Université de Montréal

Simulation-based Inference and Nonlinear Canonical Analysis in Financial Econometrics

par

Pascale Valéry

Département de sciences économiques
Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de Philosophiae Doctor (Ph.D.)
en sciences économiques

Février 2005

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée :

Simulation-based Inference and Nonlinear Canonical Analysis in Financial Econometrics

présentée par:

Pascale Valéry

a été évaluée par un jury composé des personnes suivantes :

Nour Méddahi:	président-rapporteur	(Université de Montréal)
Jean-Marie Dufour:	directeur de recherche	(Université de Montréal)
Christian Gouriéroux:	codirecteur de recherche	(University of Toronto)
Sylvia Gonçalves:	membre du jury	(Université de Montréal)
Lynda Khalaf:	examinateur externe	(Université Laval)
Christian Léger:	représentant du doyen de la FES	(Université de Montréal)

Sommaire

L'objectif de cette thèse est d'étudier des techniques d'inférence, classiques et par simulation, en échantillons finis dans le contexte de modèles utilisés en finance.

Dans le premier essai nous introduisons une méthode d'estimation simple, disponible en forme fermée, fondée sur la méthode des moments pour une famille générale de modèles de régression à volatilité stochastique, qui rend possible l'implémentation de procédures d'inférence simulées relativement couteuses en calcul. L'estimateur développé dans cet essai est fondamentalement un estimateur des moments en 2 étapes, qui utilisent les résidus d'une regression préliminaire pour évaluer les conditions de moments de deuxième étape. Sous des conditions de régularité très générales, nous montrons que cet estimateur en 2 étapes est asymptotiquement normalement distribué et en particulier sa matrice de covariance asymptotique ne dépend pas de la distribution de l'estimateur de première étape.

Dans le deuxième essai, nous exploitons la forme fermée de l'estimateur des moments proposé pour implémenter des techniques d'inférence simulée telles que la techniques des tests de Monte Carlo [cf. Dwass (1957), Barnard (1963), Birnbaum (1974)]. En particulier, les tests de Monte Carlo maximisés [cf. Dufour(2002)] autorisent des statistiques de tests dont la distribution dépend de paramètres de nuisance. Dans cette procédure, nous définissons une fonction p-value simulée comme fonction des paramètres de nuisance (sous l'hypothèse nulle), et nous montrons que maximiser cette dernière par rapport aux paramètres de nuisance rapporte un test exact, indépendamment de la taille de l'échantillon et du nombre de réplications utilisées. En particulier, nous implémentons les trois procédures de tests classiques - le test de type Wald, le test de type score et le test de type LR- ainsi que le test de type $c(\alpha)$ introduit par Neyman (1959). Nous proposons également un test de spécification pour le processus de volatilité qui distingue entre une spécification linéaire de la volatilité contre une spécification alternative à intégration fractionnaire.

Dans le troisième essai, nous estimons le modèle de volatilité stochastique par inférence indirecte [cf. Smith (1993), Gouriéroux, Monfort and Renault (1993)] sous des

conditions non regulières. En effet, la condition de rang du jacobien de la fonction de lien asymptotique n'est pas de plein rang en des valeurs isolées du paramètre d'intérêt, condition requise pour que la théorie distributionelle standard dérivée par Gouriéroux, Monfort and Renault (1993) reste valide. En particulier, l'estimateur auxiliaire entrant dans la fonction objectif du critère d'inférence indirecte est fondé sur des conditions de moment qui deviennent nonlinéairement redondantes sous l'hypothèse nulle d'homoskédasticité du processus de volatilité. La matrice de covariance de l'estimateur auxiliaire ainsi que celle des statistiques de Wald et du score deviennent singulières et non inversibles au sens usuel. Pour remédier à ce problème, nous implémentons des techniques de régularisation dont celle proposée par Lütkepohl et Burda (1997) qui consiste à prendre un estimateur de rang réduit pour la matrice de covariance de la statistique de Wald fondé sur l'inverse généralisée de Moore-Penrose. Les techniques de régularisation proposées permettent aux statistiques de test de rester calculables sous des conditions non régulières. Cependant, la théorie distributionnelle développée par Gouriéroux, Monfort et Renault (1993) n'est plus garantie sous des conditions non régulières. Par conséquent, nous combinons des techniques d'inférence par simulation telles que les tests de Monte Carlo maximisés aux statistiques de test modifiées pour rapporter une procédure inférentielle valide en présence d'estimateurs de covariance de rang réduit.

Dans le quatrième essai, nous caractérisons complètement les équations différentielles stochastiques pour lesquelles les fonctions propres du générateur infinitésimal sont des polynômes dans la variable dépendante. En particulier, des transformations affines du processus d'Ornstein-Uhlenbeck, du processus de Cox-Ingersoll-Ross et du processus de Jacobi appartiennent à cette famille d'équations différentielles stochastiques. De tels processus exhibent une structure très particulière des fonctions de dérive et de volatilité de même qu'une forme particulière des valeurs propres.

Dans le cinquième essai, diverses méthodes d'estimation à partir de données discrètes sont inspectées pour estimer un processus de Jacobi appartenant à la classe des processus de diffusion dont les fonctions propres sont des polynômes . Les propriétés

distributionnelles de ce processus autant que sa décomposition canonique non linéaire sous-tendent les méthodes d'estimation retenues. Plus précisément, nous proposons une procédure du maximum de vraisemblance approché fondée sur les fonctions propres. Cette méthode de quasi-vraisemblance est alors comparée à la méthode des moments de Kessler et Sorensen (1999). En effet, alors que nous approchons la fonction de transition inconnue de données discrètes provenant du processus de Jacobi, ces derniers utilisent la décomposition spectrale pour approcher la fonction score inconnue. Des méthodes d'estimation simulées sont aussi considérées parmi lesquelles la méthode des moments simulés et la méthode d'inférence indirecte. Les propriétés statistiques de ces divers estimateurs sont comparées dans des expériences de Monte Carlo.

Mots clés: volatilité stochastique; volatilité à intégration fractionnaire; méthode des moments; tests exacts; test $c(\alpha)$; inférence indirecte; inverses généralisées; processus de diffusion; processus de Jacobi; analyse canonique non-linéaire.

Summary

The objective of this thesis is to study standard and simulation-based inference techniques which are valid in finite samples for models used in finance.

In the first essay, we study a simple moment estimator, available in closed form for general regression models with stochastic volatility models This easy-to-use estimator allows for simulation-based inference techniques which can be computationally expensive. Using residuals from a preliminary regression, the parameters of the stochastic volatility (SV) model are then evaluated by a method-of-moment estimator based on three moments (2S-3M) for which a simple closed-form expression can be derived. Under general regularity conditions, we show the two-stage estimator is asymptotically normally distributed. An interesting and potentially useful feature of the asymptotic distribution stems from the fact its covariance matrix does not depend on the distribution of the conditional mean estimator.

In the second essay, we exploit the closed-form expression of the moment estimator for the parameters of the SV model to implement simulation-based inference techniques such as Monte Carlo (MC) tests [see Dwass (1957), Barnard (1963), Birnbaum (1974)]. More specifically, *maximized* MC tests [see Dufour(2002)] allow for test statistics whose distribution may depend on nuisance parameters. In this procedure, we define a simulated p-value function which is not pivotal under the null hypothesis and we show that maximizing this p-value w.r.t. nuisance parameters does provide an *exact* test, irrespective of the sample size and the number of replications used. We implement the three standard tests- the Wald-type test, the score-type test and the likelihood ratio-type test- but also a $c(\alpha)$-type test introduced by Neyman (1959). We also propose a specification test for the volatility process which discriminates between a *linear* Gaussian specification for the volatility against a *fractionally integrated* Gaussian alternative.

In the third essay, we estimate the SV model by indirect inference [see Smith (1993), Gouriéroux, Monfort and Renault (1993), henceforth (GMR)] under nonregular conditions. More specifically, the rank of the jacobian of the asymptotic binding func-

tion is not of full-column rank at isolated values of the parameter of interest whereas this condition is required for the standard distributional theory derived by GMR(1993) to hold. Indeed, the auxiliary estimator which enters the second step objective criterion in the indirect estimation procedure is based on moment conditions which become nonlinearly redundant under the null hypothesis of homoscedasticity of the volatility process. As a result, the covariance matrix become singular and non invertible in the usual sense. Therefore, we propose to regularize the covariance matrix by resorting to a reduced rank matrix estimator based on generalized inverse among which the Moore-Penrose inverse proposed by Lütkepohl and Burda (1997). We also propose two slightly different regularization techniques among which one that displays good power properties. Further, unlike the nonregularized test statistics, the modified statistics can always be computed under nonregular conditions. However, although the regularization techniques help in keeping the test statistics computable despite some singularity issues, they do not ensure a χ^2 distribution for the modified statistics anymore. As a result, the distributional results developed by GMR (1993) become useless when the jacobian of the asymptotic binding function does not satisfy the required rank condition. One way to overcome this difficulty and still provide valid critical points and p-values, is to resort on *maximized* Monte Carlo tests which achieves in controlling for size distortions irrespective of nuisance parameters in the distribution of the test statistic.

In the fourth essay, we characterize the one-dimensional stochastic differential equations, for which the eigenfunctions of the infinitesimal generator are polynomials in y. In particular, affine transformations of the Ornstein-Uhlenbeck process, the Cox-Ingersoll-Ross process and the Jacobi process belong to this stochastic differential equations family. Such processes exhibit specific patterns of the drift and volatility functions together with a particular form of the eigenvalues.

In the fifth essay, we consider a discretely sampled Jacobi process appropriate to specify the dynamics of a process with range [0,1], such as a discount coefficient, a regime probability, or a state price. The discrete time transition of the Jacobi process does not admit a closed form expression and therefore the exact maximum likelihood

is infeasible. We first review a characterization of the transition function based on nonlinear canonical decomposition. They allow for approximations of the log-likelihood function which can be used to define a quasi-maximum likelihood estimator. The finite sample properties of this estimator are compared with the properties of other estimators proposed in the literature, such as the Kessler and Sorensen's estimator which is a method of moments which also exploits the nonlinear canonical decomposition to approximate the unknown score function [see Kessler and Sorensen (1999)]. It is also compared with generalized method of moments (GMM) estimator, simulated method of moments (SMM) estimator, or indirect inference estimator.

Key words: stochastic volatility; fractionally integrated volatility; moment estimator; exact tests; $c(\alpha)$-test; indirect inference; generalized inverses; diffusion processes; Jacobi process; nonlinear canonical analysis.

A mes parents de la Rivière du Mât...

Remerciements

J'ai de nombreuses personnes et institutions à remercier. Tout d'abord mes directeurs de recherche, Jean-Marie Dufour et Christian Gouriéroux. J'ai eu beaucoup de chance de pouvoir travailler avec des scientifiques et des personnes de cette envergure. Ces années passées sous leur enseignement m'ont apportée un enrichissement incommensurable.

Je voudrai également remercier un de mes coauteurs, Eric Renault avec qui j'ai eu beaucoup de plaisir à travailler. J'ai apprécié également son humour et sa grande personalité.

Je voudrai également remercier Nour Méddahi pour m'avoir prodiguée des commentaires scientifiques très pertinents à plusieurs reprises ainsi que pour son soutien moral.

Je voudrai également remercier George Tauchen de m'avoir procurée les données utilisées dans cette étude.

Je voudrai également remercier les différents organismes qui m'ont accordée bourses et financements sans lesquels je n'aurai pu écrire cette thèse: conseil de recherche en sciences humaines du Canada (CRSH), Fonds québécois de la recherche sur la nature et les technologies, centre interuniversitaire de recherche en économie quantitative (CIREQ), le centre interuniversitaire de recherche en analyse des organisations (CIRANO), l'Institut de Finance Mathématique de Montréal (IFM^2). Je voudrai également remercier HEC-Montréal pour leur support lors de la rédaction finale de cette thèse.

Je voudrai remercier Abdoulaye Diaw pour son aide en informatique. Je voudrai remercier Jean-Pierre Gerbandier pour son soutien moral.

Je voudrai enfin remercier mon conjoint Sylvain Archambault de m'avoir supportée pendant toutes ces années... Je voudrai remercier finalement mes parents et ma soeur Isabelle de m'avoir encouragée tout au long de ce parcours...

Table of Contents

[1] This paper is co-authored with Jean-Marie Dufour.

[2] This paper is co-authored with Jean-Marie Dufour.

[3] This paper is co-authored with Jean-Marie Dufour.

[4] This paper is co-authored with Christian Gouriéroux and Eric Renault.

[5] This paper is co-authored with Christian Gouriéroux.

List of Tables

List of Figures

List of Definitions, Propositions and Theorems

Introduction

La thèse traite de divers sujets d'économétrie financière. Elle est divisée en deux parties. La première partie propose des tests simulés en échantillons finis dans le contexte de modèles utilisés en finance (3 essais) tandis que la seconde partie développe des méthodes d'analyse canonique non linéaire pour des processus de diffusion (2 essais).

Dans la première partie de la thèse nous nous intéressons aux propriétés asymptotiques et en échantillons finis de diverses statistiques de tests dans le cadre du modèle de volatilité stochastique lognormal introduit par Taylor (1986). Depuis, ce modèle a été largement utilisé en finance et plus particulièrement en économétrie de la finance - car il est directement relié aux processus de diffusion très populaires en finance théorique [cf. Wiggins (1987), Melino and Turnbull (1990), Chernov, Gallant, Ghysels and Tauchen (2004)]. Cependant, il reste difficile à estimer en particulier quand il est comparé aux modèles de type GARCH [cf. Engle (1982), Bollerslev (1986)] en raison de l'introduction d'un bruit inobservable dans le processus de volatilité rendant les méthodes d'estimation usuelles - telles le maximum de vraisemblance infaisable. De nombreuses techniques d'estimation alternatives, quasi-exactes [cf. Nelson (1988), Harvey, Ruiz, and Shephard (1994), Ruiz (1994)], GMM [Melino and Turnbull (1990), Andersen and Sørensen (1996)], ou des techniques d'échantillonage fondées sur la simulation telles que le maximum de vraisemblance simulé [Danielsson and Richard (1993), Danielsson (1994)], ou encore l'inférence indirecte [cf. Gouriéroux-Monfort-Renault(1993)] ou encore la méthode efficace des moments de Gallant et Tauchen (1996), [cf. Gallant, Hsieh, and Tauchen (1997), Andersen, Chung, and Sørensen (1999)] – ou encore des méthodes bayesiennes [Jacquier, Polson, and Rossi (1994), Kim, Shephard, and Chib (1998)] ont alors été proposées dans la litérature afin de contourner cette difficulté mais souvent au prix de complications computationelles importantes.

C'est la raison pour laquelle, dans le premier essai nous introduisons une méthode d'estimation simple, disponible en forme fermée, fondée sur la méthode des moments pour une famille générale de modèles de régression à volatilité stochastique, qui rend possible l'implémentation de procédures d'inférence simulées relativement couteuses

en calcul. L'estimateur développé dans cet essai est fondamentalement un estimateur des moments en 2 étapes, qui utilisent les résidus d'une regression préliminaire pour évaluer les conditions de moments de deuxième étape. Sous des conditions de régularité très générales, nous montrons que cet estimateur en 2 étapes est asymptotiquement normalement distribué et en particulier sa matrice de covariance asymptotique ne dépend pas de la distribution de l'estimateur de première étape. Suivant des résultats récents sur l'estimation de modèles autorégressifs à volatilité stochastique [cf. Goncalves-Kilian (2004)], les résultats distributionels développés dans cet essai, restent valides en particulier pour de tels modèles.

Dans le second essai, nous exploitons la forme fermée de l'estimateur des moments proposé pour implémenter des techniques d'inférence simulée telles que la techniques des tests de Monte Carlo [cf. Dwass (1957), Barnard (1963), Birnbaum (1974)]. En particulier, les tests de Monte Carlo maximisés [cf. Dufour (2002)] autorisent des statistiques de tests dont la distribution dépend de paramètres de nuisance. Dans cette procédure, nous définissons une fonction p-value simulée comme fonction des paramètres de nuisance (sous l'hypothèse nulle), et nous montrons que maximiser cette dernière par rapport aux paramètres de nuisance rapporte un test exact, indépendamment de la taille de l'échantillon et du nombre de réplications utilisées. En particulier, nous implémentons les trois procédures de tests classiques - le test de type Wald, le test de type score et le test de type LR- ainsi que le test de type $c(\alpha)$ introduit par Neyman (1959). Nous procédons alors à des comparaisons entre les techniques asymptotiques et les procédures d'inférence simulées. Les résultats exhibent une meilleure performance du test de type $c(\alpha)$. Nous proposons également un test de spécification pour le processus de volatilité qui distingue entre une spécification linéaire de la volatilité contre une spécification alternative à intégration fractionnaire qui présente un intérêt crucial en terme de mémoire longue pour la valorisation d'options [cf. Comte and Renault (1998), Comte, Coutin and Renault (2003), Ohanissian, Russel and Tsay (2003)]. Des expériences de Monte Carlo sont réalisées et suivies par une application empirique sur données journalières pour l'indice de prix composite du Standard and Poor (1928-87).

Dans le troisième essai, nous estimons le modèle de volatilité stochastique par inférence indirecte [cf. Smith (1993), Gouriéroux, Monfort and Renault (1993)] sous des conditions non regulières. En effet, la condition de rang du jacobien de la fonction de lien asymptotique n'est pas de plein rang en des valeurs isolées du paramètre d'intérêt, condition requise pour que la théorie distributionelle standard dérivée par Gouriéroux, Monfort and Renault (1993) reste valide. En particulier, l'estimateur auxiliaire entrant dans la fonction objectif du critère d'inférence indirecte est fondé sur des conditions de moment qui deviennent nonlinéairement redondantes sous l'hypothèse nulle d'homoskédasticité du processus de volatilité. La matrice de covariance de l'estimateur auxiliaire ainsi que celle des statistiques de Wald et du score deviennent singulières et non inversibles au sens usuel. Pour remédier à ce problème, nous implémentons des techniques de régularisation dont celle proposée par Lütkepohl et Burda (1997) qui consiste à prendre un estimateur de rang réduit pour la matrice de covariance de la statistique de Wald fondé sur l'inverse généralisée de Moore-Penrose. Les techniques de régularisation proposées permettent aux statistiques de test de rester calculables sous des conditions non régulières. Cependant, la théorie distributionnelle développée par Gouriéroux, Monfort et Renault (1993) n'est plus garantie sous des conditions non régulières. Par conséquent, nous combinons des techniques d'inférence par simulation telles que les tests de Monte Carlo maximisés aux statistiques de test modifiées pour rapporter une procédure inférentielle valide en présence d'estimateurs de covariance de rang réduit. Des résultats de simulation sur la performance des test modifiés sont présentés suivies d'une illustration financière pour l'indice de prix composite du Standard and Poor (1928-87).

