{(j)}_{t-1} B_{r-i,j}\left(a^{(1)}_t, \ldots, a^{(r-i-j+1)}_t \right) \right] + \bar{\bar{\Omega}}^{(i)}_{t,t} a^{(r-i)}_t \right\} = -\bar{\bar{\Omega}}^{(r-1)}_{t,t+1}. \tag{II.14}$$

This gives an expression for $a_t^{(r)}$ in terms of $a_t^{(i)}$, $i = 0, \ldots, r-1$, and $a_{t-1}^{(i)}$, $i = 0, \ldots, r$, as well as $\bar{\bar{\Omega}}^{(i)}_{t,t-1}$, $\bar{\bar{\Omega}}^{(i)}_{t,t}$, $i = 1, \ldots, r-1$, and $\bar{\bar{\Omega}}^{(r-1)}_{t,t+1}$.

We now derive a result that will give us $s_t^{(r)}$ in terms of $a_t^{(i)}$ up to order $i = r+1$ and $s_t^{(i)}$ up to order $i = r-1$. Using $\Sigma_{t|t+1} = \exp(s_{t|t+1})$ and applying Faà di Bruno formula gives

$$\begin{aligned}
\Sigma^{(r)}_{t|t+1} =& \sum_{i=1}^{r} \Sigma_{t|t+1} B_{r,i}(s^{(1)}_{t|t+1}, \ldots, s^{(r-i+1)}_{t|t+1}) \\
=& \Sigma_{t|t+1} \sum_{i=1}^{r} B_{r,i}(s^{(1)}_{t|t+1}, \ldots, s^{(r-i+1)}_{t|t+1}) \\
=& \Sigma_{t|t+1} B_r(s^{(1)}_{t|t+1}, \ldots, s^{(r)}_{t|t+1}),
\end{aligned}$$

where B_r, known as the r'th order complete Bell polynomial, is defined as $B_r \equiv \sum_{i=1}^{r} B_{r,i}$.

Using Leibniz's rule to compute $(r-1)$ derivatives of (II.10) with respect to α_{t+1} at $\alpha_{t+1} = a_{t+1}$, we obtain

$$a_t^{(r)} = -\sum_{i=0}^{r-1} \binom{r-1}{i} B_i(s_t^{(1)}, \ldots, s_t^{(i)}) \Sigma_t \bar{\bar{\Omega}}_{t,t+1}^{(r-1-i)}. \tag{II.15}$$

This gives us value of $s_t^{(r-1)}$ in term of $s_t^{(i)}$, $i = 0, \ldots, r-2$ and $a_t^{(i)}$, $i = 0, \ldots, r$.

The equations in (II.16) give expressions for $a_t^{(r)}$, $r = 1, \ldots, 5$ and $s_t^{(r)}$, $r = 1, \ldots, 4$. The development to obtain these explicit expressions is similar to appendix A.2 of Djegnene and McCausland(2011).

Let $\gamma_t = -\Sigma_t \bar{\bar{\Omega}}_{t,t-1}$, $\bar{\bar{\mathbf{\Omega}}}_t^{(i)} = \bar{\bar{\Omega}}_{t,t-1}^{(i)} a_t^{(1)} + \bar{\bar{\Omega}}_{t,t}^{(i)}$ and,

$$
\begin{aligned}
Z_{21} =& \gamma_t a_t^{(1)} a_{t-1}^{(2)} - \Sigma_t \bar{\bar{\mathbf{\Omega}}}_t^{(1)} & Z_{32} =& \gamma_t a_t^{(1)} a_{t-1}^{(3)} - 2\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_{t-1}^{(2)} \\
Z_{31} =& \gamma_t a_t^{(2)} a_{t-1}^{(2)} - \Sigma_t \bar{\bar{\mathbf{\Omega}}}_t^{(2)} & Z_{43} =& \gamma_t a_t^{(1)} a_{t-1}^{(4)} - 3\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_{t-1}^{(3)} \\
Z_{42} =& \gamma_t a_t^{(2)} a_{t-1}^{(3)} - \Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(2)} a_{t-1}^{(2)} & Z_{41} =& \gamma_t a_t^{(3)} a_{t-1}^{(2)} - 3\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_t^{(2)} a_{t-1}^{(2)} - \Sigma_t \bar{\bar{\mathbf{\Omega}}}_t^{(3)} \\
Z_{54} =& \gamma_t a_{t-1}^{(5)} a_t^{(1)} - 4\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_{t-1}^{(4)} & Z_{53} =& \gamma_t a_{t-1}^{(4)} a_t^{(2)} - \Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(2)} a_{t-1}^{(3)} \\
Z_{52} =& \gamma_t a_{t-1}^{(3)} a_t^{(3)} - \Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(3)} a_{t-1}^{(2)} \dots & Z_{51} =& \gamma_t a_{t-1}^{(2)} a_t^{(4)} + 3\gamma_t a_{t-1}^{(3)} \left(a_t^{(2)}\right)^2 - \Sigma_t \bar{\bar{\mathbf{\Omega}}}_t^{(4)} \dots \\
& - 2\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_{t-1}^{(3)} a_t^{(2)} & & - 6\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(2)} a_{t-1}^{(2)} a_t^{(2)} - 4\Sigma_t \bar{\bar{\Omega}}_{t,t-1}^{(1)} a_{t-1}^{(2)} a_t^{(3)}.
\end{aligned}
$$