La seconde partie de la thèse est consacrée à l'analyse canonique non linéaire de processus de diffusion dont le but est d'étudier la dépendance temporelle des processus d'une façon moins traditionnelle. Ainsi la décomposition canonique de la distribution conditionnelle permet d'identifier les directions de corrélation maximale entre les variables canoniques ce qui présente un intérêt statégique en finance empirique en particulier en terme de couverture des risques.

Dans le quatrième essai, nous caractérisons complètement les équations différentielles stochastiques pour lesquelles les fonctions propres du générateur infinitésimal sont des polynômes dans la variable dépendante. En particulier, des transformations affines du processus d'Ornstein-Uhlenbeck, du processus de Cox-Ingersoll-Ross et du processus de Jacobi appartiennent à cette famille d'équations différentielles stochastiques. De tels processus exhibent une structure très particulière des fonctions de dérive et de volatilité de même qu'une forme particulière des valeurs propres. En outre, des contraintes de stabilité sont imposées sur les paramètres des processus.

Dans le dernier essai, diverses méthodes d'estimation à partir de données discrètes sont inspectées pour estimer un processus de Jacobi appartenant à la classe des processus de diffusion dont les fonctions propres sont des polynômes . Ce processus prend des valeurs entre 0 et 1, et semble donc adapté pour modéliser des variables dynamiques bornées telle qu'une probabilité de changement de régime, ou capturer l'évolution d'un prix d'état. Les propriétés distributionnelles de ce processus autant que sa décomposition canonique non linéaire sous-tendent les méthodes d'estimation retenues. Plus précisément, nous proposons une procédure du maximum de vraisemblance approché fondée sur les fonctions propres. Cette méthode de quasi-vraisemblance est alors comparée à la méthode des moments de Kessler et Sorensen (1999). En effet, alors que nous approchons la fonction de transition inconnue de données discrètes provenant du processus de Jacobi, ces derniers utilisent la décomposition spectrale pour approcher la fonction score inconnue. L'estimateur de quasi-vraisemblance est aussi comparé à la méthode des moments généralisés de Hansen (1982) puisque la décomposition spectrale de l'opérateur d'espérance conditionelle [cf. Hansen and Sheinckman (1995)] et la forme polynomiale des fonctions propres associées fournissent tous les moments conditionels du processus en terme des moments marginaux. Des méthodes d'estimation simulées sont aussi considérées parmi lesquelles la méthode des moments simulés et la méthode d'inférence indirecte. Les propriétés statistiques de ces divers estimateurs sont comparées dans des expériences de Monte Carlo.

Part I

Simulation-based Inference techniques

Chapitre 1

On a simple closed-form estimator for a stochastic volatility model [1]

[1] This paper is co-authored with Jean-Marie Dufour.

1.1. Introduction

Modelling conditional heteroskedasticity is one of the central problems of financial econometrics. The two main families of models for that purpose consist of GARCH-type processes, originally introduced by Engle (1982), and stochastic volatility (SV) models proposed by Taylor (1986). Although the latter may be more attractive – because they are directly connected to diffusion processes used in theoretical finance – GARCH models are much more popular because they are relatively easy to estimate; for reviews, see Gouriéroux (1997) and Palm (1996). In particular, evaluating the likelihood function of GARCH models is simple compared to stochastic volatility models for which it is very difficult to get a likelihood in closed form; see Shephard (1996), Mahieu and Schotman (1998) and the review of Ghysels, Harvey, and Renault (1996). This is a general feature of almost all nonlinear latent variable models, due to the high dimensionality of the integral defining the likelihood function. As a result, maximum likelihood methods are prohibitively expensive from a computational viewpoint, and alternative methods appear to be required for applying such models.

Since the first discrete-time stochastic volatility models was proposed by Taylor (1986) as an alternative to ARCH models, much progress has been made regarding the estimation of nonlinear latent variable models in general and stochastic volatility models in particular. The methods suggested include quasi maximum likelihood estimation [see Nelson (1988), Harvey, Ruiz, and Shephard (1994), Ruiz (1994)], generalized method-of-moments (GMM) procedures [Melino and Turnbull (1990), Andersen and Sørensen (1996)], sampling simulation-based techniques – such as simulated maximum likelihood [Danielsson and Richard (1993), Danielsson (1994)], indirect inference and the efficient method of moments [Gallant, Hsieh, and Tauchen (1997), Andersen, Chung, and Sørensen (1999)] – and Bayesian methods [Jacquier, Polson, and Rossi (1994), Kim, Shephard, and Chib (1998),Wong (2002a), Wong (2002b)]. Note also that the most widely studied specification in this literature consists of a stochastic volatility model of order one with Gaussian log-volatility and zero (or constant) conditional mean. The most notable exception can be found in Gallant, Hsieh, and Tauchen

(1997) who allowed for an autoregressive conditional mean and considered a general autoregressive process on the log-volatility. It is remarkable that all these methods are highly nonlinear and computer intensive. Implementing them can be quite complicated and get more so as the number of parameters increases (e.g., with the orders of the autoregressive conditional mean and log-volatility).

In this paper, we consider the estimation of stochastic volatility parameters in the context of a linear regression where the disturbances follow a stochastic volatility model of order one with Gaussian log-volatility. The linear regression represents the conditional mean of the process and may have a fairly general form, which includes for example finite-order autoregressions. Our objective is to develop a computationally inexpensive estimator that can be easily exploited within a simulation-based inference procedures, such as Monte Carlo and bootstrap tests.[2] So we study here a simple two-step estimation procedure which can be described as follows: (1) the conditional mean model is first estimated by a simple consistent procedure that does take into account the stochastic volatility structure; for example, the parameters of the conditional mean can be estimated by ordinary least squares (although other estimation procedures can be used); (2) using residuals from this preliminary regression, the parameters of the stochastic model are then evaluated by a method-of-moment estimator based on three moments (2S-3M) for which a simple closed-form expression can be derived. Under general regularity conditions, we show the two-stage estimator is asymptotically normally distributed. Following recent results on the estimation of autoregressive models with stochastic volatility [see, for example, see Theorem 3.1, Gonçalves and Kilian (2004)], this then entails that the result holds for such models. An interesting and potentially useful feature of the asymptotic distribution stems from the fact its covariance matrix does not depend on the distribution of the conditional mean estimator, *i.e.*, the estimation uncertainty on the parameters of the conditional mean does not affect the distribution of the volatility parameter estimates (asymptotically). The properties of the 2S-3M estimator are also studied in a small Monte Carlo experiment and compared

[2] This feature is exploited in a companion paper [Dufour and Valéry (2004)] where various simulation-based test procedures are developed and implemented.

with GMM estimators proposed in this context. We find that the 2S-3M estimator has quite reasonable accuracy with respect to the GMM estimators: indeed, in several cases, the 2S-3M estimator has the lowest root mean square error. With respect to computational efficiency, the 2S-3M estimator always requires less than a second while GMM estimators may need several hours before convergence obtains (if it does). Finally, the proposed estimator is illustrated by applying it to the estimation of Standard and Poor's Composite Price Index (1928-87).

The paper is organized as follows. Section 1.3 sets the framework and the main assumptions made. The closed-form estimator studied is described in section 1.3. The asymptotic distribution of the estimator is established in section 1.4. In section 1.5, we report the results of a small simulation study on the performance of the estimator. Section 1.6 gives an application to the Standard and Poor's Composite Price Index return series in Section 5. We conclude in section 1.7. All proofs are gathered in the Appendix.

1.2. Framework

We consider here a regression model for a variable y_t with disturbances that follow a stochastic volatility process, which is described below following a notation similar to the one used by Gallant, Hsieh, and Tauchen (1997).

Assumption 1.2.1 LINEAR REGRESSION WITH STOCHASTIC VOLATILITY. *The process* $\{y_t : t \in \mathbb{N}_0{}^3\}$ *follows a stochastic volatility model of the type:*

$$y_t = x_t'\beta + u_t \ , \tag{1.2.1}$$

$$u_t = \exp(w_t/2) r_y z_t \ , \quad w_t = \sum_{j=1}^{L_w} a_j w_{t-j} + r_w v_t \ , \tag{1.2.2}$$

where x_t *is a* $k \times 1$ *random vector independent of the variables* $\{x_{\tau-1},\ z_\tau,\ v_\tau,\ w_\tau : \tau \leq t\}$, *and* β, r_y, $\{a_j\}_{j=1}^{L_w}$, r_w *are fixed parameters.*

[3] $\mathbb{N}_0$ refers to the nonnegative integers.

Typically y_t denotes the first difference over a short time interval, a day for instance, of the log-price of a financial asset traded on securities markets. The regression function $x_t'\beta$ represents the conditional mean of y_t (given the past) while the stochastic volatility process determines a varying conditional variance. A common specification here consists in assuming that $x_t'\beta$ has an autoregressive form as in the following restricted version of the model described by Assumption **1.2.1**.

Assumption 1.2.2 AUTOREGRESSIVE MODEL WITH STOCHASTIC VOLATILITY. *The process* $\{y_t : t \in \mathbb{N}_0\}$ *follows a stochastic volatility model of the type:*

$$y_t - \mu_y = \sum_{j=1}^{L_y} c_j(y_{t-j} - \mu_y) + u_t \ , \tag{1.2.3}$$

$$u_t = \exp(w_t/2)r_y z_t \ , \quad w_t = \sum_{j=1}^{L_w} a_j w_{t-j} + r_w v_t \ , \tag{1.2.4}$$

where $\beta, \{c_j\}_{j=1}^{L_y}, r_y, \{a_j\}_{j=1}^{L_w}$ *and* r_w *are fixed parameters.*

We shall refer to the latter model as an AR-SV(L_y, L_w) model. The lag lengths of the autoregressive specifications used in the literature are typically short, e.g.: $L_y = 0$ and $L_w = 1$ [Andersen and Sørensen (1996), Jacquier, Polson, and Rossi (1994), Andersen, Chung, and Sørensen (1999)], $0 \leq L_y \leq 2$ and $0 \leq L_w \leq 2$ [Gallant, Hsieh, and Tauchen (1997)]. In particular, we will devote special attention to the AR-SV(1, 1) model:

$$y_t - \mu_y = c(y_{t-1} - \mu_y) + \exp(w_t/2)r_y z_t \ , \quad |c| < 1 \tag{1.2.5}$$

$$w_t = a w_{t-1} + r_w v_t \ , \quad |a| < 1 \ . \tag{1.2.6}$$

so that

$$Cov(w_t, w_{t+\tau}) = a^\tau \gamma \tag{1.2.7}$$

where $\gamma = r_w^2/(1 - a^2)$. The basic assumptions described above will be completed by a Gaussian distributional assumption and stationarity condition.

Assumption 1.2.3 GAUSSIAN NOISE. *The vectors* $(z_t, v_t)'$, $t \in \mathbb{N}_0$ *are i.i.d. according to a* $\mathrm{N}[0, I_2]$ *distribution.*

Assumption 1.2.4 STATIONARITY. *The process* $s_t = (y_t, w_t)'$ *is strictly stationary.*

The process defined above is Markovian of order $L_s = max(L_y, L_w)$. Under these assumptions, the AR-SV(L_y, L_w) is a parametric model with parameter vector

$$\rho = (\mu_y, c_1, \ldots, c_{L_y}, r_y, a_1, \ldots, a_{L_w}, r_w)'. \tag{1.2.8}$$

Due to the fact that the model involves a latent variable (w_t), the joint density of the vector of observations $\bar{y} = (y_1, \ldots, y_T)$ is not available in closed-form because the latter would involve evaluating an integral with dimension equal to the whole path of the latent volatilities.

1.3. Closed-form method-of-moments estimator

In order to estimate the parameters of the volatility model described in the previous section, we shall consider the moments of the residual process in (1.2.1), which can be estimated relatively easily from regression residuals. Specifically, we will focus on stochastic volatility model of order one ($L_w = 1$). Set

$$\theta = (a, r_y, r_w)', \tag{1.3.1}$$

$$v_t(\theta) \equiv \exp(\frac{aw_{t-1} + r_w v_t}{2}) r_y z_t\,, \ \forall t\,. \tag{1.3.2}$$

Model (1.2.1) - (1.2.2) may then be conveniently rewritten as the following identity:

$$y_t - x_t'\beta = v_t(\theta)\,, \ \forall t\,. \tag{1.3.3}$$

The estimator we will study is based the moments of the process $u_t \equiv v_t(\theta)$. The required moments are given in the following lemma. [4]

Lemma 1.3.1 MOMENTS AND CROSS-MOMENTS OF THE VOLATILITY PROCESS. *Under the assumptions* **1.2.1**, **1.2.3** *and* **1.2.4** *with* $L_w = 1$, *the moments and cross-moments of* $u_t = \exp(w_t/2)r_y z_t$ *are given by the following formulas: for* k, l *and* m *nonnegative integers,*

$$\begin{aligned} \mu_k(\theta) \equiv \mathsf{E}(u_t^k) &= r_y^k \frac{k!}{2^{(k/2)}(k/2)!} \exp\left[\frac{k^2}{8} r_w^2/(1-a^2)\right], \quad \textit{if } k \textit{ is even,} \\ &= 0, \quad \textit{if } k \textit{ is odd,} \end{aligned} \tag{1.3.4}$$

$$\begin{aligned} \mu_{k,l}(m|\theta) &\equiv \mathsf{E}(u_t^k u_{t+m}^l) \\ &= r_y^{k+l} \frac{k!}{2^{(k/2)}(k/2)!} \frac{l!}{2^{(l/2)}(l/2)!} \exp\left[\frac{r_w^2}{8(1-a^2)}(k^2 + l^2 + 2kla^m)\right] \end{aligned} \tag{1.3.5}$$

if k *and* l *are even, and* $\mu_{k,l}(m|\theta) = 0$ *if* k *or* l *is odd.*

On taking $k = 2$, $k = 4$, $k = l = 2$ and $m = 1$, we get:

$$\mu_2(\theta) = \mathsf{E}(u_t^2) = r_y^2 \exp[r_w^2/2(1-a^2)], \tag{1.3.6}$$

$$\mu_4(\theta) = \mathsf{E}(u_t^4) = 3r_y^4 \exp[2r_w^2/(1-a^2)], \tag{1.3.7}$$

$$\mu_{2,2}(1|\theta) = \mathsf{E}[u_t^2 u_{t-1}^2] = r_y^4 \exp[r_w^2/(1-a)]. \tag{1.3.8}$$

An important observation here comes from the fact the above equations can be explicitly solved for a, r_y and r_w. The solution is given in the following lemma.

Lemma 1.3.2 MOMENT EQUATIONS SOLUTION. *Under the assumptions of Propo-*

[4] Expressions for the autocorrelations and autocovariances of u_t^2 were derived by Taylor (1986, Section 3.5) and Jacquier, Polson, and Rossi (1994). The latter authors also provide the higher-order moments $E[|u_t^m|]$, while general formulas for the higher-order cross-moments of a stochastic volatility process are reported (without proof) by Ghysels, Harvey, and Renault (1996). For completeness, we give a relatively simple proof in the Appendix.

sition **1.3.1**, *we have:*

$$a = \left[\frac{\log[\mu_{2,2}(1|\theta)] + \log\left[\mu_4(\theta)/\left(3\mu_2(\theta)^4\right)\right]}{\log\left[\mu_4(\theta)/\left(3\mu_2(\theta)^2\right)\right]} \right] - 1\,, \tag{1.3.9}$$

$$r_y = \frac{3^{1/4}\mu_2(\theta)}{\mu_4(\theta)^{1/4}}\,, \tag{1.3.10}$$

$$r_w = \left[(1-a^2)\log\left[\mu_4(\theta)/\left(3\mu_2(\theta)^2\right)\right]\right]^{1/2}. \tag{1.3.11}$$

From lemmas **1.3.1** and **1.3.3**, it is easy to derive higher-order autocovariance functions. In particular, for later reference, we will find useful to spell out the second and fourth-order autocovariance functions.

Lemma 1.3.3 HIGHER-ORDER AUTOCOVARIANCE FUNCTIONS. *Under the assumptions of Proposition* **1.3.1**, *let* $X_t = (X_{1t},\, X_{2t},\, X_{3t})'$ *where*

$$X_{1t} = u_t^2 - \mu_2(\theta),\; X_{2t} = u_t^4 - \mu_4(\theta),\; X_{3t} = u_t^2 u_{t-1}^2 - \mu_{2,2}(1|\theta)\,. \tag{1.3.12}$$

Then the covariances $\gamma_i(\tau) = \mathrm{Cov}(X_{i,t},\, X_{i,t+\tau})$, $i = 1, 2, 3$, *are given by:*

$$\gamma_1(\tau) = \mu_2^2(\theta)[\exp(\gamma a^\tau) - 1] \tag{1.3.13}$$

$$\gamma_2(\tau) = \mu_4^2(\theta)[\exp(4\gamma a^\tau) - 1]\,,\; \forall \tau \geq 1, \tag{1.3.14}$$

$$\gamma_3(\tau) = \mu_{2,2}^2(1|\theta)[\exp(\gamma(1+a)^2 a^{\tau-1}) - 1]\,,\; \forall \tau \geq 2, \tag{1.3.15}$$

where $\gamma = r_w^2/(1-a^2)$.

Suppose now we have a preliminary estimator $\hat{\beta}$ of β. For example, for the autoregressive model (1.2.3) - (1.2.4), estimation of the equation (1.2.3) yields consistent asymptotically normal estimators of β; see Theorem 3.1, Gonçalves and Kilian (2004) and Kuersteiner (2001). Of course, other estimators of the regression coefficients may be considered. Given the residuals

$$\hat{u}_t = y_t - x_t'\hat{\beta}\,,\; t = 0, 1, \ldots, T\,, \tag{1.3.16}$$

it is then natural to estimate $\mu_2(\theta)$, $\mu_4(\theta)$ and $\mu_{2,2}(1|\theta)$ by the corresponding empirical moments:

$$\hat{\mu}_2 = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2\,, \quad \hat{\mu}_4 = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^4\,, \quad \hat{\mu}_2(1) = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2\hat{u}_{t-1}^2\,.$$

This yields the following estimators of the stochastic volatility coefficients:

$$\hat{a} = \left[\frac{\log[\hat{\mu}_2(1)] + \log\left[\hat{\mu}_4/\left(3\hat{\mu}_2^4\right)\right]}{\log\left[\hat{\mu}_4/\left(3\hat{\mu}_2^2\right)\right]}\right] - 1\,, \tag{1.3.17}$$

$$\hat{r}_y = \frac{3^{1/4}\hat{\mu}_2}{\hat{\mu}_4^{1/4}} = \left(\frac{3\hat{\mu}_2^4}{\hat{\mu}_4}\right)^{1/4}\,, \tag{1.3.18}$$

$$\hat{r}_w = \left[(1-\hat{a}^2)\log\left[\hat{\mu}_4/\left(3\hat{\mu}_2^2\right)\right]\right]^{1/2}\,. \tag{1.3.19}$$

Clearly the latter estimates can be quite easy to compute as soon as the estimator $\hat{\beta}$ used to compute the residuals $\hat{u}_t = y_t - x_t'\hat{\beta}$ is also easy to obtain (e.g., it is a least squares estimator).

1.4. Asymptotic distribution

We will now study the asymptotic distribution of the moment estimator defined in (1.3.17) - (1.3.19). For that purpose, it will be convenient to view the latter as a special case of the general class of estimators obtained by minimizing a quadratic form of the type:

$$M_T(\theta) = [\bar{g}_T(\hat{U}_T) - \mu(\theta)]'\hat{\Omega}_T[\bar{g}_T(\hat{U}_T) - \mu(\theta)] \tag{1.4.1}$$

where $\mu(\theta)$ is a vector of moments, $\bar{g}_T(\hat{U}_T)$ is the corresponding vector of empirical moments based on the residual vector $\hat{U}_T = (\hat{u}_1, \ldots, \hat{u}_T)'$, and $\hat{\Omega}_T$ is a positive-definite (possibly random) matrix. Of course, this estimator belongs to the general family of moment estimators, for which a number of general asymptotic general results do exist; see Volume 1, Chapter 9, Gouriéroux and Monfort (1995b) and Newey and McFadden (1994). However, we need to account here for two specific features, namely:

(1) the disturbances in (1.2.1) follow a stochastic volatility model, and the satisfaction of the relevant regularity conditions must be checked; (2) the two-stage nature of the procedure where the estimator of the parameter β of the conditional mean equation is obtained separately and may not be based on the same objective function as the one used to estimate θ. In particular, it is important to known whether the the estimator of conditional mean parameter β has an effect on the asymptotic distribution of the estimator of θ.

To spell out the properties of the estimator $\hat{\theta}_T(\hat{\Omega}_T)$ obtained by minimizing $M_T(\theta)$, we will consider first the following generic assumptions, where θ_0 denotes the "true" value of the parameter vector θ.

Assumption 1.4.1 ASYMPTOTIC NORMALITY OF EMPIRICAL MOMENTS.

$$\sqrt{T}\left[\bar{g}_T(U_T) - \mu(\theta_0)\right] \xrightarrow{D} \mathrm{N}[0,\, \Omega_*] \tag{1.4.2}$$

where $U_T \equiv (u_1, \ldots, u_T)'$ *and*

$$\Omega_* = \lim_{T\to\infty} \mathsf{E}\{T\left[\bar{g}_T(U_T) - \mu(\theta_0)\right]\left[\bar{g}_T(U_T) - \mu(\theta_0)\right]'\}\,. \tag{1.4.3}$$

Assumption 1.4.2 ASYMPTOTIC EQUIVALENCE FOR EMPIRICAL MOMENTS. *The random vector* $\sqrt{T}[\bar{g}_T(\hat{U}_T) - \mu(\theta_0)]$ *is asymptotically equivalent to* $\sqrt{T}[\bar{g}_T(U_T) - \mu(\theta_0)]$, i.e.

$$\operatorname*{plim}_{T\to\infty}\{\sqrt{T}\left[\bar{g}_T(\hat{U}_T) - \mu(\theta_0)\right] - \sqrt{T}\left[\bar{g}_T(U_T) - \mu(\theta_0)\right]\} = 0\,. \tag{1.4.4}$$

Assumption 1.4.3 ASYMPTOTIC NONSINGULARITY OF WEIGHT MATRIX. $\operatorname*{plim}_{T\to\infty}(\hat{\Omega}_T) = \Omega$ *where* $\det(\Omega) \neq 0$.

Assumption 1.4.4 ASYMPTOTIC NONSINGULARITY OF WEIGHT MATRIX. $\mu(\theta_0)$ *is twice continuously differentiable in an open neighborhood of* θ_0 *and the Jacobian matrix* $P(\theta_0)$ *has full rank, where* $P(\theta) = \frac{\partial \mu'}{\partial \theta}$.

Given these assumptions, the asymptotic distribution of $\hat{\theta}_T(\hat{\Omega}_T)$ is determined by a standard argument on method-of-moments estimation.

Proposition 1.4.5 ASYMPTOTIC DISTRIBUTION OF METHOD-OF-MOMENTS ESTIMATOR. *Under the assumptions* **1.4.1** *to* **1.4.4**,

$$\sqrt{T}\left[\hat{\theta}_T(\Omega) - \theta_0\right] \xrightarrow{D} \mathrm{N}[0,\, V(\theta_0|\Omega)] \tag{1.4.5}$$

where

$$V(\theta|\Omega) = \left[P(\theta)\Omega P(\theta)'\right]^{-1} P(\theta)\Omega\Omega_*\Omega P(\theta)' \left[P(\theta)\Omega P(\theta)'\right]^{-1} \tag{1.4.6}$$

$P(\theta) = \frac{\partial \mu'}{\partial \theta}$. *If, furthermore, (i)* $P(\theta)$ *is a square matrix or (ii)* Ω_* *is nonsingular and* $\Omega = \Omega_*^{-1}$, *then*

$$V(\theta|\Omega) = \left[P(\theta)\Omega_*^{-1}P(\theta)'\right]^{-1} \equiv V_*(\theta)\,. \tag{1.4.7}$$

As usual, $V_*(\theta_0)$ is the smallest possible asymptotic covariance matrix for a method-of-moments estimator based on $M_T(\theta)$. The latter, in particular, is reached when the dimensions of μ and θ are the same, in which case the estimator is obtained by solving the equation

$$\bar{g}_T(\hat{U}_T) = \mu\big(\hat{\theta}_T\big)\,.$$

Consistent estimators $V(\theta_0|\Omega)$ and $V_0(\theta_0)$ can be obtained on replacing θ_0 and Ω_* by consistent estimators.

A consistent estimator of Ω_* can easily be obtained [see Newey and West (1987b)] by a Bartlett kernel estimator, *i.e.*:

$$\hat{\Omega}_* = \hat{\Gamma}_0 + \sum_{k=1}^{K(T)} \Big(1 - \frac{k}{K(T)+1}\Big)(\hat{\Gamma}_k + \hat{\Gamma}_k') \tag{1.4.8}$$

where

$$\hat{\Gamma}_k = \frac{1}{T}\sum_{t=k+1}^{T} [g_{t-k}(\hat{u}) - \mu(\theta)][g_t(\hat{u}) - \mu(\theta)]'$$

with θ replaced by a consistent estimator $\tilde{\theta}_T$ of θ. The truncation parameter $K(T) = \tilde{c}T^{1/3}$ is allowed to grow with the sample size such that:

$$\lim_{T\to\infty} \frac{K(T)}{T^{1/2}} = 0 \ ,$$

[see White and Domowitz (1984)]. A consistent estimator of $V_*(\theta_0)$ is then given by:

$$\hat{V}_* = \left[P(\hat{\theta}_T)\hat{\Omega}_*^{-1}P(\hat{\theta}_T)'\right]^{-1} . \tag{1.4.9}$$

The main problem here consists in showing that the relevant regularity conditions are satisfied for the estimator $\hat{\theta} = (\hat{a}, \hat{r}_y, \hat{r}_w)'$ given by (1.2.5)-(1.2.6) for the parameters of a stochastic volatility model of order one. In this case, we have $\mu(\theta) = [\mu_2(\theta), \mu_4(\theta), \mu_{2,2}(1|\theta)]'$,

$$\bar{g}_T(\hat{U}_T) = \frac{1}{T}\sum_{t=1}^{T} g_t(\hat{U}_T) = \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2 \\ \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^4 \\ \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2\hat{u}_{t-1}^2 \end{pmatrix} \tag{1.4.10}$$

$$\bar{g}_T(U_T) = \frac{1}{T}\sum_{t=1}^{T} g_t(U_T) = \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T}u_t^2 \\ \frac{1}{T}\sum_{t=1}^{T}u_t^4 \\ \frac{1}{T}\sum_{t=1}^{T}u_t^2u_{t-1}^2 \end{pmatrix} \tag{1.4.11}$$

where $g_t(\hat{U}_T) = [\hat{u}_t^2, \hat{u}_t^4, \hat{u}_t^2\hat{u}_{t-1}^2]'$, and $g_t(U_T) = [u_t^2, u_t^4, u_t^2u_{t-1}^2]'$.

Since the number of moments used is equal to the number of parameters (three), the moment estimator can be obtained by taking $\hat{\Omega}_T$ equal to an identity matrix so that Assumption **1.4.3** automatically holds. So the main problem consists in showing that the assumptions **1.4.1** and **1.4.2** are satisfied. For that, it will useful to show the following lemma.

Assumption 1.4.6 EXISTENCE OF MOMENTS. *Let:*

$$\operatorname*{plim}_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}x_t x_t' = \sigma_{2,x}(0)\ , \tag{1.4.12}$$

$$\operatorname*{plim}_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}x_t u_t^2 x_t' = \sigma_{2,x,u}(0,0)\ , \tag{1.4.13}$$

$$\operatorname*{plim}_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}x_t u_{t-1}^2 x_t' = \sigma_{2,x,u}(0,1)\ , \tag{1.4.14}$$

$$\operatorname*{plim}_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}x_{t-1} u_t^2 x_{t-1}' = \sigma_{2,x,u}(1,0)\ , \tag{1.4.15}$$

where the $k \times k$ *matrices* $\sigma_{2,x}(0)$, $\sigma_{2,x,u}(0,0)$, $\sigma_{2,x,u}(0,1)$ *and* $\sigma_{2,x,u}(1,0)$ *are bounded.*

Proposition 1.4.7 ASYMPTOTIC DISTRIBUTION FOR EMPIRICAL MOMENTS. *Under the assumptions* **1.2.1**, **1.2.3** *and* **1.2.4**, *with* $L_w = 1$, *we have:*

$$\sqrt{T}\left[\bar{g}_T(U_T) - \mu(\theta_0)\right] \xrightarrow{D} \mathrm{N}[0,\ \Omega_*] \tag{1.4.16}$$

where $\bar{g}_T(U_T) = \sum_{t=1}^{T} g_t/T$, $g_t = [u_t^2,\ u_t^4,\ u_t^2 u_{t-1}^2]'$, *and*

$$\Omega_* = V[g_t] = \mathsf{E}[g_t g_t'] - \mu(\theta_0)\mu(\theta_0)'\ . \tag{1.4.17}$$

Proposition 1.4.8 ASYMPTOTIC EQUIVALENCE FOR EMPIRICAL MOMENTS. *Suppose the assumptions* **1.2.1**, **1.2.3**, **1.2.4** *and* **1.4.6** *hold with* $L_w = 1$, *let* $\hat{\beta}$ *be an estimator of* β *such that*

$$\sqrt{T}(\hat{\beta} - \beta) \text{ is asymptotically bounded,} \tag{1.4.18}$$

and $\hat{u}_t = y_t - x_t'\hat{\beta}$. *Then* $\sqrt{T}\left[\bar{g}_T(\hat{U}_T) - \mu(\theta_0)\right]$ *is asymptotically equivalent to* $\sqrt{T}\left[\bar{g}_T(U_T) - \mu(\theta_0)\right]$.

The fact that condition (1.4.18) is satisfied by the least squares estimator can be easily seen from earlier published results on the estimation of regression models with stochastic volatility; see Theorem 3.1, Gonçalves and Kilian (2004) and Kuersteiner (2001). Concerning equation 1.4.12 it holds in particular for the AR(p) case with $x_t = Y_{t-1} = (y_{t-1}, \ldots, y_{t-p})'$, [see the proofs of Theorem 3.1, and 3.4, Gonçalves and Kilian (2004)].

On assuming that the matrices Ω_* and $P(\theta_0)$ have full rank, the asymptotic normality of $\hat{\theta}_T$ follows as described in Proposition **1.4.5**. Concerning the latter, it is interesting and potentially useful to note that this asymptotic distribution does not depend on the asymptotic distribution of the first-step estimator of the autoregressive coefficient ($\hat{\beta}$) in the conditional mean equation.

1.5. Simulation study

In this section we study the statistical properties in terms of root mean square error, variance and bias of our moment estimator by simulation. We have considered two different sets of parameters, one set with a low dependency in the autoregressive dynamics of both processes, namely $c = 0.3$ and $a = 0$ while the other one sets $c = 0.95$ and $a = 0.95$. For both sets the scale parameters have been fixed at $r_y = 0.5$ and $r_w = 0.5$. The RMSE are computed on 1000 replications. Our unrestricted estimator available in closed form is denoted by $\hat{\theta}_T$ with 3 moments. As a benchmark, we have taken the moment design used by Jacquier, Polson, and Rossi (1994) and Andersen and Sørensen (1996). In particular we compare our estimator available in closed form to the GMM estimator of Andersen and Sorensen obtained with 5 moments and 24 moments. Globally, the optimality of one estimator over the other one is not so clear since in some situations we are doing better in terms of bias and RMSE than the optimal GMM estimator with 24 moments. the GMM estimator with 5 moments is clearly dominated by our 2S-3M estimator. In terms of variance the GMM estimator with 24 moments performs this time quite better than ours. Indeed, including more moment conditions usually helps in reducing the variance but introduces more bias. In this respect, Ander-

sen and Sorensen did address the choice of the number of moments to include in the overidentified estimation procedure and found that it depends critically on sample size. According to these authors, one should exploit additional moment restrictions when the sample size increases. This advice is not so clear here since our estimator based on the three minimal (for identification) moments performs better than their estimator when the sample size is getting larger, namely for $T = 1000, 2000, 5000$. In this respect, our just identified estimator enhances the widespread idea that one should not include too many instruments increasing thereby the chance of including irrelevant ones in the estimation procedure. This assertion is largely documented in the literature on asymptotic theory [see for example, Buse (1992), Chao and Swanson (2000)]. In particular overidentification increases bias of IV and GMM estimators in finite samples. Dufour and Taamouti (2003) give evidence on that through Monte Carlo methods. Further, when 24 moments are used, it implies to estimate $24(24+1)/2$ separate entries of the weighting matrix along with the sample moments and the GMM estimator becomes thereby computationally cumbersome compared to our estimator available in closed form. Furthermore, when the values of the autoregressive parameters get close to the boundaries of the domain, this creates some numerical instability in estimating the weighting matrix and the situation is getting worse in small samples ($T = 100, 200$). Note that when the sample size is very small ($T = 100, 200$), the RMSE is critically high (between 55% and 84%) especially for the autoregressive parameter a and is due to the extremely poor behavior of sample moments in small samples. A GARCH filter for the volatility process is known to have rather good filtering properties, However, Bayesian estimation of the volatility process is largely considered to be the more efficient way to estimate this process but relies strongly on the choice of an a priori distribution.

1.6. Application to Standard and Poor's price index

In this section, we apply our moment estimator on the Standard and Poor's Composite Price Index (SP), 1928-87.The data have been provided by Tauchen where Efficient Method of Moments have been used by Gallant, Hsieh and Tauchen to fit a standard

stochastic volatility model. The data to which we fit the univariate stochastic volatility model is a long time series comprised of 16,127 daily observations, $\{\tilde{y}_t\}_{t=1}^{16,127}$, on adjusted movements of the Standard and poor's Composite Price Index, 1928-87. The raw series is the Standard and Poor's Composite Price Index (SP),daily, 1928-87. We use a long time series, because, among other things, we want to investigate the long-term properties of stock market volatility through a persistence test. The raw series is converted to a price movements series, $100[\log(SP_t) - \log(SP_{t-1})]$, and then adjusted for systematic calendar effects, that is, systematic shifts in location and scale due to different trading patterns across days of the week, holidays, and year-end tax trading. This yields a variable we shall denote y_t.

The unrestricted estimated value of ρ from the data is:

$$\hat{\rho}_T = (0.129,\ 0.926,\ 0.829,\ 0.427)' ,$$

$$\hat{\sigma}_T = [0.007,\ 2.89,\ 1.91,\ 8.13]' ,$$

where the method-of-moments estimated value of a corresponds to $\hat{a}_T = 0.926$. We may conjecture that there is some persistence in the data during the period 1928-87 what has been statistically checked by performing the three standard tests in a companion paper [see Dufour and Valéry (2004)].

1.7. Conclusion

We provide a computationally simple moment estimator available in close form and derive its asymptotic distribution for the parameters of the stochastic volatility model. Compared with the GMM estimator of Andersen and Sorensen, it demonstrates good statistical properties in terms of bias and RMSE in many situations. Further, it casts some doubt on their advice that one should increase the number of moments to some extent as the sample size grows. In this respect, our just identified estimator enhances the widespread idea that one should not include too many instruments increasing thereby the chance of including irrelevant ones in the estimation procedure. This assertion

is largely documented in the literature on asymptotic theory [see for example, Buse (1992), Chao and Swanson (2000)]. In particular overidentification increases bias of IV and GMM estimators in finite samples. Dufour and Taamouti (2003) give evidence on that through Monte Carlo methods. Further, our closed-form estimator can underlie computationally costly inference techniques like simulation-based inference techniques when asymptotic approximations do not provide reliable inference. Further, our closed-form estimator can be the basis for a easy-to-implement restricted estimator which is deduced from the unrestricted one by simply imposing the constraint in the analytical expression of the former one. This easy-to-implement restricted estimator is very attractive in particular for its simplicity and allows for implementing $C(\alpha)$ tests [see Neyman (1959)] based on any root-n consistent restricted estimator [see Dufour and Valéry (2004)].