Then, we have

$$
\begin{aligned}
a_t^{(2)} =& Z_{21} a_t^{(1)} - \Sigma_t \bar{\bar{\Omega}}_{t,t+1}^{(1)} \\
s_t^{(1)} =& Z_{21} \\
a_t^{(3)} =& Z_{32} \left(a_t^{(1)}\right)^2 + Z_{31} a_t^{(1)} - \Sigma_t \bar{\bar{\Omega}}_{t,t+1}^{(2)} + 2 s_t^{(1)} a_t^{(2)} \\
s_t^{(2)} =& Z_{32} a_t^{(1)} + Z_{31} + \left(s_t^{(1)}\right)^2 \\
a_t^{(4)} =& Z_{43} \left(a_t^{(1)}\right)^3 + 3 Z_{42} \left(a_t^{(1)}\right)^2 + Z_{41} a_t^{(1)} - \Sigma_t \bar{\bar{\Omega}}_{t,t+1}^{(3)} \\
& + 3 \left(s_t^{(2)} - \left(s_t^{(1)}\right)^2\right) a_t^{(2)} + 3 s_t^{(1)} a_t^{(3)} \\
s_t^{(3)} =& Z_{43} \left(a_t^{(1)}\right)^2 + 3 Z_{42} a_t^{(1)} + Z_{41} - \left(s_t^{(1)}\right)^3 + 3 s_t^{(1)} s_t^{(2)} \\
a_t^{(5)} =& Z_{54} \left(a_t^{(1)}\right)^4 + 6 Z_{53} \left(a_t^{(1)}\right)^3 + 4 Z_{52} \left(a_t^{(1)}\right)^2 + Z_{51} a_t^{(1)} - \Sigma_t \bar{\bar{\Omega}}_{t,t+1}^{(4)} \\
& + 6 s_t^{(1)} a_t^{(4)} + 6 \left(s_t^{(2)} - \left(s_t^{(1)}\right)^2\right) a_t^{(3)} + 5 \left(s_t^{(3)} + \left(s_t^{(1)}\right)^3 - 3 s_t^{(1)} s_t^{(2)}\right) a_t^{(2)} \\
s_t^{(4)} =& Z_{54} \left(a_t^{(1)}\right)^3 + Z_{53} \left(a_t^{(1)}\right)^2 + Z_{52} a_t^{(1)} + Z_{51} - \Sigma_t \bar{\bar{\Omega}}_{t,t+1}^{(4)} \\
& + \left(s_t^{(1)}\right)^4 + 4 s_t^{(1)} s_t^{(3)} + 3 \left(s_t^{(2)} - 2 \left(s_t^{(1)}\right)^2\right) s_t^{(2)}.
\end{aligned}
\tag{II.16}
$$

II.3 Conditional mode and mean.

II.3.1 Gradient of the log conditional density

In this subsection, we derive an exact expression of the gradient of the log conditional density $\log f(\alpha_t|\alpha_{t+1}, y)$. For $t = 1$, Bayes' rule gives

$$\begin{aligned}\frac{\partial \log f(\alpha_1|\alpha_2, y)}{\partial \alpha_1} =& \frac{\partial \log f(\alpha_2, y_1|\alpha_1)}{\partial \alpha_1} + \frac{\partial \log f(\alpha_1)}{\partial \alpha_1} \\ &= \psi_1^{(1,0)}(\alpha_1, \alpha_2) + \psi_0^{(1)}(\alpha_1).\end{aligned}$$

For $t = 2, \ldots, n-1$, we first write

$$f(\alpha_t|\alpha_{t+1}, y) = f(y_{t+1:n}|\alpha_{t+1}) f(\alpha_{t+1}, y_t|\alpha_t) f(\alpha_t|y_{1:t-1}) f(y_{1:t-1}).$$

Taking the logarithm of $f(\alpha_t|\alpha_{t+1}, y)$ and then its derivative with respect to α_t gives

$$\frac{\partial \log f(\alpha_t|\alpha_{t+1}, y)}{\partial \alpha_t} = \frac{\partial \log f(\alpha_{t+1}, y_t|\alpha_t)}{\partial \alpha_t} + \frac{\partial \log f(\alpha_t|y_{1:t-1})}{\partial \alpha_t}. \tag{II.17}$$