Table 1.1. BIAS

BIAS									
	$(c = 0.3,\ a = 0,\ r_y = 0.5,\ r_w = 0.5)$								
	$\hat{\theta}_T$								
	T=100			*T=200*			*T=500*		
	3 mm.	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**
a	-0.2106	-0.0767	0.0780	-0.1554	-0.0522	0.0901	-0.0805	-0.0233	0.0717
r_y	0.0047	-0.0117	-0.0152	0.0044	-0.0021	-0.0064	0.0023	0.0017	-0.0012
r_w	-0.2988	-0.4016	-0.3315	-0.2384	-0.3643	-0.3070	-0.1360	-0.3210	-0.2218
	T=1000			*T=2000*			*T=5000*		
	3 mm.	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**
a	-0.0332	0.0052	0.0186	-0.0204	0.0149	0.0186	-0.0062	0.0191	0.0186
r_y	0.0012	0.0026	0.0009	0.0006	0.0019	0.0009	0.0003	0.0012	0.0009
r_w	-0.0685	-0.3097	-0.0485	-0.0328	-0.3026	-0.0485	-0.0127	-0.2074	-0.0485
	$(c = 0.95,\ a = 0.95,\ r_y = 0.5,\ r_w = 0.5)$								
	T=100			*T=200*			*T=500*		
	3 mm.	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**
a	-0.2490	-0.2904	-0.3400	-0.1576	-0.2652	-0.1327	-0.0921	-0.3209	-0.0257
r_y	0.2063	0.0801	0.0178	0.1754	0.0422	0.0339	0.1379	0.0124	0.0284
r_w	-0.1240	-0.3307	-0.3024	-0.0817	-0.2240	-0.3146	-0.0687	-0.0843	-0.3215
	T=1000			*T=2000*			*T=5000*		
	3 mm.	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**
a	-0.0610	-0.3391	-0.0156	-0.0480	-0.3593	0.0071	-0.0299	-0.3813	0.0256
r_y	0.1149	0.0056	0.0253	0.0890	0.0061	0.0262	0.0639	0.0141	0.0305
r_w	-0.0746	-0.0104	-0.3105	-0.0583	0.0676	-0.2856	-0.0683	0.1988	-0.2461

Table 1.2. VARIANCE

VARIANCE									
	$(c = 0.3,\ a = 0,\ r_y = 0.5,\ r_w = 0.5)$								
	$\hat{\theta}_T$								
	T=100			*T=200*			*T=500*		
	3 mm.	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**
a	0.6482	0.3712	0.2914	0.5434	0.3819	0.2986	0.3346	0.3373	0.2947
ry	0.0019	0.0056	0.0024	0.0010	0.0018	0.0008	0.0005	0.0004	0.0003
rw	0.0572	0.0423	0.0360	0.0593	0.0557	0.0321	0.0436	0.0827	0.0233
	T=1000			*T=2000*			*T=5000*		
	3 mm.	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**
a	0.1686	0.2103	0.0354	0.0862	0.1027	0.0354	0.0276	0.0304	0.0354
ry	0.0002	0.0001	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
rw	0.0200	0.1119	0.0030	0.0092	0.1432	0.0030	0.0029	0.1252	0.0030
	$(c = 0.95,\ a = 0.95,\ r_y = 0.5,\ r_w = 0.5)$								
	T=100			*T=200*			*T=500*		
	3 mm.	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**
a	0.1796	0.3538	0.3019	0.0751	0.3217	0.1634	0.0343	0.3339	0.0426
r_y	0.1184	0.0815	0.0691	0.0647	0.0458	0.0497	0.0284	0.0177	0.0225
r_w	0.1574	0.0607	0.0633	0.1679	0.0979	0.0481	0.1649	0.1254	0.0325
	T=1000			*T=2000*			*T=5000*		
	3 mm.	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**
a	0.0210	0.3336	0.0414	0.0143	0.3309	0.0172	0.0093	0.2911	0.0003
r_y	0.0143	0.0089	0.0115	0.0073	0.0047	0.0056	0.0040	0.0020	0.0021
r_w	0.1522	0.1484	0.0213	0.1432	0.1546	0.0189	0.1312	0.1709	0.0108

Table 1.3. RMSE

RMSE									
	$(c = 0.3,\ a = 0,\ r_y = 0.5,\ r_w = 0.5)$								
	$\hat{\theta}_T$								
	T=100			*T=200*			*T=500*		
	3 mm.	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**
a	0.8318	0.6138	0.5459	0.7530	0.6205	0.5536	0.5837	0.5818	0.5475
ry	0.0439	0.0759	0.0513	0.0320	0.0434	0.0295	0.0226	0.0203	0.0199
rw	0.3827	0.4512	0.3822	0.3408	0.4335	0.3555	0.2491	0.4313	0.2694
	T=1000			*T=2000*			*T=5000*		
	3 mm.	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**
a	0.4118	0.4590	0.4759	0.2942	0.3211	0.3561	0.1662	0.1754	0.1891
r_y	0.0155	0.0140	0.0137	0.0113	0.0101	0.0098	0.0078	0.0070	0.0068
r_w	0.1571	0.4559	0.2000	0.1014	0.4852	0.1393	0.0556	0.4100	0.0732
	$(c = 0.95,\ a = 0.95,\ r_y = 0.5,\ r_w = 0.5)$								
	T=100			*T=200*			*T=500*		
	3 mm.	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**
a	0.4914	0.6617	0.6459	0.3159	0.6351	0.4252	0.2069	0.6607	0.2079
r_y	0.4010	0.2964	0.2634	0.3089	0.2180	0.2255	0.2178	0.1338	0.1527
r_w	0.4155	0.4123	0.3933	0.4176	0.3847	0.3835	0.4116	0.3638	0.3686
	T=1000			*T=2000*			*T=5000*		
	3 mm.	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**
a	0.1573	0.6696	0.2041	0.1291	0.6780	0.1314	0.1014	0.6605	0.0312
r_y	0.1659	0.0944	0.1102	0.1234	0.0686	0.0797	0.0900	0.0460	0.0553
r_w	0.3970	0.3852	0.3431	0.3828	0.3988	0.3170	0.3685	0.4586	0.2673

Chapitre 2

Finite and Large Sample Inference for a Stochastic Volatility Model [1]

[1] This paper is co-authored with Jean-Marie Dufour.

2.1. Introduction

Evaluating the likelihood function of ARCH models is relatively easy compared to Stochastic Volatility models (SV) for which it is impossible to get an explicit closed-form expression for the likelihood function [see Shephard (1996), Mahieu and Schotman (1998)]. This is a generic feature common to almost all nonlinear latent variable models due to the curse of the high dimensionality of the integral appearing in the likelihood function of the stochastic volatility model. This is the reason why econometricians were reluctant to use this kind of models in their applications for a long time since in this setting, maximum likelihood methods are computationally intensive. But ever since progress has been made regarding the estimation of nonlinear latent variable models in general and stochastic volatility models in particular. It mainly exists three types of methods, namely, quasi-exact methods, simulation-based-estimation methods and bayesian methods. Thus, we can mention the Quasi Maximum Likelihood (QML) approach suggested by Nelson (1988) and Harvey, Ruiz and Shephard (1994), Ruiz (1994), a Generalized Method of Moments (GMM) procedure proposed by Melino and Turnbull (1990). On the other hand, increased computer power has made simulation-based estimation methods more attractive among which we can mention the Simulated Method of Moments (SMM) proposed by Duffie and Singleton (1993), the indirect inference approach of Gouriéroux, Monfort and Renault (1993) and the moment matching methods (EMM) of Gallant and Tauchen (1996). But computer intensive Markov Chain Monte Carlo methods applied to SV models by Jacquier, Polson and Rossi (1994) and Kim and Shephard (1994), Kim, Shephard and Chib (1998), Wong(2002a,2002b) and simulation-based Maximum Likelihood (SML) method proposed by Danielsson and Richard (1993), Danielsson (1994), are the most efficient methods to estimate this kind of models. In particular, Danielsson (1994), Danielsson and Richard (1993) develop an importance sampling technique to estimate the integral appearing in the likelihood function of the SV model. In a Bayesian setting, Jacquier, Polson and Rossi (1994), Kim, Shephard and Chib (1998) combine a Gibbs sampler with the Metropolis-Hastings algorithm to obtain the marginal posterior densities of the parameters of the

SV model.

In contrast, the major contribution of this paper is to provide asymptotic and exact inference techniques for testing hypotheses on the parameters of the log-normal stochastic volatility model with an autoregressive mean part. Indeed, the standard form as set forth, for instance, in Harvey, Ruiz, and Shephard (1994), Jacquier, Polson, and Rossi (1994), Danielsson (1994), takes the form of an autoregression whose innovations are scaled by an unobservable volatility process, usually distributed as a lognormal autoregression but other distributions (Student, mixture of normal distributions) can be considered [see Kim, Shephard and Chib (1998), Mahieu and Schotman (1998), Wong (2002a,2002b)]. The stochastic volatility specification we have chosen here comes from Gallant, Hsieh, Tauchen (1997), Tauchen (1997). Whereas all the authors quoted above, mainly focus on estimation performance for the stochastic volatility model, often preoccupied by efficiency considerations [e.g. bayesian methods, Efficient Method of Moments], our paper instead is mostly motivated by inference techniques applied to the stochastic volatility model. Our concern for inference, in particular for simulation-based inference such as the technique of Monte Carlo tests introduced by Dwass (1957) for permutation tests, and later extended by Barnard (1963) and Birnbaum (1974), requires an estimation method easy to implement. Thus, the estimation method used in this paper is mainly a method of moments [see Taylor (1986)] in two steps which coincides with the GMM procedure in the particular case that the autoregressive mean part vanishes. For a detailed presentation of the estimation technique applied to the SV model with an autoregressive conditional mean part, see Dufour and Valéry (2004). As econometricians previously quoted mainly focused on efficient estimation procedures to estimate the SV model, they mostly examined specification tests such as the χ^2 tests for goodness of fit in Andersen and Sorensen (1996), Andersen, Chung and Sorensen (1999), specification tests with diagnostics in Gallant, Hsieh and Tauchen (1997), χ^2 specification tests through Indirect Inference criterion in Monfardini (1997), or likelihood ratio tests statistics for comparative fit in Kim, Shephard and Chib (1998). As a result, inference techniques for testing hypotheses on parameters

of the stochastic volatility model remained underdeveloped, apart from standard t-tests for individual parameters in Andersen and Sorensen (1996), in Andersen, Chung and Sorensen (1999) often performed with size distortions.

In this setting, the aim of the paper is to fulfill the gap for testing hypotheses on parameters of the SV model, more precisely, to propose exact tests in the sense that the tests have correct levels in small samples. To do this, we implement the three standard test statistics that is the Wald-type, score-type and likelihood-ratio-type test based on a computationally simple method-of-moments estimator available in closed form [see Dufour and Valéry (2004)]. We further consider a $c(\alpha)$-type test [see Neyman (1959), Ronchetti (1987), Berger and Wallenstein (1989), Kocherlakota and Kocherlakota (1991)] which is very easy to implement in our framework and demonstrates good size and power properties. Using these test procedures, we test the null hypothesis of no persistence in the volatility against alternatives of strong persistence in the volatility process.Testing for the presence or not of strong serial correlation in the volatility process is relevant mostly for speculative returns which tend to display systematic long-range volatility dependencies in general and more specifically for option pricing predictions. Indeed, a strong serial correlation in the underlying volatility process will help minimizing the pricing error of future option prices computed on the basis of both current realized and implied volatilities. In this respect, a stream of the option pricing literature has seized the importance of this issue by allowing for long-range dependence in the volatility process when compared with the standard stochastic volatility model or the ARCH family, using thereby a fractional integration process whose autocorrelation function is known to decrease at a much slower rate, a hyperbolic decay rate, than that of the standard stochastic volatility process or the ARCH-type family [see Breidt,Crato, Lima (1998) for detection and estimation of a long-memory feature in a discrete time stochastic volatility model , see Comte and Renault (1998) for the continuous time stochastic volatility and Comte, Coutin and Renault (2003), Ohanissian, Russel and Tsay (2003) for its applications to option pricing]. In this regards, we propose a specification test for testing the null hypothesis of linearity in the volatility process against a

fractionally integrated volatility process by means of a likelihood-ratio-type test statistic for comparative fit. Furthermore,we also provide a joint test for testing homoscedasticity in the volatility process. In this respect, a statistical check for homoscedasticity in the stochastic volatility model could be viewed as a relevant pre-test before trying to include a latent factor to drive the dynamic of the volatility process which makes its estimation much more complicated. Testing for homoscedasticity arises strong anomalies as the moment conditions become no more identifying under the null. In presence of such irregularities, the standard asymptotic distribution is known to fail and one has to resort to nonstandard inference techniques or simulation-based inference techniques such as Monte Carlo tests to control for the size.

In a Monte Carlo study we compare the finite sample properties of the standard asymptotic techniques to the technique of Monte Carlo tests which is valid in finite samples and allow for test statistics whose null distribution may depend on nuisance parameters. In particular maximized Monte Carlo tests (MMC) introduced by Dufour (1995) have the exact level in finite samples when the p-value function is maximized over the entire set of nuisance parameters. In contrast to MMC tests which are highly computer intensive, simplified (asymptotically justified) approximate versions of Monte Carlo tests provide a halfway solution which achieves to control the level of the tests while being less computationally demanding. We finally illustrate the test procedures by providing an application on a long time return series on the Standard and Poor's Composite Price Index.

The paper is organized as follows. Section 2 sets the framework and the assumptions underlying the model and reviews the estimation procedure used to implement the tests. Section 3 is devoted to the specification test of linear volatility against fractionally integrated volatility. Hypothesis testing is examined in Section 4 where we also discuss how to build confidence sets by inverting the test statistics. In Section 5 we review the technique of Monte Carlo tests. Simulation results are displayed in Section 6 while empirical results on the Standard and Poor's Composite Price Index return series are discussed in Section 7. We finally conclude in Section 8.

2.2. Framework

The basic form of the stochastic volatility model we study here for y_t comes from Gallant, Hsieh, Tauchen (1997). Let y_t denote the first difference over a short time interval, a day for instance, of the log-price of a financial asset traded on securities markets.

Assumption 2.2.1 *The process* $\{y_t,\ t \in \mathbb{N}\}$ *follows a stochastic volatility model of the type:*

$$y_t - \mu_y = \sum_{j=1}^{L_y} c_j(y_{t-j} - \mu_y) + exp(w_t/2)r_y z_t\ , \tag{2.2.1}$$

$$w_t - \mu_w = \sum_{j=1}^{L_w} a_j(w_{t-j} - \mu_w) + r_w v_t\ , \tag{2.2.2}$$

where μ_y, $\{c_j\}_{j=1}^{L_y}$, r_y, μ_w, $\{a_j\}_{j=1}^{L_w}$ *and* r_w *are the parameters of the two equations, called the mean and volatility equations respectively.* $s_t = (y_t, w_t)'$ *is initialized from its stationary distribution.*

The lag lengths of the autoregressive specifications used in the literature are typically short, e.g. $L_w = 1$, $L_y = 1$, or $L_y = 0$, or $L_w = 2$, $L_y = 2$ [see e.g. Andersen and Sorensen (1996), Gallant, Hsieh, Tauchen (1997), Andersen, Chung and Sorensen (1999)]. In this regards, a simplified version of model (2.2.1)-(2.2.2) consists in setting $\mu_w = 0$ and $c_j = a_j = 0, \forall j \geq 2$, and $\rho = (c, \theta')'$ with $\theta = (a, r_y, r_w)'$. We then have:

$$y_t - \mu_y = c(y_{t-1} - \mu_y) + \exp(w_t/2)r_y z_t\ , \quad |c| < 1 \tag{2.2.3}$$

$$w_t = aw_{t-1} + r_w v_t\ , \quad |a| < 1\ . \tag{2.2.4}$$

We shall call the model represented by equations (2.2.3)-(2.2.4) the stochastic volatility model with an autoregressive mean part of order one [AR(1)-SV for short].

Assumption 2.2.2 *The vectors* $(z_t, v_t)'$, $t \in \mathbb{N}$ *are i.i.d. according to a* $N(0, I_2)$ *distribution.*

Assumption 2.2.3 *The process* $s_t = (y_t, w_t)'$ *is strictly stationary.*

The process is Markovian of order $L_s = max(L_y, L_w)$. Let

$$\rho = (\mu_y, c_1, \ldots, c_{L_y}, r_y, \mu_w, a_1, \ldots, a_{L_w}, r_w)' \tag{2.2.5}$$

denote the parameter vector of the stochastic volatility model. The process $\{y_t\}$ is observed whereas $\{w_t\}$ is regarded as latent. Accordingly, the joint density of the vector of observations $\bar{y} = (y_1, \ldots, y_T)$ is not available in closed form since it requires evaluating an integral with dimension equal to the whole path of the latent volatilities. Let $F(\bar{y}) = F(y_1, \ldots, y_T) = P[Y_1 \leq y_1, \ldots, Y_T \leq y_T|\rho]$ denote its unknown distribution function

To estimate the AR(1)-SV model above, we consider a two-step method whose first step consists in applying ordinary least squares (OLS) to the mean equation which yields a consistent estimate of the autoregressive parameter c and of the mean parameter μ_y, denoted by $\hat{c}_T$, $\hat{\mu}_{yT}$ and the residuals $\hat{u}_t \equiv u_t(\hat{c}_T) = y_t - \mu_y - \hat{c}_T(y_{t-1} - \mu_y)$. Then, we apply in a second step a method of moments to the residuals $\hat{u}_t$ to get the estimate of the parameter $\theta = (a, r_y, r_w)'$ of the mean and volatility equations. In the sequel we will focus on the particular case where $\mu_y = 0$ but all the results still hold in the general case. In the two propositions below, we recall the moments of the volatility process as well as the estimating equations defining the moment estimator of θ. For a detailed proof of these propositions, the reader is referred to Dufour and Valéry (2004).

Proposition 2.2.4 MOMENTS OF THE VOLATILITY PROCESS.
Under Assumptions **2.2.1,2.2.2,2.2.3**, *with* $\mu_y = \mu_w = 0$ *and* $c_j = a_j = 0, \forall\ j \geq 2$. *Then* u_t *has the following moments for even values of* k *and* l*:*

$$\mu_k(\theta) \equiv E(u_t^k) = r_y^k \frac{k!}{2^{(k/2)}(k/2)!} \exp[\frac{k^2}{8} r_w^2/(1 - a^2)] \ , \tag{2.2.6}$$

$$\mu_{k,l}(m|\theta) \equiv E(u_t^k u_{t+m}^l)$$

$$= r_y^{k+l} \frac{k!}{2^{(k/2)}(k/2)!} \frac{l!}{2^{(l/2)}(l/2)!} \exp[\frac{r_w^2}{8(1-a^2)}(k^2+l^2+2kla^m)] \,. \tag{2.2.7}$$

The odd moments are equal to zero.

In particular, for $k = 2$, $k = 4$ and $k = l = 2$ and $m = 1$, we get as in Jacquier, Polson and Rossi (1994):

$$\mu_2(\theta) = E(u_t^2) = r_y^2 \exp[r_w^2/2(1-a^2)] \,, \tag{2.2.8}$$

$$\mu_4(\theta) = E(u_t^4) = 3r_y^4 \exp[2r_w^2/(1-a^2)] \,, \tag{2.2.9}$$

and

$$\mu_{2,2}(1|\theta) = E[u_t^2 u_{t-1}^2] = r_y^4 \exp[r_w^2/(1-a)] \,. \tag{2.2.10}$$

Solving the above moment equations corresponding to $k = 2$, $k = 4$ and $m = 1$ yields the following proposition.

Proposition 2.2.5 ESTIMATING EQUATIONS.

Under the assumptions of Proposition **2.2.4**, *we have:*

$$a = \frac{[\log(\mu_{2,2}(1|\theta)) - \log(3) - 4\log(\mu_2) + \log(\mu_4)]}{\log(\frac{\mu_4}{3(\mu_2)^2})} - 1 \,, \tag{2.2.11}$$

$$r_y = \frac{3^{1/4}\mu_2}{\mu_4^{1/4}} \,, \tag{2.2.12}$$

$$r_w = \left(\log(\frac{\mu_4}{3(\mu_2)^2})(1-a^2)\right)^{1/2} \,. \tag{2.2.13}$$

Given the latter proposition, it is easy to compute a method-of-moments estimator for $\theta = (a, r_y, r_w)'$ replacing the theoretical moments by sample counterparts based on the residuals $\hat{u}_t$. Let $\hat{\theta}_T$ denote the method-of-moments estimator of θ. Typically, $E(u_t^2)$, $E(u_t^4)$ and $E(u_t^2 u_{t-1}^2)$ are approximated by:

$$\hat{\mu}_2 = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t^2 \quad \hat{\mu}_4 = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t^4 \,, \quad \hat{\mu}_2(1) = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t^2 \hat{u}_{t-1}^2$$

respectively. $\hat{\theta}_T$ is consistent and asymptotically normally distributed. See Dufour and Valéry (2004) for an exhaustive presentation of its asymptotic properties.

2.3. Specification test

In this section we propose a specification test to test the null hypothesis of linearity in the volatility process as stated in equation (2.2.4) against the alternative of a fractionally integrated Gaussian process for the volatility where equation (2.2.4) is replaced by:

$$(1-B)^d w_t = \eta_t \quad , \quad \eta_t \overset{i.i.d.}{\sim} N(0, \sigma_\eta^2) \tag{2.3.14}$$

where $d \in (-0.5, 0.5)$. When d is restricted to this domain, w_t is stationary and invertible [see Hosking (1981)]. By denoting $v_t(\tilde{\theta}) = \exp(w_t/2) r_y z_t$ where $\tilde{\theta} = (d, r_y, \sigma_\eta^2)'$, we review the first two moments of $v_t(\tilde{\theta})$ obtained from properties of the log-normal distribution as it is stated in Breidt, Crato and de Lima (1998):

$$\mu_2(\tilde{\theta}) = E(v_t(\tilde{\theta})^2) = r_y^2 \exp[\gamma(0)/2] \, , \tag{2.3.15}$$

$$\mu_4(\tilde{\theta}) = E(v_t(\tilde{\theta})^4) = 3 r_y^4 \exp[2\gamma(0)] \, , \tag{2.3.16}$$

and

$$\mu_{2,2}(h|\tilde{\theta}) = E[v_t(\tilde{\theta})^2 v_{t-h}(\tilde{\theta})^2] = r_y^4 \exp[\gamma(0)(1+\rho(h))] \, , \tag{2.3.17}$$

where the auto-covariance and autocorrelation functions for the long-memory process $\{w_t\}$, denoted by $\gamma(\cdot)$ and $\rho(\cdot)$ are given by:

$$\gamma(0) = \sigma_\eta^2 \frac{\Gamma(1-2d)}{\Gamma^2(1-d)} \, , \tag{2.3.18}$$

$$\rho(h) = \frac{\Gamma(h-d)\Gamma(1-d)}{\Gamma(h-d+1)\Gamma(d)} \, , \quad , h = 1, 2, \ldots , \tag{2.3.19}$$

[see Brockwell and Davis, (1991), p.522]. Then, the likelihood-ratio-type test statistic for comparative fit that is investigated here is given by:

$$\tilde{\xi}_T^C = T[M_T^*(\hat{\theta}_T|\mathcal{M}_0) - M_T^*(\tilde{\theta}_T|\mathcal{M}_1)] \tag{2.3.20}$$

where

$$M_T^*(\theta|\mathcal{M}_i) \equiv [\bar{g}_T(\hat{U}) - \mu(\theta|\mathcal{M}_i)]'\hat{\Omega}^{*-1}[\bar{g}_T(\hat{U}) - \mu(\theta|\mathcal{M}_i)], \quad i = 0, 1 \tag{2.3.21}$$

to test the null hypothesis that the true model, denoted by $\mathcal{M}_0$ is the linear volatility process against the alternative $\mathcal{M}_1$ which is the fractionally integrated gaussian volatility process.

2.4. Tests and confidence sets

In this section we shall set the framework for testing general hypotheses as $H_0 : F \in \mathcal{H}_0$, where $\mathcal{H}_0$ is a subset of all possible distributions for the stochastic volatility model (2.2.3)- (2.2.4), that is,

$$\mathcal{H}_0 \equiv \{F(.) : F(\bar{y}) = F_0(\bar{y}|\psi(\theta)) \; and \; \psi(\theta) = 0\} , \tag{2.4.22}$$

where $\psi(\theta)$ is a $p \times 1$ continuously differentiable function of θ. H_0 is usually abbreviated as: $H_0 : \psi(\theta) = 0$. The derivative of the constraints $P(\theta) = \frac{\partial \psi}{\partial \theta'}$ has full row rank. Let $\hat{\theta}_T$ be the unrestricted estimator and $\hat{\theta}_T^c$ the constrained estimator obtained by minimizing the following criterion under H_0:

$$M_T^*(\theta) \equiv [\bar{g}_T(\hat{U}) - \mu(\theta)]'\hat{\Omega}^{*-1}[\bar{g}_T(\hat{U}) - \mu(\theta)] . \tag{2.4.23}$$

The Wald statistic is defined as

$$\xi_T^W = T\psi(\hat{\theta}_T)'[\hat{P}(\hat{J}'\hat{I}^{-1}\hat{J})^{-1}]^{-1}\psi(\hat{\theta}_T) \tag{2.4.24}$$

where $\hat{P} = P(\hat{\theta}_T)$, $\hat{I} = I(\hat{\theta}_T) = \Omega^*(\hat{\theta}_T)$, $\hat{J} = J(\hat{\theta}_T) = \frac{\partial \mu}{\partial \theta'}(\hat{\theta}_T)$.

The score statistic is defined from the gradient of the objective function with respect to θ evaluated at the constrained estimator. This gradient is given by:

$$\mathcal{D}_T = \frac{\partial \mu'}{\partial \theta}(\hat{\theta}_T^c)\hat{\Omega}^{*-1}(\mu(\hat{\theta}_T^c) - \bar{g}_T(\hat{U})) \tag{2.4.25}$$

and the test statistic is given by

$$\xi_T^S = T\mathcal{D}_T'(\hat{J}_0'\hat{I}_0^{-1}\hat{J}_0)^{-1}\mathcal{D}_T \,, \tag{2.4.26}$$

where $\hat{I}_0 = I(\hat{\theta}_T^c) = \Omega^*(\hat{\theta}_T^c)$, $\hat{J}_0 = J(\hat{\theta}_T^c) = \frac{\partial \mu}{\partial \theta'}(\hat{\theta}_T^c)$. Finally, we can introduce the difference between the optimal values of the objective function that we shall call the LR-type test in the simulations:

$$\xi_T^C = T[M_T^*(\hat{\theta}_T^c) - M_T^*(\hat{\theta}_T)] \,. \tag{2.4.27}$$

The three standard test statistics ξ_T^W , ξ_T^S, and ξ_T^C are known to be asymptotically equivalent and to follow a χ^2 distribution under the null hypothesis.

We also consider the $c(\alpha)$-type test statistic defined by:

$$PC(\tilde{\theta}_T^c) = T[\mu(\tilde{\theta}_T^c) - \bar{g}_T(\hat{U})]'\tilde{W}_0[\mu(\tilde{\theta}_T^c) - \bar{g}_T(\hat{U})] \tag{2.4.28}$$

where $\tilde{W}_0 = \tilde{I}_0^{-1}\tilde{J}_0(\tilde{J}_0'\tilde{I}_0^{-1}\tilde{J}_0)^{-1}\tilde{P}_0'[\tilde{P}_0(\tilde{J}_0'\tilde{I}_0^{-1}\tilde{J}_0)^{-1}\tilde{P}_0']^{-1}\tilde{P}_0(\tilde{J}_0'\tilde{I}_0^{-1}\tilde{J}_0)^{-1}\tilde{J}_0'\tilde{I}_0^{-1}$, with $\tilde{J}_0 = J(\tilde{\theta}_T^c) = \frac{\partial \mu}{\partial \theta'}(\tilde{\theta}_T^c)$, $\tilde{I}_0 = I(\tilde{\theta}_T^c) = \Omega^*(\tilde{\theta}_T^c)$, and $\tilde{P}_0 = P(\tilde{\theta}_T^c)$. $\tilde{\theta}_T^c$ is any root-n consistent estimator of θ that satisfies $\psi(\tilde{\theta}_T^c) = 0$. For our concern, $\tilde{\theta}_T^c$ will be obtained by imposing the constraints in the analytic expressions of the unrestricted method-of-moments estimator $\hat{\theta}_T$ given at equations (2.2.11) to (2.2.13), yielding a consistent restricted estimator without any optimization step. It is known [see Dufour and Trognon (2001, p.8, Proposition 3.1)] that the $c(\alpha)$-type test statistic is asymptotically distributed as a χ^2 variable under the null hypothesis. In the simulations, we

will focus on a particular form of the constraint, i.e. $\psi(\theta) = (1,0)\begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix} = \theta_1$ and the null hypothesis $H_0 : \psi(\theta) = 0$ simplifies to $H_0 : \theta_1 = 0$, (e.g. $\theta_1 \equiv a$, $\theta_1 \equiv (a, r_w)'$).We shall discuss at this stage a few anomalies arising when testing the joint null hypothesis of no heteroscedasticity $H_0 : (a, r_w)' = \underline{0}$ against an alternative of stochastic volatility. We shall stress two interesting findings. The first one is when trying implementing the null hypothesis no heteroscedasticity, the score-type test statistics such as the score statistic and the $c(\alpha)$ statistic become identically null by construction through the derivatives of the moments of the volatility process. In that sense, the score-type test statistics are no longer meaningful under weaker regularity conditions. As a consequence, the test of no heteroscedasticity against an alternative of stochastic volatility is performed by means of the Wald statistic and the LR-type statistic. However, a serious singularity issue arises when implementing the null hypothesis of homoscedasticity, since under the null the moment conditions become nonlinearly redundant. Indeed, the three moment conditions (2.2.8), (2.2.9) and (2.2.10) reduces to only two relevant moment conditions. Hence, the Jacobian of the moment conditions is no more of full-column rank and therefore some singularity problems arise. In such a framework, it is known that the standard asymptotic theory does not provide reliable inference any longer. A simulation exercise strongly highlights the failure of the asymptotic theory when the usual regularity conditions do not hold anymore. In particular, the Wald statistic exhibits severe size distortions for any length of the simulated path. As for the LR-type statistic, it tends to under-reject the null but remains globally valid under nonregular conditions. Indeed, it is known [see Dufour (1997)] that the Wald statistic is not reliable in nonstandard situations whereas the LR statistic still provides reliable inference. In such a context, simulation-based inference such as the technique of Monte Carlo tests presented in the next section, is the solution to correct for these extreme size distortions observed for its asymptotic counterparts.

We also provide confidence sets by inverting the test statistics. Let $S_0 = S(\psi, \bar{y})$ note one of the four previous tests statistics computed from the sample points $\bar{y} = (y_1, \ldots, y_T)$ and under the hypothesis $H_0 : \psi(\theta) = 0$. It is known that there is

a correspondence between confidence sets and tests. The acceptance region of the hypothesis test, the set in the *sample space* for which $H_0 : \psi(\theta) = 0$ is accepted, is given by

$$A(\psi) = \{\bar{y} = (y_1, \ldots, y_T) : S(\psi, \bar{y}) \leq \chi^2_{1-\alpha}\} \tag{2.4.29}$$

for a α level test, and the confidence set, the set in the *parameter space* with plausible values of $\psi(\theta)$, is given by

$$C(y_1, \ldots, y_T) = \{\psi(\theta) : S(\psi, \bar{y}) \leq \chi^2_{1-\alpha}\} = \{\psi(\theta) : G(S(\psi, \bar{y})) \geq \alpha\} , \tag{2.4.30}$$

where $G(.)$ denotes the p-value function. These sets are connected to each other by the tautology

$$(y_1, \ldots, y_T) \in A(\psi) \Leftrightarrow \psi(\theta) \in C(y_1, \ldots, y_T) .$$

The hypothesis test fixes the parameter and asks what sample values (the acceptance region) are consistent with that fixed value. The confidence set fixes the sample value and asks what parameter values (the confidence set) make this sample value most plausible. Thus, if $A(\psi)$ is an acceptance region with level α, we have:

$$P_F[Y \notin A(\psi)] \leq \alpha \quad , \forall F \in \mathcal{H}_0$$

and hence,

$$P_F[Y \in A(\psi)] \geq 1 - \alpha \quad , \forall F \in \mathcal{H}_0 .$$

Then, the coverage probability of the set $C(Y)$ is given by:

$$P_F[\psi(\theta) \in C(Y)] = P_F[Y \in A(\psi)] \geq 1 - \alpha$$

showing that $C(Y)$ is a $1 - \alpha$ confidence set for $\psi(\theta)$.

Following this methodology, we will build confidence sets for the autoregressive parameter of the volatility process by retaining all the values of the parameter for which the p-value function is greater than or equal to $1-\alpha$, yielding a $(1-\alpha)$-level confidence

set.

2.5. Monte Carlo testing

The technique of Monte Carlo tests has originally been suggested by Dwass (1957) for implementing permutation tests, and did not involve nuisance parameters. This technique has been later extended by Barnard (1963) and Birnbaum (1974). This technique has the great attraction of providing *exact* (randomized) tests based on any statistic whose finite sample distribution may be intractable but can be simulated.

We review in this section the methodology of Monte Carlo tests as it is exposed in Dufour (2002),[see also Dufour and Kiviet (1996), Kiviet and Dufour (1997),Dufour and Khalaf (1997), Dufour and Khalaf (2002b), Dufour and Khalaf (2002a), ...] where the distribution of the test statistic S may depend on nuisance parameters. For the test statistics exposed in section 2.4, their asymptotic distribution is asymptotically pivotal (Chi-square distribution), but their finite sample distribution remains unknown. At this stage, we need to make an effort of formalization to clearly expose the procedure. We consider a family of probability spaces $\{(\mathcal{Z}, \mathcal{A}_{\mathcal{Z}}, P_\rho) : \rho \in \Omega\}$ and suppose that S is a real valued $\mathcal{A}_{\mathcal{Z}}$-measurable function whose distribution is determined by $P_{\bar{\rho}}$ where $\bar{\rho}$ is the "true" parameter vector. We wish to test the hypothesis

$$H_0 : \bar{\rho} \in \Omega_0,$$

where Ω_0 is a nonempty subset of Ω. We consider a critical region of the form $S \geq c$, where c is a constant which does not depend on ρ. The critical region $S \geq c$ has *level* α if and only if

$$P_\rho[S \geq c] \leq \alpha, \forall \rho \in \Omega_0,$$

or equivalently,

$$\sup_{\rho \in \Omega_0} P_\rho[S \geq c] \leq \alpha.$$

Furthermore, $S \geq c$ has *size* α when

$$\sup_{\rho \in \Omega_0} P_\rho[S \geq c] = \alpha.$$

We consider a real random variable S_0 and random vectors of the form

$$S(N, \rho) = (S_1(\rho), \ldots, S_N(\rho))', \rho \in \Omega,$$

all defined on a common probability space $(\mathcal{Z}, \mathcal{A}_\mathcal{Z}, P)$, such that the variables $S_0, S_1(\bar{\rho}), \ldots, S_N(\bar{\rho})$ are i.i.d. or exchangeable for some $\bar{\rho} \in \Omega$, each one with distribution function $F[x|\bar{\rho}] = P[S_0 \leq x]$. Typically, S_0 will refer to a test statistic computed from the observed data when the true parameter vector is $\bar{\rho}$ (i.e., $\rho = \bar{\rho}$), while $S_1(\rho), \ldots, S_N(\rho)$ will refer to i.i.d replications of the test statistic obtained independently (e.g., by simulation) under the assumption that the parameter vector is ρ (i.e., $P[S_i(\rho) \leq x] = F[x|\rho]$). In other words, the observed statistic S_0 is simulated by first generating an "observation" vector y according to

$$y = g(\rho, z, v) \tag{2.5.31}$$

where the function g has the bivariate AR(1)-SV specification as stated in equations (2.2.3) and (2.2.4), with $\rho = (c, \mu_y, \theta)'$, $\theta = (a, r_y, r_w)'$. The perturbations z and v have known distributions, which can be simulated ($N(0,1)$ or student, or mixtures, e.g.). We can then compute

$$S(\rho) \equiv S[g(\rho, z, v)] \equiv g_S(\rho, z, v)\ . \tag{2.5.32}$$

The observed statistic S_0 is then computed as $S_0 = S[g(\bar{\rho}, z_0, v_0)]$ and the simulated statistics as $S_i(\rho) = S[g(\rho, z_i, v_i)]$, $i = 1, \ldots, N$ where the random vectors $z_0, z_1, \ldots, z_N$ are i.i.d. (or exchangeable) and $v_0, v_1, \ldots, v_N$ are i.i.d. (or exchangeable) as well.