Now consider the derivative of $\log f(\alpha_t|y_{1:t-1})$ with respect to α_t. We have

$$\begin{aligned}f(\alpha_t|y_{1:t-1}) &= \int f(\alpha_t, \alpha_{t-1}|y_{1:t-1}) d\alpha_{t-1} \\ &= \int f(\alpha_t, y_{t-1}|\alpha_{t-1}) \frac{f(\alpha_{t-1}, y_{1:t-2})}{f(y_{1:t-1})} d\alpha_t.\end{aligned}$$

Then, taking the derivative of $f(\alpha_t|y_{1:t-1})$ with respect to α_t gives

$$\begin{aligned}\frac{\partial f(\alpha_t|y_{1:t-1})}{\partial \alpha_t} &= \int \frac{\partial f(\alpha_t, y_{t-1}|a_{t-1})}{\partial \alpha_t} \frac{f(\alpha_{t-1}, y_{1:t-2})}{f(y_{1:t-1})} d\alpha_{t-1} \\ &= \int \frac{\partial \log f(\alpha_t, y_{t-1}|\alpha_{t-1})}{\partial \alpha_t} \frac{f(\alpha_t, \alpha_{t-1}, y_{1:t-1})}{f(y_{1:t-1})} d\alpha_{t-1} \\ &= \int \frac{\partial \log f(\alpha_t, y_{t-1}|\alpha_{t-1})}{\partial \alpha_t} f(\alpha_{t-1}|\alpha_t, y_{1:t-1}) f(\alpha_t|y_{1:t-1}) d\alpha_{t-1}.\end{aligned}$$

Dividing both sides by $f(\alpha_t|y_{1:t-1})$ leads to

$$\frac{1}{f(\alpha_t|y_{1:t-1})}\frac{\partial f(\alpha_t|y_{1:t-1})}{\partial \alpha_t} = \int \frac{\partial f(\alpha_t, y_{t-1}|\alpha_{t-1})}{\partial \alpha_t} f(\alpha_{t-1}|\alpha_t, y_{1:t-1}) d\alpha_{t-1},$$

or

$$\begin{aligned}\frac{\partial \log f(\alpha_t|y_{1:t-1})}{\partial \alpha_t} &= \int \frac{\partial \log f(\alpha_t, y_{t-1}|a_{t-1})}{\partial \alpha_t} f(\alpha_{t-1}|\alpha_t, y_{1:t-1}) d\alpha_{t-1} \\ &= E\left[\frac{\partial \log f(\alpha_t, y_{t-1}|a_{t-1})}{\partial \alpha_t}|\alpha_t, y_{1:t-1}\right] \\ &= E\left[\psi_{t-1}^{0,1}(\alpha_{t-1}, \alpha_t)|\alpha_t, y_{1:t-1}\right]\end{aligned} \tag{II.18}$$

Using (II.18) in (II.17), we obtain finally

$$h_t^{(1)}(\alpha_t; \alpha_{t+1}) = \frac{\partial \log f(\alpha_t|\alpha_{t+1}, y)}{\partial \alpha_t} = \psi_t^{(1,0)}(\alpha_t, \alpha_{t+1}) + x_{t-1|t}(\alpha_t), \tag{II.19}$$

where $x_{t-1|t}(\alpha_t) = E\left[\psi_{t-1}^{(0,1)}(\alpha_{t-1}, \alpha_t)|\alpha_t, y_{1:t-1}\right]$.

For the case $t = n$, we have $f(\alpha_n|y) \propto f(y_n|\alpha_n) f(\alpha_n|y_{1:n-1})$, so that

$$\frac{\partial \log f(\alpha_n|y)}{\partial \alpha_n} = \psi_n^1(\alpha_n) + x_{n-1|n}(\alpha_n) \tag{II.20}$$

II.3.2 Approximation of the conditional derivatives

We cannot easily compute the derivative $h_t^{(1)}(\alpha_t; \alpha_{t+1})$ due to the conditional expectation $x_{t-1|t}(\alpha_t)$. Thus, we propose an approximation $H_t^{(1)}(\alpha_t; \alpha_{t+1})$, with

$$H_t^{(1)}(\alpha_t; \alpha_{t+1}) = \Psi_t^{(1,0)}(\alpha_t, \alpha_{t+1}) + X_{t-1|t}(\alpha_t), \tag{II.21}$$

where $X_{t-1|t}(\alpha_t)$ is an approximation of $x_{t-1|t}(\alpha_t)$ and $\Psi_t^{(p,q)}(\alpha_t, \alpha_{t+1})$ is an approximation of $\psi_t^{(p,q)}(\alpha_t, \alpha_{t+1})$.