The technique of Monte Carlo tests provides a simple method allowing one

to replace the theoretical distribution $F(x|\rho)$ by its sample analogue based on $S_1(\rho), \ldots, S_N(\rho)$:

$$\hat{F}_N[x; S(N,\rho)] = \frac{1}{N}\sum_{i=1}^{N} s(x - S_i(\rho)) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{[0,\infty]}(x - S_i(\rho))$$

where $s(x) = 1_{[0,\infty]}(x)$ and $1_A(x)$ is the indicator function associated with the set A. We also consider the corresponding sample tail area function:

$$\hat{G}_N[x; S(N,\rho)] = \frac{1}{N}\sum_{i=1}^{N} s(S_i(\rho) - x)\cdot$$

and the p-value function

$$\hat{p}_N[x|\rho] = \frac{N\hat{G}_N[x|\rho] + 1}{N+1}\,.$$

The sample distribution function is related to the ranks $R_1, \cdots, R_N$ of the variables $S_1(\rho), \ldots, S_N(\rho)$ (when put in ascending order) by the expression:

$$R_j = N\hat{F}_N[S_j; S(N,\rho)] = \sum_{i=1}^{N} s(S_j(\rho) - S_i(\rho)),\ j = 1, \ldots, N.$$

The central property which is exploited here is the following: to obtain critical values or compute p-values, the "theoretical" null distribution $F[x|\bar{\rho}]$ can be replaced by its simulation-based "estimate" $\hat{F}_N[x|\rho] \equiv \hat{F}_N[x; S(N,\rho)]$ in a way that will preserve the level of the test in *finite samples, irrespective of the number N of replications used.* At this stage we shall refer the reader to Dufour (2002, p.13, Proposition 4.1) in which the author states the finite sample validity of Monte Carlo tests when the p-value function is maximized over the entire set of the nuisance parameters.

Therein, the author shows that the critical region $\sup\{\hat{G}_N[S_0|\rho] : \rho \in \Omega_0\} \leq \alpha_1$ has level α irrespective of the presence of nuisance parameters in the distribution of the test statistic S under the null hypothesis $H_0 : \bar{\rho} \in \Omega_0$. Likewise, the (almost) equivalent randomized critical regions $\inf\{\hat{F}_N[S_0|\rho] : \rho \in \Omega_0\} \geq 1 - \alpha_1$ or $S_0 \geq \sup\{\hat{F}_N^{-1}[1 - \alpha_1|\rho] : \rho \in \Omega_0\}$ are shown to have the same level α as their non-randomized analogues.

Dufour (2002) calls such tests maximized Monte Carlo (MMC) tests. The function $\hat{G}_N[S_0|\rho]$ (or $\hat{p}_N[S_0|\rho]$) is then maximized with respect to $\rho \in \Omega_0$, keeping the observed statistic S_0 and the simulated disturbance vectors $z_1, ..., z_N$ and $v_1, ..., v_N$ fixed. The function $\hat{G}_N[S_0|\rho]$ is a step-type function which typically has zero derivatives almost everywhere, except on isolated points (or manifolds) where it is not differentiable. So it cannot be maximized with usual derivative-based algorithms. However, the required maximizations can be performed by using appropriate optimization algorithms that do not require differentiability, such as *simulated annealing*. For further discussion of such algorithms, the reader may consult Goffe, Ferrier, and Rogers(1994).

On the other hand, Dufour (2002) also proposes simplified (asymptotically justified) approximate versions of Monte Carlo tests where the p-value function may be evaluated either at a consistent point estimate and defines thereby a bootstrap version, or at a consistent set estimate of ρ and defines instead confidence-set-Monte Carlo tests. The author shows [see Dufour, (2002, p.16, Proposition 5.1 and p.19, Proposition 6.3)] that both tests are asymptotically valid in the sense that they have the correct level α asymptotically and the estimated p-values converge to the true p-values. He also assesses the validity of the MMC tests and the asymptotic Monte Carlo tests based on consistent set estimators for general distributions , when ties have non-zero probability [see Dufour, (2002, p.14, Proposition 4.2 and p.17, Proposition 5.2)].

In the remaining of the paper we will implement the maximized and bootstrap versions of the Monte Carlo technique and investigate in a comparative Monte Carlo study their actual size and power performances with respect to those of the standard asymptotic tests developed in section 2.4.

2.6. Simulation results

Here we test the null hypothesis of no-persistence in the volatility, which corresponds to $H_0 : a = 0$ against the alternatives $H_1 : a = 0.8, 0.99$. The nominal level of the tests has been set to $\alpha = 5\%$. M represents the number of replications used to assess the actual size of the tests and has been fixed to $M = 1000$ for all tests. N represents

the number of simulated statistics used in the Monte Carlo tests. T is the sample size of the series y_t whose data generating process is assumed to be specified as in equations (2.2.3)-(2.2.4). Implementation is performed with the GAUSS software version 3.2.37. Note that the autoregressive parameter a in the autoregressive specification of the volatility process is restricted to (-1,1) to ensure the stationarity of the volatility process. At this end, each time the estimate of a falls outside of its domain we truncate the estimator by setting it to $a = 0.99$ when $a >= 1$ and to $a = -0.99$ when $a <= -1$.

The Wald statistic as defined at equation (2.4.24) is evaluated at the unrestricted method-of-moments estimator $\hat{\theta}_{1T}$. The Score statistic as defined at equation (2.4.26) is evaluated at the restricted estimator $\hat{\theta}_T^c$ which minimizes the criterion $M_T^*(\theta)$ defined at equation (2.4.23) submitted to the constraint $a = 0$ whereas $\tilde{\theta}_T^c$ represents another restricted estimator of θ obtained by setting $a = 0$ in the analytic expressions of the unrestricted method-of-moments estimator $\hat{\theta}_T$ given at equations (2.2.11)- (2.2.13). The $c(\alpha)$-type statistic as defined at equation (2.4.28) is evaluated at this restricted estimator $\tilde{\theta}_T^c$ of θ. Further, the LR-type test statistic corresponds to the difference between the optimal values of the objective function. Let $LR(\hat{\Omega}) \equiv \xi_T^C$ [see equation (2.4.27)] where $\hat{\Omega} \equiv \Omega(\hat{\theta}_T)$. The weighting matrix $\hat{\Omega}$ is estimated by a kernel estimator with a fixed-Bandwith Bartlett Kernel, where the lag truncation parameter K has been set to $K = 2$.

Let S denote the test statistic which alternatively will take the form of one of the four test statistics earlier mentioned and let S_0 denote the statistic computed from the "pseudo-true" data obtained by simulation under the data generating process evaluated at the "true" value of the parameter. The asymptotic critical regions used to perform the asymptotic tests are of the form:

$$\mathcal{R}_c = \{S_0 > \chi^2_{1-\alpha}(i) = c_i\}, \quad i = 1, 2, 3$$

with $c_1 = 3.84$, $c_2 = 5.99$ and $c_3 = 7.81$. The critical regions used to perform the

Monte Carlo tests has the following form:

$$\mathcal{R}_c = \{\hat{p}_N[S_0|\hat{\rho}_T^c] < \alpha\}$$

with the p-value function given by:

$$\hat{p}_N[S_0|\rho] = \frac{N\hat{G}_N[S_0|\rho] + 1}{N+1} ,$$

and the survival function given by:

$$\hat{G}_N[S_0; S(N,\rho)] = \frac{1}{N}\sum_{i=1}^{N} s(S_i(\rho) - S_0) .$$

The p-value function is evaluated at a consistent restricted estimator of $\rho = (c, \theta')' = (c, a, r_y, r_w)'$. The critical region used to implement the *maximized* Monte Carlo test correspond to:

$$\mathcal{R}_c = \{\max_{\rho \in \Theta_0} \hat{p}_N[S_0|\rho] < \alpha\} ,$$

where the p-value function is maximized on a neighborhood of the restricted estimate of ρ. We use a grid with increment equal to 0.1 to compute the p-value function in the neighborhood. The simulated statistics $S_i(\rho)$ $i = 1, ..., N$ will always be evaluated under the null hypothesis in the Monte Carlo tests whatever the hypothesis to be tested. α has been set to $\alpha = 5\%$. Monte Carlo tests whose p-value function is evaluated at a consistent point estimate of the nuisance parameters follow the methodology presented in section 2.5.

2.6.1. Size investigation

We study the actual size of the various tests compare them to their nominal size fixed at $\alpha = 5\%$. Concerning the specification test, we study in Table 2.1 the actual size of rejecting the null hypothesis of a linear autoregressive volatility specification against an alternative fractionally integrated gaussian volatility process. The parameters have

been set to $c = 0.3$, $r_y = r_w = 0.5$ and the autoregressive volatility parameter $a = 0.3$. As usually encountered in specification tests, the test underreject the null in small samples and requires at least $T = 5000$ observations to reach the nominal level stated at $\alpha = 5\%$.

The results reported in the top part of Table 2.2 for rejecting the null hypothesis $H_0 : a = 0$ display evidence for the asymptotic tests of under-rejecting H_0 for the Wald and the $C(\alpha)$ tests particularly in small samples, whereas the score-type and the LR-type tests tend to over-reject. In particular the underrejection under the null tends to induce a loss of power under the alternative. By contrast, we can see in the bottom part of Table 2.2 that the technique of MC tests achieves in correcting for the size distortions of the asymptotic tests. We also investigate in Table 2.3, a joint test of homoscedasticity in the stochastic volatility model by testing the null hypothesis $H_0 : a = 0, r_w = 0$ by means of the Wald-type and LR-type statistics. The score-type test statistics have been evacuated here since they are identically null by construction. The asymptotic critical value is given by the 95%-quantile of the chi-square distribution with two degrees of freedom which correspond to $c_2 = 5.99$. Note the extremely huge over-rejection (more than 90%) displayed by the asymptotic Wald test when usual regularity conditions are not satisfied. Whatever sample size is considered, the situation is not getting better. Concerning the LR statistic behavior, it tends to slightly overreject in small samples and underreject in large samples. Once again we can note in Table 2.3,that Monte Carlo tests achieve in correcting the severe size distortions observed for the asymptotic tests. More specifically, the Wald statistic performs extremely poorly for the joint null hypothesis $H_0 : a = 0, r_w = 0$ whereas the LR statistic is more reliable. Indeed, the estimators used to construct the test statistics, are based on the moments of the volatility process but under this joint null hypothesis these moment conditions become nonlinearly redundant. As a consequence, the Jacobian of the moment conditions is no more of full-column rank under the null causing some singularity issue for the covariance matrices. It is known [see Dufour (1997)] that the Wald statistic is not reliable under nonregular conditions whereas the LR statistic still provides reliable inference.

It is worth noting in Table 2.4 that when the Monte Carlo tests (MC) evaluated at a consistent restricted estimate of the nuisance parameter fail to correct for the size distortions observed in small samples ($T = 50, 100$) for the LR statistic, its maximized version (MMC) does correct for the size distortions. Indeed, we observe in Table 2.4 that MMC test achieves in reaching the correct level stated at $\alpha = 5\%$ in small samples ($T = 50, 100$) whereas MC tests remains around 10%. The MMC version is performed by maximizing the p-value function on a neighborhood of the restricted estimate of the nuisance parameters which are c and r_w.

2.6.2. Power investigation

Here we study the actual power of the different tests. Note that the standard asymptotic tests for testing the null hypothesis $H_0 : a = 0$ have been corrected for size distortions using the corresponding simulated critical values computed on $M = 10,000$ replications, as reported in Table 2.6 which yields exact 5%-level tests under the null hypothesis. Concerning the specification test, to simulate the model under the alternative of a fractionally integrated gaussian process, we follow Bollerslev and Mikkelsen (1996) [see also Baillie, Bollerslev, and Mikkelsen (1996)] and truncate the moving average filter and then let the process run for a long while to attenuate the effects of transients. Bollerslev and Mikkelsen suggest to truncate at $k = 1000$ but since the moving average coefficients become very small after 160, we chose to truncate at $k = 160$ yielding the moving average filter $\sum_{k=0}^{160} \psi_k B^k$. We then trim off the first 10000 observations. All parameters have been kept to the same values as under the null hypothesis with the long memory parameter $d = 0.3$ replacing the autoregressive parameter $a = 0.3$. We then observe that the simulations averaged over 1000 replications, require at least 1000 observations to exhibit sufficient power. Note also that the Monte Carlo tests do gradually loose power when compared to their asymptotic analogues due to some noise introduced by lengthy simulations. In Table 2.7, we observe that both inference techniques, that is the asymptotic and Monte Carlo tests, suffer from a lack of power when the sample sizes are very small ($T = 50, 100, 200$). Note also the increase in power

when we switch from one type of alternative: $H_1 : a = 0.8$ to a more persistent one: $H_1 : a = 0.99$. The power of Monte Carlo tests could be improved in small samples by increasing the number of simulated statistics from $N = 99$ to $N = 299, 499, 999$. Note that although the asymptotic procedure seems in some cases to exhibit more power w.r.t. Monte Carlo tests, the former however remains a not feasible benchmark for real data whose data generating process (DGP) is generally unknown. In this respect the simulation-based inference technique appears more robust to any DGP. Both test procedures have more power when the sample size grows which is intuitive since both tests are asymptotically justified. Further, note that the $c(\alpha)$ test outperforms its competitors at any sample sizes. In particular the $c(\alpha)$ test performs better than the score test statistic whereas both belong to the same score-type family. The $c(\alpha)$ test statistic has besides the advantage of being the easiest to implement since it does not require in our case any optimization procedure. Indeed the restricted estimate of θ is obtained by simply imposing the constraint in the analytical expressions available for the unrestricted moment estimator.

We also examine in Tables 2.8 and 2.9, the power of the joint test of the null hypothesis of homoscedasticity against the alternative $H_1 : a = 0.5, r_w = 0.5$. The Wald-type test has little power compared to the LR-type test which still remains valid under nonstandard conditions. Indeed, the Wald test after being corrected for the size distortions, is not consistent at all when increasing the sample size. In this respect, it is known [see Dufour (1997)], that Wald tests are not reformable in nonstandard situations, whatever asymptotic, Monte Carlo or *maximized* MC tests, exhibit the same inconsistent behavior for the Wald test. By contrast, the LR-type test remains consistent despite some singularity issues, even though its finite and asymptotic distribution may be modified.

Finally, we also provide some plots of the power functions for asymptotic (in dashed line) and Monte Carlo (in cubic line) Wald and LR tests in Figure 2.1, and for score-type and $c(\alpha)$-type tests in Figure 2.2, respectively. Once again, we observe that the $c(\alpha)$ test has more power than its counterparts and displays a much smoother power function when compared to the tests involving the unrestricted estimator (the LR or the

Wald tests). The score-type test also performs better than the LR or the Wald tests.

2.7. Empirical application

In this subsection we test the null hypothesis of no-persistence in the volatility and also the hypothesis of linear specification for the volatility process against the alternative of a fractionally integrated specification from real data (Standard and Poor's Composite Price Index (SP), 1928-87).

2.7.1. Data

The data have been provided by Tauchen where Efficient Method of Moments have been used by Gallant, Hsieh and Tauchen to fit a standard stochastic volatility model. The data to which we fit the univariate stochastic volatility model is a long time series comprised of 16,127 daily observations, $\{\tilde{y}_t\}_{t=1}^{16,127}$, on adjusted movements of the Standard and poor's Composite Price Index, 1928-87. The raw series is the Standard and Poor's Composite Price Index (SP),daily, 1928-87. We use a long time series, because, among other things, we want to investigate the long-term properties of stock market volatility through a persistence test. The raw series is converted to a price movements series, $100[\log(SP_t) - \log(SP_{t-1})]$, and then adjusted for systematic calendar effects, that is, systematic shifts in location and scale due to different trading patterns across days of the week, holidays, and year-end tax trading. This yields a variable we shall denote y_t.

2.7.2. Results

The unrestricted estimated value of ρ from the data is:

$$\hat{\rho}_T = (0.129,\ 0.926,\ 0.829,\ 0.427)' ,$$

$$\hat{\sigma}_T = [0.007,\ 2.89,\ 1.91,\ 8.13]' ,$$

where the method-of-moments estimated value of a corresponds to $\hat{a}_T = 0.926$. We may conjecture that there is some persistence in the data during the period 1928-87 what is statistically checked by performing the tests below. The restricted estimated values of ρ from the data are:

$$\hat{\rho}_T^c = (0.129,\ 0,\ 0.785,\ 1.152)' ,$$

$$\hat{\sigma}_T = [0.007,\ -, 1.95,\ 1.77]' ,$$

and

$$\tilde{\rho}_T^c = (0.129,\ 0, 0.829,\ 1.133)' ,$$

$$\hat{\sigma}_T = [0.007,\ -,\ 1.91,\ 1.66]' .$$

Note the large discrepancy between the unrestricted and restricted estimates of r_w where the restricted estimates are not consistent if the null hypothesis $H_0 : \ a = 0$ is false.

In Table 2.10, we observe that all standard asymptotic tests reject indeed the null hypothesis of no-persistence in the volatility since $S_0 > \chi^2_{1-\alpha}(1) = 3.84$ as well as all the bootstrap tests whose p-value is equal or less than 5%, whatever length of the simulated statistics is used to implement them. Concerning the specification test, the results shown in the bottom part of Table 2.10 give evidence in favor of the null hypothesis of linear volatility against the alternative of a fractionally integrated volatility process as given by the statistic $\tilde{\xi}_T^C$ defined in equations (2.3.20) and (2.3.21). Indeed, the observed statistic ($\tilde{\xi}_T^C = 0.00345$) is much below the asymptotic critical value of $\chi^2_{95}(3) = 7.81$. The same hold for the MC p-values which are around 0.8 and greater than $\alpha = 0.05$.

We also provide in Table 2.11 confidence sets by inverting the corresponding test statistics as exposed in section 2.4. The coverage probabilities for the confidence sets are $1 - \alpha = 95\%$. We can observe that all tests do cover the estimated value of a ,($\hat{a} = 0.926$), at the confidence level of 95%, except for the bootstrap version of the

score test statistic that covers at a confidence level of 93% and is empty at 95%. We may conclude by saying that the data seem to exhibit some persistence features as usually expected from financial data.