Equation (2.9) constructs $\Psi_t^{(p,q)}(\alpha_t, \alpha_{t+1})$ as a multivariate Taylor expansion of $\psi_t^{(p,q)}(\alpha_t, \alpha_{t+1})$ around a_t and a_{t+1}. Appendix C.2 in Djegnene and McCausland[2011] gives details on the construction of the approximation $X_{t-1|t}(\alpha_t)$. A minor adjustment is required to

compute the values $M_{t-1}^{(j)}, j = 0, \ldots, r$ included in the coefficients $X_{t-1}^{(r)}, r = 0, \ldots, 4$ of the polynomial approximation $X_{t-1|t}(\alpha_t)$. We just need to replace equation (67) in Appendix C.4 of Djegnene and McCausland (2011), with $h_t^{(1)}(b_{t|t+1}; \alpha_{t+1}) b_{t|t+1}^{(1)} = -\bar{\bar{\Omega}}_{t,t+1}(b_{t|t+1}, \alpha_{t+1})$, and define $D_1(\alpha_{t+1}) \doteq \bar{\bar{\Omega}}_{t,t+1}(b_{t|t+1}, \alpha_{t+1})$.

II.3.3 Coefficients of the polynomial approximation of the conditional mode $b_{t|t+1}$

The mode $b_{t|t+1}$ of $\log f(\alpha_t|\alpha_{t+1}, y_t)$ is the root of $h_t(\alpha_t; \alpha_{t+1})$. This root is well approximated using one iteration of the Newton-Raphson algorithm for root finding, starting at $a_{t|t+1}$. Thus, we have

$$b_{t|t+1} \approx a_{t|t+1} - \frac{h_t^{(1)}(a_{t|t+1}; \alpha_{t+1})}{h_t^{(2)}(a_{t|t+1}; \alpha_{t+1})}, \tag{II.22}$$

where $h_t^{(1)}(\alpha_t; \alpha_{t+1})$ is the first order derivative of $h_t(\alpha_t; \alpha_{t+1})$ with respect to α_t.

Using equation (II.22), we define the approximation $B_t^{(r)}$ of $b_t^{(r)}$, $r = 0, \ldots, R-1$, as

$$B_t^{(r)} \doteq a_t^{(r)} - \frac{\partial^r}{\partial \alpha_{t+1}^r} \left(\frac{H_t^{(1)}(a_{t|t+1}(\alpha_{t+1}); \alpha_{t+1})}{H_t^{(2)}(a_{t|t+1}(\alpha_{t+1}); \alpha_{t+1})} \right) \Bigg|_{\alpha_{t+1}=a_{t+1}}. \tag{II.23}$$

The second term on the RHS of (II.23) is the r'th order derivative of a quotient, which we compute using the quotient rule for derivatives, equation (81) in Appendix E. of Djegnene and McCausland[2011].

Subtracting the first order condition (II.6) for $(a_{1|t+1}, \ldots, a_{t|t+1})$ to be the conditional mode of $f(\alpha_1, \ldots, \alpha_t|\alpha_{t+1}, y)$ from equation (II.21), evaluated at $a_{t|t+1}$, gives

$$\begin{aligned} H_t^{(1)}(a_{t|t+1}; \alpha_{t+1}) = &- \psi_{t-1}^{(0,1)}(a_{t-1|t}(a_{t|t+1}), a_{t|t+1}) + X_{t-1|t}(a_{t|t+1}) \\ &+ \left(\Psi_t^{(1,0)}(a_{t|t+1}, \alpha_{t+1}) - \psi_t^{(1,0)}(a_{t|t+1}, \alpha_{t+1}) \right). \end{aligned}$$

It is easy to show that

$$\frac{\partial^{(r)} \Psi_t^{(1,0)}(a_{t|t+1}, \alpha_{t+1})}{\partial \alpha_{t+1}^r} \Bigg|_{\alpha_{t+1}=a_{t+1}} = \frac{\partial^{(r)} \psi_t^{(1,0)}(a_{t|t+1}, \alpha_{t+1})}{\partial \alpha_{t+1}^r} \Bigg|_{\alpha_{t+1}=a_{t+1}}.$$

Consequently,

$$\left.\frac{\partial^{(r)} H_t^{(1)}(a_{t|t+1};\alpha_{t+1})}{\partial \alpha_{t+1}^r}\right|_{\alpha_{t+1}=a_{t+1}} = \left.\frac{\partial^{(r)}\left[-\psi_{t-1}^{(0,1)}(a_{t-1|t}(a_{t|t+1}),a_{t|t+1})+X_{t-1|t}(a_{t|t+1})\right]}{\partial \alpha_{t+1}^r}\right|_{\alpha_{t+1}=a_{t+1}}.$$

In the case of $t = n$, we have a value b_n and not a function to approximate. Using a development analogous to the case $t = 2, \ldots, n-1$, we define the following approximation of b_n :

$$B_n \doteq a_n - \frac{H_n(a_n)}{H_n^{(1)}(a_n)}. \tag{II.24}$$

II.4 ASCD models derivatives

In this section we provide analytic expression of $\psi_t(\alpha_t, \alpha_{t+1})$, the log conditional density of (α_{t+1}, y_t) given α_t, along with its partial derivatives with respect to α_t and α_{t+1}. We compute $\log f(y_t|\alpha_t)$ and $\log f(\alpha_{t+1}|y_t, \alpha_t)$ and then combine these two densities to obtain $\psi_t(\alpha_t, \alpha_{t+1})$.

We described in Section 2.3.1 the asymmetric stochastic conditional duration models (ASCD) used as empirical illustrations. The observation innovation, ε_t, is either Exponential, Gamma or Weibull. The three distributions are scale-normalized. The Gamma and Weibull distributions have shape parameter ν. The density function of each of these distributions can be written as the generic function

$$f(\varepsilon_t; \lambda, \nu, \delta) = \lambda \varepsilon_t^{\nu-1} \exp(-\varepsilon_t^{\delta}),$$

where the choice of the parameters λ, ν and δ determines the distribution. For the Exponential distribution, we have (λ, ν, δ)=(1,1,1). For Gamma and Weibull distribution, λ and δ is function of ν, which is the only free parameter. We have $(\lambda, \delta) = (\nu, \nu)$ for the Weibull distribution and $(\lambda, \delta) = (\Gamma(\nu)^{-1}, 1)$ for the Gamma distribution.

Recall from Equation (2.11), the observation equation $y_t = \exp(\alpha_t)\varepsilon_t$. Using a

change of measure, the log conditional density of the observation variable is

$$f(y_t|\alpha_t) = \lambda y_t^{\nu-1} \exp(-\nu\alpha_t) \exp\left(-y_t^{\delta} \exp(-\delta\alpha_t)\right).$$

The log density equals :

$$\log f(y_t|\alpha_t) = \log\lambda + (\nu-1)\log y_t - \nu\alpha_t - \varphi_t(\alpha_t), \tag{II.25}$$

with $\varphi_t(\alpha_t) = y_t^{\delta}\exp(-\delta\alpha_t)$.

The state equation in (2.11) implies that the conditional distribution of α_{t+1}, given α_t and y_t, is Gaussian with mean $(1-\phi)\bar{\alpha}+(\phi-\rho)\alpha_t+\rho\log(y_t)$ and variance σ^2. Thus, the log conditional density of α_{t+1} equals

$$\log f(\alpha_{t+1}|y_t,\alpha_t) = -0.5\log(2\pi\sigma^2) - 0.5\sigma^{-2}u_t^2(\alpha_t,\alpha_{t+1}), \tag{II.26}$$

with $u_t(\alpha_t,\alpha_{t+1}) = \alpha_{t+1} - (1-\phi)\bar{\alpha} - (\phi-\rho)\alpha_t - \rho\log(y_t)$.

To compute $\psi_t(\alpha_t,\alpha_{t+1})$, the log conditional density of (α_{t+1},y_t) given α_t, we use the probability decomposition rule, $f(\alpha_{t+1},y_t|\alpha_t) = f(y_t|\alpha_t)f(\alpha_{t+1}|y_t,\alpha_t)$, apply the logarithm to each side of this decomposition, and then add the RHS of (II.25) and (II.26). We obtain

$$\psi_t(\alpha_t,\alpha_{t+1}) = \log(\lambda)+(\nu-1)\log(y_t)-\nu\alpha_t-\varphi_t(\alpha_t)-0.5\log(2\pi\sigma^2)-0.5\sigma^{-2}u_t^2(\alpha_t,\alpha_{t+1}).$$

The log-density $\psi_t(\alpha_t,\alpha_{t+1})$ is the sum of an exponential function and a quadratic function. Its derivatives with respect to α_t and α_{t+1} at order respectively p and q are

easily computed and are given by :

$$\psi_t^{p,q}(\alpha_t, \alpha_{t+1}) = \begin{cases} -\frac{1}{\sigma^2} u_t & p = 0,\ q = 1, \\ -\frac{1}{\sigma^2} & p = 0,\ q = 2, \\ -\nu + \delta\varphi_t(\alpha_t) + \frac{\phi - \rho}{\sigma^2} u_t(\alpha_t, \alpha_{t+1}) & p = 1,\ q = 0, \\ \frac{\phi - \rho}{\sigma^2} & p = 1,\ q = 1, \\ -(-\delta)^p \varphi_t(\alpha_t) - \frac{(\phi - \rho)^2}{\sigma^2} 1_{\{p=2\}} & p \geq 2,\ q = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{II.27}$$

For $t = 1$, we have

$$\psi_0(\alpha_1) = -0.5 \log\left(2\pi \frac{\sigma^2}{1 - \phi^2}\right) - 0.5 \frac{1 - \phi^2}{\sigma^2} (\alpha_1 - \bar{\alpha})^2,$$

$$\psi_0^1(\alpha_1) = -\frac{1 - \phi^2}{\sigma^2}(\alpha_1 - \bar{\alpha}), \quad \psi_0^2(\alpha_1) = -\frac{1 - \phi^2}{\sigma^2}, \quad \psi_0^p(\alpha_1) = 0,\ p \geq 3.$$

For $t = n$, we have

$$\psi_n(\alpha_n) = \log(\lambda) + (\nu - 1)\log(y_n) - \nu\alpha_n - \varphi_n(\alpha_n),$$

and

$$\psi_n^1(\alpha_n) = -\nu + \delta\varphi_n(\alpha_n), \quad \psi_n^p(\alpha_n) = -(-\delta)^p \varphi_n(\alpha_n),\ p \geq 2.$$