2.8. Concluding remarks

The $c(\alpha)$ test outperforms the other types of tests while being the easiest to implement since it does not require in our framework any optimization procedure. It has good statistical properties: a good level and a high power for sufficiently large sample sizes. On the other hand, Monte Carlo tests and *maximized* MC tests appear as a good alternative to the standard asymptotic tests, specifically when the standard asymptotic approach fails - in situations of almost-unidentified models where the modified distribution of the test statistic remains unknown. We may consider as further research an extension of our approach to asymmetric and fat-tailed distributions such as the asymmetric student distribution and shall test the hypothesis of leverage effect in the stochastic volatility model. We may also consider a continuous-time specification of stochastic volatility since all the moments are already available in Meddahi (2002).

Table 2.1. Size of asymptotic and Monte Carlo tests, specification test

LEVELS in % specification test								
	T=50		*T=100*		*T=200*		*T=500*	
	Asy	**MC**	**Asy**	**MC**	**Asy**	**MC**	**Asy**	**MC**
$LR(\hat{\Omega})$	0.2	0.3	0	0.1	0.1	0.1	0	0.1
	T=1000		*T=2000*		*T=5000*			
	Asy	**MC**	**Asy**	**MC**	**Asy**	**MC**	**Asy**	**MC**
$LR(\hat{\Omega})$	0	0.7	0.1	0.7	5.1	1.3	-	-

Table 2.2. Size of asymptotic and Monte Carlo tests, ,$H_0 : \ a = 0$

LEVELS in % (under $H_0 : \ a = 0$)						
	Asymptotic tests					
	T=50	T=100	T=200	T=500	T=1000	T=2000
Wald	0.1	0.7	0.9	2.1	2.4	3.2
$Score(\hat{\Omega}_C)$	7.7	6	2.6	2.8	3.2	3
$LR(\hat{\Omega})$	7.5	4.8	3.8	2.5	3	3.7
$C(\alpha)$	0.4	0.7	2.6	3	2.9	2.9
	Monte Carlo tests					
	T=50	T=100	T=200	T=500	T=1000	T=2000
Wald	5.4	5.1	3	2.6	5.1	5.5
$Score(\hat{\Omega}_C)$	5.2	5.1	6	6	4.7	3
$LR(\hat{\Omega})$	4.2	5.6	5.8	6.6	5.5	4.8
$C(\alpha)$	4.7	4.4	6	6.9	5.4	4

Table 2.3. Size of asymptotic and Monte Carlo tests, $H_0 : \ a = 0, r_w = 0$

LEVELS in % ($H_0 : a = 0, r_w = 0$),(nuisance:$c = 0.3, r_y = 0.5$)						
	Asymptotic joint tests					
	T=50	T=100	T=500	T=1000	T=2000	T=5000
Wald	94.8	91.6	90.7	90	90.2	92.3
$LR(\hat{\Omega})$	8.8	8.9	1.4	0.7	0.5	0.6
	Monte Carlo joint tests					
	T=50	T=100	T=500	T=1000	T=2000	T=5000
Wald	5.5	4.6	3.6	5.8	4.4	4.3
$LR(\hat{\Omega})$	8.1	7.3	4.7	4.5	3.2	4

Table 2.4. Size of asymptotic and Monte Carlo tests, $H_0: a = 0, r_w = 0$

LEVELS in % ($H_0: a = 0, r_w = 0$), **(nuisance:**$c = 0.95, r_y = 0.5$**)**									
	T=50			*T=100*			*T=500*		
	Asy	**MC**	**MMC**	**Asy**	**MC**	**MMC**	**Asy**	**MC**	**MMC**
Wald	93.8	4.3	4.5	92.2	5	4.2	91.1	3	2.9
$LR(\hat{\Omega})$	9.4	10.5	3.3	8.2	9.9	5.2	1.50	6.4	4.9
	T=1000			*T=2000*			*T=5000*		
	Asy	**MC**	**MMC**	**Asy**	**MC**	**MMC**	**Asy**	**MC**	**MMC**
Wald	88.8	5.6	5	90.8	4.4	4.3	91	3.9	3.9
$LR(\hat{\Omega})$	0.6	5.6	4.1	0.4	3.2	3.1	0.6	4.7	4.1

Table 2.5. Power of asymptotic and Monte Carlo tests, specification test

POWER in % specification test								
	T=50		*T=100*		*T=200*		*T=500*	
	Asy	**MC**	**Asy**	**MC**	**Asy**	**MC**	**Asy**	**MC**
$LR(\hat{\Omega})$	7.2	1.5	2	1.8	0.4	8.4	6.8	26
	T=1000		*T=2000*		*T=5000*			
	Asy	**MC**	**Asy**	**MC**	**Asy**	**MC**	**Asy**	**MC**
$LR(\hat{\Omega})$	32.5	33.2	74.4	41.1	83.3	46.5	-	-

Table 2.6. Simulated critical values, under $H_0: a = 0$

Simulated critical values						
	M=10,000 replications					
	T=50	T=100	T=200	T=500	T=1000	T=2000
Wald	0.8458	1.4295	2.8303	2.5826	2.7878	3.0203
$Score(\hat{\Omega}_C)$	1.7051	2.3336	2.6773	2.9260	2.9472	2.9523
$LR(\hat{\Omega})$	5.7228	3.7033	2.7759	3.0385	3.1352	2.9970
$C(\alpha)$	1.7974	2.3030	2.6901	2.8807	2.8879	2.9133

Table 2.7. Power of size-corrected asymptotic and Monte Carlo tests

POWER in % (under H_1)						
	Size-corrected Asymptotic tests $H_1 : a = 0.8$					
	T=50	T=100	T=200	T=500	T=1000	T=2000
Wald	10.9	17	23.4	60.4	84.5	93.2
$Score(\hat{\Omega}_C)$	16.8	25	47	78.6	93.9	97.8
$LR(\hat{\Omega})$	10.3	16.8	37.6	71.5	88.9	96.6
$C(\alpha)$	19.7	30.9	51.8	81.8	96	99.5
	$H_1 : a = 0.99$					
Wald	31.2	59.5	81.5	90.9	99	99.6
$Score(\hat{\Omega}_C)$	39.7	55.7	85.4	97.7	99.3	99.9
$LR(\hat{\Omega})$	25	44.6	77.3	96.7	99.2	99.3
$C(\alpha)$	41.5	68.8	91.6	99.2	99.7	100
	Monte Carlo tests $(N = 99)$ $H_1 : a = 0.8$					
	T=50	T=100	T=200	T=500	T=1000	T=2000
Wald	10.1	11.8	19.4	44.8	68.3	84
$Score(\hat{\Omega}_C)$	15	18.2	27.9	63.3	89.7	96.8
$LR(\hat{\Omega})$	9.4	10	23.4	60.5	83.5	92.4
$C(\alpha)$	21.6	28.8	43.4	74.1	93.5	98.5
	$H_1 : a = 0.99$					
Wald	28.7	54.1	74.6	87.5	96.3	96.5
$Score(\hat{\Omega}_C)$	11.9	22.3	39.6	82.7	94.4	97.8
$LR(\hat{\Omega})$	15.8	29.8	55.6	72.6	98.5	99.2
$C(\alpha)$	36.1	62.6	78.8	91.6	99.6	99.9

Table 2.8. Power of asymptotic and Monte Carlo tests, $H_1 : a = 0.5, r_w = 0.5$, set I

POWER in % (under H_1)						
	Asymptotic joint tests $H_1 : a = 0.5, r_w = 0.5$					
	T=50	T=100	T=500	T=1000	T=2000	T=5000
Wald	15.8	17.6	18.1	12.7	6.7	1.3
$LR(\hat{\Omega})$	10.9	13.3	84.8	99.4	99.9	100
	Monte Carlo joint tests $(N = 499)$					
	T=50	T=100	T=500	T=1000	T=2000	T=5000
Wald	16.1	18.8	18	12.6	6.9	1.6
$LR(\hat{\Omega})$	14.5	15.7	86.5	99.1	99.9	100

Table 2.9. Power of asymptotic and Monte Carlo tests, $H_1 : a = 0.5, r_w = 0.5$, set II

POWER in % ($H_1 : a = 0.5, r_w = 0.5$), **(nuisance:** $c = 0.95, r_y = 0.5$)									
	T=50			*T=100*			*T=500*		
	Asy	**MC**	**MMC**	**Asy**	**MC**	**MMC**	**Asy**	**MC**	**MMC**
Wald	18	16.8	12.8	20.2	17.2	16.6	17.6	16.4	16.2
$LR(\hat{\Omega})$	11	14	3.8	15.4	17.4	11.6	84.6	85.6	85.4
	T=1000			*T=2000*			*T=5000*		
	Asy	**MC**	**MMC**	**Asy**	**MC**	**MMC**	**Asy**	**MC**	**MMC**
Wald	12	11.6	11.5	6	6	6	1	0.8	0.8
$LR(\hat{\Omega})$	99.6	99	99	100	100	100	100	100	100

Table 2.10. Empirical application

data				
$H_0 : a = 0$				
	Asymptotic tests	*Monte Carlo tests*		
	S_0	N=19	N=99	N=999
Wald	206.03	0.05	0.01	0.001
$Score(\hat{\Omega}_C)$	1039.04	0.05	0.01	0.001
$LR(\hat{\Omega})$	63.20	0.05	0.01	0.001
$C(\alpha)$	854.55	0.05	0.01	0.001
specification test				
$\tilde{\xi}_T^C$	0.00345	0.80	0.80	0.789

Table 2.11. Confidence sets

Confidence sets for a, $(1 - \alpha = 95\%)$		
	Asymptotic	*Monte Carlo*
Wald	]0.92,0.93]	[0.92,0.93]
$Score(\hat{\Omega}_C)$	]0.92,0.93]	[0.92,0.93]*
$LR(\hat{\Omega})$	]0.92,0.93]	[0.92,0.93]
$C(\alpha)$	]0.92,0.93]	[0.92,0.93]

Figure 2.1. Asymptotic and Monte Carlo Power functions, Wald and LR tests

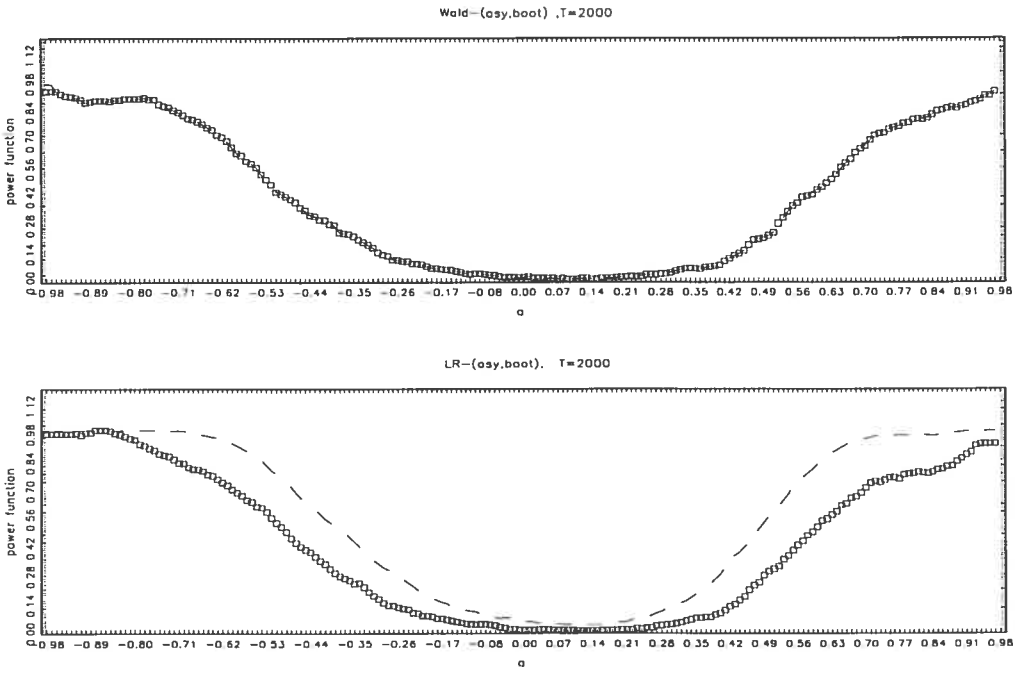
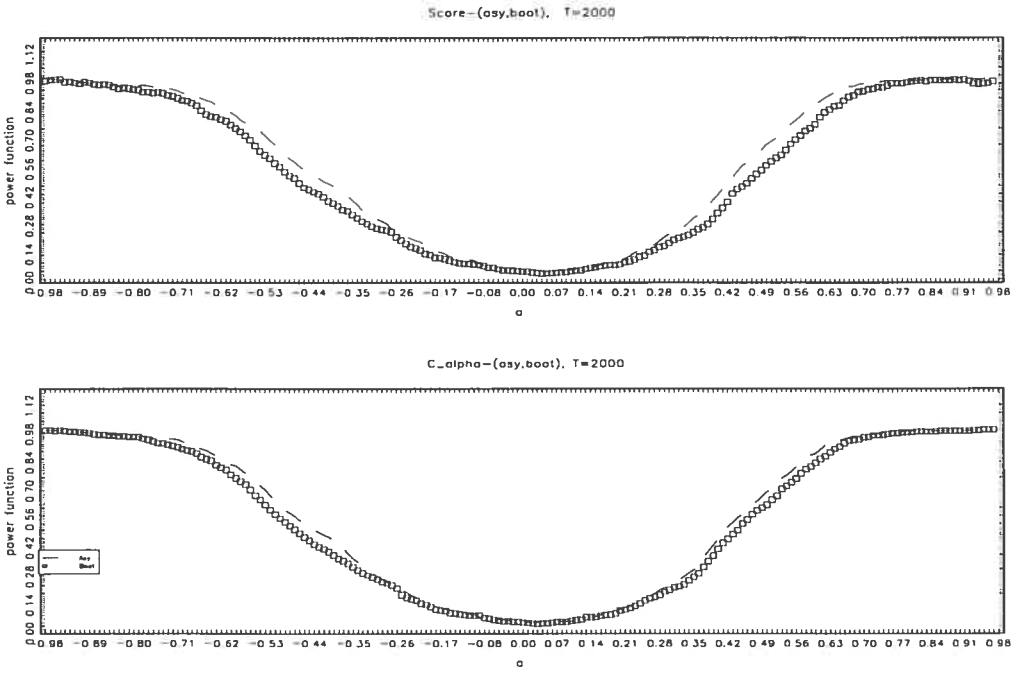


Figure 2.2. Asymptotic and Monte Carlo Power functions, score and $C(\alpha)$ tests



Chapitre 3

Monte Carlo Tests and Regularized Indirect Inference for a Stochastic Volatility Model [1]

[1] This paper is co-authored with Jean-Marie Dufour.

3.1. Introduction

Indirect estimation was proposed by Smith (1993), Gouriéroux, Monfort and Renault (1993) [henceforth GMR] as an estimation and inference procedure for models having complex formulations or untractable likelihood functions. Basically, it consists in optimizing an auxiliary criterion that does not directly provide a consistent estimator of the parameter of interest. A consistent estimator is then obtained by simulation. Indirect inference techniques belong to the class of modern statistical procedures which exploit Monte Carlo methods to derive estimators and tests for complex models. Bootstrap and Monte Carlo Markov chain methods belong to this class and, more generally, any simulation-based inference technique is a potential candidate. The only requirement for implementing simulation-based procedures is that the model or the statistic can be simulated. In this framework, the three standard test statistics have been proposed to make inference on the parameters of interest of the structural model, which are a Wald-type statistic, a likelihood ratio-type statistic and a score-type statistic [see GMR (1993)].

However, the distributional theory associated with those statistics is asymptotic and the choice of the existing statistics importantly depends on the possibility to obtain an asymptotic nuisance-parameter free distribution under the null hypothesis. This opens up the way for approximation errors of any magnitude [see Dufour (1997)]. Further, under nonregular conditions, asymptotic tests are known to have incorrect size even asymptotically on a subset of the parameter space [see Andrews (1987), Gregory and Veall (1985), Breusch and Schmidt (1988), Lutkepohl and Burda (1997, henceforth LB)]. More specifically, LB examined the behavior of the Wald statistic for multi-step causality for finite order vector autoregressive (VAR) processes. In such a setup, multi-step noncausality entails a set of highly nonlinear restrictions on the VAR coefficient matrices. For this type of nonlinear restrictions, standard Wald tests fail to have limiting χ^2-distributions in general. In this respect, LB proposed modifications to the Wald statistic which ensure an asymptotic χ^2-distribution under the null hypothesis. Indeed, Andrews (1987) derived a necessary and sufficient rank condition to be satisfied by the asymptotic covariance matrix and its estimator to ensure a limiting χ^2-distribution for

the Wald statistic under the null.

In this paper, we examine the behavior of the indirect inference procedure and of the resulting test statistics as proposed by GMR (1993) under nonregular conditions, when the simulated binding function does not satisfy the same rank condition (derived by Andrews (1987)) as the population binding function whose rank is getting lower at isolated values θ_0 of the parameter of interest θ. This rank condition can be violated at least in two ways: first,in situations where some parameters become unidentified under the null hypothesis while the other situation is concerned with some (possibly nonlinearly) redundant restrictions under the null. Thus under the null hypothesis, the covariance matrix of the auxiliary estimator and that of the Wald and score statistics become singular. As a result, the usual invertibility technique breaks down making the statistics non implementable. To remedy this problem, we propose to modify the indirect objective function in a way that accounts for singularity problems under the null hypothesis. To do so, we exploit two alternative regularization techniques: the first one was originally proposed by LB (1997) for multi-step noncausality and amounts to estimating a reduced rank covariance matrix and then modifying the Wald statistic accordingly. When the covariance matrix becomes singular, we replace (like LB) the usual inverse by its Moore-Penrose generalized inverse, by setting to zero the inverses of its eigenvalues of the estimated covariance matrix when they drop below a threshold. Alternatively, we propose two slightly different regularization techniques which consists in keeping the eigenvalues of the estimated covariance matrix which are greater than a predetermined threshold and setting the smaller ones to the threshold, instead of zero. Then we can still be proceeding as usual to invert the covariance matrix thus regularized. The third regularization technique is particularly attractive from a power viewpoint. Unlike LB who did regularize the singular covariance matrix of the Wald statistic in the testing problem, we implement these regularization techniques at two levels: one to regularize the indirect estimator at the estimation stage and the other one to regularize the covariance matrices of the test statistics like LB(1997). Indeed, we resort to g-inverted matrices to regularize the Wald-type statistic, the score-type statistic

and the indirect criterion in a situation where the standard regularity conditions required for the parameter estimator to be consistent and asymptotically normal are no more satisfied under the null hypothesis. There have been several papers in the econometric literature dealing with nonregular inference problems such as unidentified parameters under the null, for instance. Andrews (1993b) analyzed tests for structural change. Andrews and Ploberger (1994) explore optimal testing but do not discuss methods to obtain critical values in practice. Andrews (1993a) discusses econometric examples which suffer from the problem of unidentified nuisance parameters.