Annexe III

Appendix to Chapter 3

III.1 Reputation Appendix

The FOCs' proposal e^t supports a pooling equilibrium if

$$\begin{aligned}(1) \quad \Pi^t(e^t,\omega;\theta) \geq \quad \tilde{\Pi}^t &= \max_e \Pi^t(e,\omega;\beta=0)\\ (2) \quad \Pi^u(e^t,\omega;\theta) \geq \quad \hat{\Pi}^u &= \Pi^u(\hat{e}^u,\omega;\beta=0)\end{aligned}$$

Let $e^t = e^{\star t} = \arg\max_e \Pi^t(e,\omega;\theta)$. Given that $\Pi^t(e,\omega;\theta) \geq \Pi^t(e,\omega;0)$ for all $e \in [0,1]$, then $e^{\star t}$ satisfies the first constraint. For the second constraint, we have :

$$\Pi^u(e^{\star t},\omega;\theta) = \frac{1-\omega}{6b(1-\theta)}\left[\frac{(1+\omega+\omega^2)-6ba_u(1-\pi^t)^2}{\pi^t}\right].$$

Using the fact that :

$$(\pi^u)^2 = 1 - \frac{1+\omega+\omega^2}{6ba_u}$$

we have :

$$\begin{aligned}\Pi^u(e^{\star t},\omega;\theta) =& \frac{(1-\omega)a_u}{(1-\theta)}\left[\frac{1-(\pi^u)^2-(1-\pi^t)^2}{\pi^t}\right]\\ =& \frac{(1-\omega)a_u}{(1-\theta)}\left[\frac{1-(\pi^u)^2-(1-\pi^t)^2}{\pi^t}\right].\end{aligned} \tag{III.1}$$

Recall that from (3.9) :

$$\hat{\Pi}^u = \Pi(\hat{e}^u,\omega;\beta=0) = 2(1-\omega)a_u(1-\pi^u).$$

So, we have :

$$\begin{aligned}\Pi^u(e^{\star t},\omega;\theta) - \hat{\Pi}^u =& \frac{(1-\omega)a_u}{(1-\theta)\pi^t}\left[1-(\pi^u)^2-(1-\pi^t)^2-2(1-\theta)\pi^t(1-\pi^u)\right]\\ =& \frac{(1-\omega)a_u}{(1-\theta)\pi^t}\left[2\theta\pi^t(1-\pi^u)-(\pi^u-\pi^t)^2\right]\end{aligned}$$

Let $\Delta(\omega;\theta,a_u,a_t,b) = 2\theta\pi^t(1-\pi^u) - (\pi^u-\pi^t)^2$. We have

$$\Pi^u(e^{\star t},\omega;\theta) - \hat{\Pi}^u \geq 0 \Leftrightarrow \Delta(\omega;\theta,a_u,a_t,b) \geq 0$$

As a function of ω, $\Delta(\omega;\theta,a_u,a_t,b)$ is continuous on $[0,1]$. So it has a minimal value on $[0,1]$. The function $\Delta(\omega;\theta,a_u,a_t,b)$ is also differentiable on $(0,1)$. Let us assume that the minimum, $\omega^\star(\theta,a_u,a_t,b) \in (0,1)$ and let $\Delta^\star(\theta,a_u,a_t,b)$ denote the minimal value

$$\Delta^\star(\theta,a_u,a_t,b) = \Delta(\omega^\star(\theta,a_u,a_t,b);\theta,a_u,a_t,b).$$

Then, by the envelope theorem, we have :

$$\frac{\partial\Delta^\star(\theta,a_u,a_t,b)}{\partial\theta} = 2\pi^{\star t}(1-\pi^{\star u}) \geq 0$$

The minimal value is increasing in θ. If the minimum value is attained on the boundaries, $\{0,1\}$, $\Delta^\star(\theta,a_u,a_t,b)$ is still an increasing function of θ. However, we have for $(a_u,a_t,b) \in \Lambda$, $\Delta^\star(1,a_u,a_t,b) > 0$ and $\Delta^\star(0,a_u,a_t,b) < 0$. So, there exists $\underline{\theta} \in (0,1)$ such that $\Delta^\star(\underline{\theta},a_u,a_t,b) = 0$ and for all $\theta \geq \underline{\theta}$ we have :

$$\Delta(\theta,a_u,a_t,b) \geq \Delta^\star(\theta,a_u,a_t,b) \geq \Delta^\star(\underline{\theta},a_u,a_t,b) = 0.$$

BIBLIOGRAPHIE FOR CHAPITRE 1

[1] C. K. Carter and R. Kohn. On gibbs sampling for state space models. *Biometrika*, 81(3) :541–553, 1994.

[2] J. C. C. Chan and I. Jeliazkov. Efficient simulation and integrated likelihood estimation in state space models. 2009. Working paper.

[3] S. Chib, F. Nardari, and N. Shephard. Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics*, 108 :281–316, 2002.

[4] P. de Jong and N. Shephard. The simulation smoother for time series models. *Biometrika*, 82(1) :339–350, 1995.

[5] J. Durbin and S. J. Koopman. A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89(3) :603–615, 2002.

[6] D. Feng, J. J. Jiang, and P. Song. Stochastic conditional durations models with "Leverage Effect" for financial transaction data. *Journal of Financial Econometrics*, 2 :390–421, 2004.

[7] S. Frühwirth-Schnatter. Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15 :183–202, 1994.

[8] S. Frühwirth-Schnatter and H. Wagner. Auxiliary mixture sampling for parameter-driven models of time series of counts with applications to state space modelling. *Biometrika*, 93 :827–841, 2006.

[9] S. Frühwirth-Schnatter, R. Früwirth, L. Held, and H. Rue. Improved auxiliary mixture sampling for hierarchical models of non-Gaussian data. *Statistics and Computing*, 19 :479–492, 2009.

[10] J. Geweke. Bayesian inference in econometric models using monte carlo integration. *Econometrica*, 57 :1317–1340, 1989.

[11] J. Geweke. Getting it right : Joint distribution tests of posterior simulators. *Journal of the American Statistical Association*, 99 :799–804, 2004.

[12] A. C. Harvey and N. Shephard. The estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business and Economic Statistics*, 14 : 429–434, 1996.

[13] E. Jacquier, N. Polson, and P. Rossi. Bayesian analysis of stochastic volatility models with leverage effect and fat tails. *Journal of Econometrics*, pages 185–212, 2004.

[14] B. Jungbacker and S. J. Koopman. Monte Carlo estimation for nonlinear non-Gaussian state space models. *Biometrika*, 94 :827–839, 2008.

[15] S. Kim, N. Shephard, and S. Chib. Stochastic volatility : Likelihood inference and comparison with ARCH models. *Review of Economic Studies*, 65(3) :361–393, 1998.

[16] W. J. McCausland. The hessian method : Highly efficient simulation smoothing, in a nutshell. *Journal of Econometrics*, 168(2) :189 – 206, 2012. ISSN 0304-4076.

[17] W. J. McCausland, S. Miller, and D. Pelletier. Simulation smoothing for state-space models : A computational efficiency analysis. *Computational Statistics and Data Analysis*, 55 :199–212, 2011.

[18] J. Nakajima and Y. Omori. Leverage, heavy-tails and correlated jumps in stochastic volatility models. *Computational Statistics and Data Analysis*, 53.

[19] Y. Omori and T. Watanabe. Block sampler and posterior mode estimation for asymmetric stochastic volatility models. *Computational Statistics and Data Analysis*, 52 :2892–2910, 2008.

[20] Y. Omori, S. Chib, N. Shephard, and J. Nakajima. Stochastic volatility with leverage : fast and efficient likelihood inference. *Journal of Econometrics*, 140 : 425–449, 2007.

[21] J. Richard and W. Zhang. Efficient high-dimensional importance sampling. *Journal of Econometrics*, 141 :1385–1411, 2007.

[22] H. Rue. Fast sampling of Gaussian Markov random fields with applications. *Journal of the Royal Statistical Society Series B*, 63 :325–338, 2001.

[23] H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations. *Journal of the Royal Statistical Society Series B*, 71 :319âĂŞ–392, 2009.

[24] N. Shephard and M. K. Pitt. Likelihood analysis of non-Gaussian measurement time series. *Biometrika*, 84(3) :653–667, 1997.

[25] C. M. Strickland, C. S. Forbes, and G. M. Martin. Bayesian analysis of the stochastic conditional duration model. *Computational Statistics and Data Analysis*, 50 :2247–2267, 2006.

[26] J. R. Stroud, P. Müller, and N. G. Polson. Nonlinear state-space models with state-dependent variances. *Journal of the American Statistical Association*, 98 :377–386, 2003.

[27] T. Watanabe and Y. Omori. A multi-move sampler for estimating non-Gaussian time series models : Comments on Shephard and Pitt (1997). *Biometrika*, 91 : 246–248, 2004.

[28] J. Yu. On leverage in a stochastic volatility model. *Journal of Econometrics*, 127 : 165–178, 2005.

BIBLIOGRAPHIE FOR CHAPITRE 2

[1] C. K. Carter and R. Kohn. On gibbs sampling for state space models. *Biometrika*, 81(3) :541–553, 1994.

[2] J. C. C. Chan and I. Jeliazkov. Efficient simulation and integrated likelihood estimation in state space models. 2009. Working paper.

[3] P. de Jong and N. Shephard. The simulation smoother for time series models. *Biometrika*, 82(1) :339–350, 1995.

[4] B. Djegnene and W. J. McCausland. The hessian method for models with leverage like effect. Thesis chapter, Economics Department, University of Montreal, 2012.

[5] J. Durbin and S. J. Koopman. A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89(3) :603–615, 2002.

[6] D. Feng, J. J. Jiang, and P. Song. Stochastic conditional durations models with "Leverage Effect" for financial transaction data. *Journal of Financial Econometrics*, 2 :390–421, 2004.

[7] S. Frühwirth-Schnatter. Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15 :183–202, 1994.

[8] B. Jungbacker and S. J. Koopman. Monte Carlo estimation for nonlinear non-Gaussian state space models. *Biometrika*, 94 :827–839, 2008.

[9] S. Kim, N. Shephard, and S. Chib. Stochastic volatility : Likelihood inference and comparison with ARCH models. *Review of Economic Studies*, 65(3) :361–393, 1998.

[10] W. J. McCausland. The hessian method : Highly efficient simulation smoothing, in a nutshell. *Journal of Econometrics*, 168(2) :189 – 206, 2012. ISSN 0304-4076.

[11] W. J. McCausland, S. Miller, and D. Pelletier. Simulation smoothing for state-space models : A computational efficiency analysis. *Computational Statistics and Data Analysis*, 55 :199–212, 2011.

[12] Y. Omori and T. Watanabe. Block sampler and posterior mode estimation for asymmetric stochastic volatility models. *Computational Statistics and Data Analysis*, 52 :2892–2910, 2008.

[13] Y. Omori, S. Chib, N. Shephard, and J. Nakajima. Stochastic volatility with leverage : fast and efficient likelihood inference. *Journal of Econometrics*, 140 : 425–449, 2007.

[14] J. Richard and W. Zhang. Efficient high-dimensional importance sampling. *Journal of Econometrics*, 141 :1385–1411, 2007.

[15] H. Rue. Fast sampling of Gaussian Markov random fields with applications. *Journal of the Royal Statistical Society Series B*, 63 :325–338, 2001.

[16] H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations. *Journal of the Royal Statistical Society Series B*, 71 :319âĂŞ–392, 2009.

[17] N. Shephard and M. K. Pitt. Likelihood analysis of non-Gaussian measurement time series. *Biometrika*, 84(3) :653–667, 1997.

[18] C. M. Strickland, C. S. Forbes, and G. M. Martin. Bayesian analysis of the stochastic conditional duration model. *Computational Statistics and Data Analysis*, 50 :2247–2267, 2006.

[19] R. Tsay. *Analysis of Financial Time Series*. Wiley, 2010.

[20] T. Watanabe and Y. Omori. A multi-move sampler for estimating non-Gaussian time series models : Comments on Shephard and Pitt (1997). *Biometrika*, 91 : 246–248, 2004.

[21] D. Xu, J. Knight, and T. S. Wirjanto. Asymmetric stochastic conditional duration modelŮa mixture-of-normal approach. *Journal of Financial Econometrics*, 9(3) :469–488, 2011. doi : 10.1093/jjfinec/nbq026. URL `http://jfec.oxfordjournals.org/content/9/3/469.abstract`.

BIBLIOGRAPHIE FOR CHAPITRE 3

[1] S. Albæk and C. Schultz. On the relative advantage of cooperatives. *Economic Letter*, 59(3) :397 – 401, 1998.

[2] K. J. Arrow. Gifts and exchanges. *Philosophy and Public Affairs*, 1(4) :343–362, 1972.

[3] A. Banerjee, D. Mookherjee, k. Munshi, and D. Ray. Inequality, control rights, and rent seeking : Sugar cooperatives in maharashtra. *Journal of Political Economy*, 109(1) :138–190, 2001.

[4] J. Berg, J. Dickhaut, and K. McCabe. Trust, reciprocity, and social history. *Games and Economic Behavior*, 10(1) :122 – 142, 1995.

[5] D. Gambetta. Can we trust trust ? In *Trust : Making and Breaking Cooperative Relations*, pages 213–237. Basil Blackwell, 1988.

[6] R. Hardin. Conceptions and explanations of trust. In *Trust in Society*, pages 3–39. Russell Sage Foundation Publications, 2003.

[7] H. S. James and M. E. Sykuta. Farmer trust in producer- and investor-owned firms : Evidence from missouri corn and soybean producers. *CORI Working Paper*, No. 2004-14., 2004.

[8] H. S. James and M. E. Sykuta. Property right and organizational characteristics of producer-owned firms and organizational trust. *Annals of Public & Cooperative Economics*, 76(4), 2005.

[9] K. Karantininis and A. Zago. Cooperatives and membership commitment : Endogenous membership in mixed duopsonies. *American Journal of Agricultural Economics*, 83(5) :1266–1272, 2001.

[10] C. LeVay. Agricultural cooperative theory : a review. *Journal of Agricultural Economics*, 34(1), 1983.

[11] R. J. Sexton. Imperfect competition in agricultural markets and the role of cooperatives : A spatial analysis. *American Journal of Agricultural Economics*, 72(3) : 709–720, 1990.