As an example, we consider testing a null hypothesis of homoskedasticity in the volatility process of a lognormal stochastic volatility (SV) model under which the gradient of the simulated binding function does not satisfy the same rank condition as the gradient of the population binding function. Indeed, the auxiliary estimator which enters the second step objective criterion in the indirect estimation procedure is based on moment conditions which become nonlinearly redundant under the null of homoskedasticity of the volatility process. To account for this singularity issue, we implement the proposed regularization techniques at two distinct levels: one to overcome the singularity problem of the covariance matrix of the auxiliary estimator appearing in the indirect criterion whereas the second-step regularization handles singularity problems occurring for the Wald statistic and the score statistic. Unlike the nonregularized test statistics, the modified statistics can always be computed. They also demonstrate more power than their nonregularized counterparts. These power advantages have already been pointed out by Gallant (1977), Gallant and Tauchen (1989) for taking care of unidentified parameters under H_0 and redundant restrictions. Therefore the regularization techniques appear very useful in two ways, by keeping the statistics computable in nonregular conditions and further by increasing power performances when compared with their nonregularized counterparts. However, although the regularization techniques help in keeping the test statistics computable in such situations, they do not ensure a limiting χ^2 distribution for the modified statistics anymore. As a result, the distributional results developed by GMR (1993) become useless under nonregular

conditions. One way to overcome this difficulty and obtain valid critical points and p-values is to resort on simulation-based inference techniques such as Monte Carlo tests. In the same spirit, Dufour, Khalaf, Bernard and Genest (2004) resort to Monte Carlo tests in non-standard test problems such as the ARCH-M case to circumvent an unidentified nuisance parameter problem and obtain valid p-values. By contrast, Hansen (1996) propose to use a conditional transformation which is analogous to an asymptotic p-value but yields an asymptotic distribution free of nuisance parameters. Then Hansen shows that this transformation can be easily approximated via simulation.

To summarize, there are two main contributions in the paper: the first one consists in modifying the objective function and test statistics in order to account for singular covariance matrices under nonregular conditions, the second one consists in applying the technique of Monte Carlo tests (MC, henceforth)[see Dwass (1957), Barnard (1963), Birnbaum (1974)], and *maximized* Monte Carlo (MMC, henceforth) [see Dufour (2002)] tests to the modified test statistics in order to provide valid critical points and p-values to offset a standard distributional theory which may be misleading under nonregular condition.

The paper is organized as follows. In Section 2, we review the standard indirect inference procedure while in Section 3, we document some singularity issues arising when estimating a log-normal SV model under the null hypothesis of homoscedasticity in the volatility process. In Section 4, we describe the techniques to regularize the singular covariance matrices. In Section 5, we briefly review the methodology of Monte Carlo tests which still provides reliable inference for distributions which are not pivotal even asymptotically. We then provide some simulation results in Section 6 before illustrating the methodology on the Standard and Poor's Composite Price Index (SP),daily, 1928-87 in Section 7. We conclude in Section 8.

3.2. Estimation by Indirect Inference

In this section we review the indirect estimation procedure chosen to estimate a parameter of interest θ. For a more complete description of the method, see Gouriéroux,

Monfort and Renault (1993). The method is proposed for situations where the likelihood function of the structural model is unknown or untractable. To solve this difficulty, one resorts to an approximate model called the auxiliary model which is simpler to estimate. The auxiliary model should closely approximate the distribution of the observed data but does not have to nest it. However, if the auxiliary model nests the structural model then the estimator is as efficient as maximum likelihood [see Gallant and Tauchen (1996)]. Let $M_T(\beta)$ denote the auxiliary criterion parameterized by the auxiliary parameter β. Let us denote by $\hat{\beta}_T$ the solution to this problem:

$$\hat{\beta}_T = \arg\max_{\beta \in B} M_T(\beta) \ . \tag{3.2.1}$$

Then in a second step we can obtain the indirect estimator $\hat{\theta}_T$ by minimizing the second step criterion $M_T(\theta)$ defined by :

$$M_T(\theta) \equiv [\hat{\beta}_T - \frac{1}{S}\sum_{s=1}^{S} \tilde{\beta}_T^{(s)}(\theta)]' \hat{\Omega}_2 [\hat{\beta}_T - \frac{1}{S}\sum_{s=1}^{S} \tilde{\beta}_T^{(s)} \hat{\theta}_T] \ , \tag{3.2.2}$$

where $\hat{\Omega}_2$ is a positive definite matrix defining the metric. $\hat{\beta}_T$ denotes the estimate of the auxiliary parameter based on the observed data whereas $\tilde{\beta}_T^{(s)}(\theta)$ denotes the corresponding estimate for a data set simulated under the structural model for a value θ. Under standard regularity conditions, $\hat{\theta}_T$ is a consistent estimator of the true unknown value θ_0. A consistent estimator of the metric is given by:

$$\hat{\Omega}_2 = J(\hat{\theta})' I(\hat{\theta})^{-1} J(\hat{\theta}) \tag{3.2.3}$$

where

$$J(\theta) = -\frac{\partial^2 M_T(\beta)}{\partial \beta \partial \beta'}(y_T(\theta), \hat{\beta}_T) \ , \tag{3.2.4}$$

$$I(\theta) = \Gamma_0(\theta) + \sum_{k=1}^{K} (1 - \frac{k}{K+1})(\Gamma_k(\theta) + \Gamma_k'(\theta)) \ , \tag{3.2.5}$$

and

$$\Gamma_k(\theta) = \frac{1}{T}\sum_{t=k+1}^{T} \frac{\partial M_{t-k}(\beta)}{\partial \beta'}(y_T(\theta), \hat{\beta}_T)\frac{\partial M_t'(\beta)}{\partial \beta}(y_T(\theta), \hat{\beta}_T) \ . \tag{3.2.6}$$

The metric $\hat{\Omega}_2$ defined at equation (3.2.3) is the metric which minimizes the asymptotic variance-covariance matrix of the indirect estimator, yielding the optimal estimator. This asymptotic variance-covariance matrix is given by

$$W_S^* = (1+\frac{1}{S})\left(\frac{\partial^2 M_\infty}{\partial\theta\partial\beta'}(F(\theta_0),\theta_0,\beta_0)I(\theta_0)^{-1}\frac{\partial^2 M_\infty}{\partial\beta\partial\theta'}(F(\theta_0),\theta_0,\beta_0)\right)^{-1} \tag{3.2.7}$$

where $F(\theta_0)$ is the true unknown probability measure associated with the structural model. A consistent estimator of W_S^* is given by

$$W_S^* = (1+\frac{1}{S})\left(\frac{\partial^2 M_T}{\partial\theta\partial\beta'}(\hat{\theta}_T,\hat{\beta}_T)I(\hat{\theta}_T)^{-1}\frac{\partial^2 M_T}{\partial\beta\partial\theta'}(\hat{\theta}_T,\hat{\beta}_T)\right)^{-1}, \tag{3.2.8}$$

as soon as we can compute the derivative of $\frac{\partial M_T}{\partial\beta}$ with respect to θ. The computation of such a derivative has to be made numerically.

Let us now consider the problem of testing general hypotheses such as $H_0 : F \in \mathcal{H}_0$, where $\mathcal{H}_0$ is a subset of all possible distributions, that is,

$$\mathcal{H}_0 \equiv \{F(.) \ : F(\bar{y}) = F_0(\bar{y}|\psi(\theta)) \ and \ \psi(\theta) = 0\} \ , \tag{3.2.9}$$

where $\psi(\theta)$ is a $p \times 1$ continuously differentiable function of θ. H_0 is usually abbreviated as: $H_0 : \ \psi(\theta) = 0$. The derivative of the constraints $P(\theta) = \frac{\partial\psi}{\partial\theta'}$ has full row rank. Let $\hat{\theta}_T$ be the unrestricted indirect estimator and $\hat{\theta}_T^c$ the constrained estimator obtained by minimizing the second step indirect criterion $M_T(\theta)$ defined in equation (3.2.2) under H_0. To test the null hypothesis we shall consider the three standard test statistics, such as a Wald-type statistic, a likelihood ratio-type statistic and a score-type statistic. The Wald statistic is defined as

$$W = T\psi(\hat{\theta}_T)'[\hat{P}(\hat{J}'\hat{I}^{-1}\hat{J})^{-1}\hat{P}']^{-1}\psi(\hat{\theta}_T) \tag{3.2.10}$$

where $\hat{P} = P(\hat{\theta}_T)$, $\hat{I} = I(\hat{\theta}_T)$, $\hat{J} = J(\hat{\theta}_T)$. The likelihood ratio statistic is the difference between the optimal values of the objective function as defined below:

$$LR = \frac{TS}{1+S}[M_T(\hat{\theta}_T^c) - M_T(\hat{\theta}_T)] \ . \tag{3.2.11}$$

The score-type statistic is defined from the gradient of the indirect objective function with respect to θ evaluated at the restricted estimator $\hat{\theta}_T^c$. This gradient is given by:

$$\mathcal{D}_T = \frac{\partial \tilde{\beta}_T^{(s)}}{\theta}(\hat{\theta}_T^c)\hat{\Omega}_2[\hat{\beta}_T - \frac{1}{S}\sum_{s=1}^{S}\tilde{\beta}_T^{(s)}\hat{\theta}_T^c] \ , \tag{3.2.12}$$

and the test statistic is

$$S = T\mathcal{D}_T'\hat{\Sigma}_S\mathcal{D}_T \ . \tag{3.2.13}$$

Under standard regularity conditions for the estimator and the testing problem, those statistics have been shown [see GMR (1993)] to be asymptotically χ^2-distributed. However, if certain regularity conditions are somehow relaxed, there is no guarantee anymore that the indirect estimator be asymptotically normally distributed, and standard distributional theory for making valid inference collapses.

3.3. Singularity issues: example of a stochastic volatility model

The main purpose of this section is to investigate some *degenerate* testing problems in the sense that some regularity conditions defining the indirect estimator are not satisfied under the null hypothesis. In this respect, one condition required for the indirect estimator to be consistent is the true binding function

$$b[F(\theta_0), \theta_0] = \beta_0$$

being a one-to-one mapping and satisfying the following rank condition that is,

$$\frac{\partial b}{\partial \theta'}[F(\theta_0), \theta_0] \quad \textit{is of full-column rank.} \tag{3.3.14}$$

If this rank condition is not satisfied, then singularity problems can arise. In this respect, the rank condition above can be violated at least in two ways: first, in situations where some parameters become unidentified under the null hypothesis while the other situation is concerned with some (possibly nonlinearly) redundant restrictions under the null. In other words, in order to ensure identification of the parameter of interest θ, the dimension of the auxiliary parameter β is required to be equal or greater than the one of θ. If this condition does not hold, the structural parameter is no more fully identified and so, the standard distributional theory for the indirect estimator and the resulting test statistics may be misleading. Bound et al. (1995), Hall et al. (1996), Maddala and Jeong (1992), Nelson and Startz (1990a, 1990b), Staiger and Stock (1997), and Zivot et al. (1998) give evidences on size distortions when conducting inference with instrumental variables when *weak* instruments are involved. As an example, we are interested in testing the null hypothesis of no stochastic volatility ($H_0 : \ a = 0, \ r_w = 0$) in the stochastic volatility model described below.

Let the structural model be a stochastic volatility model with an autoregressive mean part of order one [AR(1)-SV for short][2]:

$$y_t = \alpha + c y_{t-1} + \exp(w_t/2) r_y z_t \ , \quad |c| < 1 \tag{3.3.15}$$

$$w_t = a w_{t-1} + r_w v_t \ , \quad |a| < 1 \ , \tag{3.3.16}$$

and let $\theta = (c, r_y, a, r_w)'$ denote the parameter of interest. Let $\alpha = \mu_y(1-c)$ where μ_y be the conditional mean of y. The perturbations z_t and v_t are mutually independent and identically distributed $N(0,1)$. Let $\hat{\beta}_T(\hat{U}, \theta)$ denote the functional estimator[3] defined

[2]This AR(p)-SV specification comes from Gallant, Hsieh and Tauchen (1997). The SV model without autoregressive mean part, has extensively been used, in particular, in Harvey, Ruiz, and Shephard (1994), Jacquier, Polson, and Rossi (1994), Danielsson (1994).

[3]See Dufour and Valéry (2004) for a more exhaustive description of the moment estimator used for the AR(1)-SV model.

by:

$$\hat{\beta}_T(\hat{U},\theta) = \arg\min_{\beta\in B}[\bar{g}_T(\hat{U}) - \mu(\theta)]'\hat{\Omega}_1[\bar{g}_T(\hat{U}) - \mu(\theta)] \ , \tag{3.3.17}$$

where $\bar{g}_T(\hat{U}) = \frac{1}{T}\sum_{t=1}^{T} g_t(\hat{U})$ with $g_t(\hat{U}) = (y_t, \hat{u}_t^2, \hat{u}_t^4, \hat{u}_t^2\hat{u}_{t-1}^2)'$, and $\mu(\theta) = (\mu_y, \mu_2, \mu_4, \mu_{2,2}(1|\theta))'$ with

$$\mu_2(\theta) = E(u_t^2) = r_y^2 \exp[r_w^2/2(1-a^2)] \ , \tag{3.3.18}$$

$$\mu_4(\theta) = E(u_t^4) = 3r_y^4 \exp[2r_w^2/(1-a^2)] \ , \tag{3.3.19}$$

and

$$\mu_{2,2}(1|\theta) = E[u_t^2 u_{t-1}^2] = r_y^4 \exp[r_w^2/(1-a)] \ . \tag{3.3.20}$$

As the sample moments used to compute the estimator are computed from residuals from a preliminary regression yielding a $\sqrt{T}$-consistent estimator for the mean parameter, and not from true perturbations, we can correct for the approximation error by simulating the true binding function $b(F(\theta_0),\theta_0)$. In consequence, the functional estimator $\hat{\beta}_T(\hat{U},\theta)$ will tend asymptotically to the true binding function $\beta(U(\theta_0),\theta_0) = b(F(\theta_0),\theta_0)$.

To get an insight on the singularity issue here, let us focus on the simplified model with $c = 0$ and $\theta = (a, r_w, r_y)$. In this context, the binding function depends on the moment conditions given in equations (3.3.18)-(3.3.20), namely

$$\hat{\beta}_T(\hat{U},\theta) = b_T[\hat{U},\mu(\theta)] \ . \tag{3.3.21}$$

In this context, we are interested in testing hypotheses of the form

$$\psi(\theta) = (0,1)\begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix} = \theta_2$$

and the null hypothesis $H_0: \ \psi(\theta) = 0$ simplifies to $H_0: \ \theta_2 = 0$, (e.g. $\theta_2 \equiv (a, r_w)'$). This specific form $H_0: \ (a, r_w)' = \underline{0}$ of the constraint corresponds to testing no heteroskedasticity in the volatility process against an alternative of stochastic volatility.

However, when implementing the null hypothesis of homoskedasticity, some singularity issues arise since under the null, the moment conditions which define the auxiliary estimator become *nonlinearly redundant*. Indeed, the moment conditions (3.3.18) and (3.3.20) reduce to only one relevant moment condition under H_0 up to a nonlinear transformation causing the rank of $\frac{\partial b}{\partial \theta}[U, \theta]$ being lower at the isolated value $\theta_0 = r_y$ under $H_0 : (a, r_w)' = \underline{0}$. Indeed, the determinant of $\frac{\partial b}{\partial \theta}[U, \theta]$ where

$$\frac{\partial b}{\partial \theta}[U, \theta] = \begin{bmatrix} \frac{\partial \mu_2}{\partial a} & \frac{\partial \mu_2}{\partial r_w} & \frac{\partial \mu_2}{\partial r_y} \\ \frac{\partial \mu_4}{\partial a} & \frac{\partial \mu_4}{\partial r_w} & \frac{\partial \mu_4}{\partial r_y} \\ \frac{\partial \mu_{2,2}}{\partial a} & \frac{\partial \mu_{2,2}}{\partial r_w} & \frac{\partial \mu_{2,2}}{\partial r_y} \end{bmatrix} \tag{3.3.22}$$

when evaluated under $H_0 : (a, r_w)' = \underline{0}$, that is

$$\frac{\partial b}{\partial \theta}(U, \theta) = \begin{bmatrix} 0 & 0 & 2r_y \\ 0 & 0 & 12r_y^3 \\ 0 & 0 & 4r_y^3 \end{bmatrix} \tag{3.3.23}$$

is equal to zero and the rank becomes equal to one at $\theta = \theta_0 = r_y$ [see Appendix for the analytical expressions of the derivatives]. But the rank of $\frac{\partial \tilde{\beta}_T^{(s)}}{\partial \theta}(\hat{U}, \theta)$ when evaluated at any value $\theta \neq \theta_0$ is greater than that of $\frac{\partial b}{\partial \theta}(U, \theta_0)$. In consequence, the rank of the gradient of the simulated binding function generally exceed that of the population binding function causing the equality of the rank condition between the population quantity and its estimator to fail. [See Andrews (1987), LB(1997)]. Therefore the whole standard distributional theory derived by GMR (1993) may not hold anymore.

3.4. Regularized Inference

In this section, we examine the singularity problem highlighted in the previous section when studying null hypotheses which causes the rank condition to fail. We will investigate to what extent this may affect standard inference procedures and propose solutions to still conduct valid inference when dealing with singular matrices. More specifically,

the redundant moment conditions under the null hypothesis, creates some singularity problems for the covariance matrix:

$$Var(\sqrt{T}[\hat{\beta}_T - \frac{1}{S}\sum_{s=1}^{S}\tilde{\beta}_T^{(s)}(\theta)]) = (1 + 1/S)\Omega_2^{-1}$$

through the non-invertibility of $I(\theta)$ defined in equations (3.2.5) and (3.2.6). Hence, the usual invertibility of the matrix fails occasionally.

To remedy this problem, we propose to modify the indirect objective function $M_T(\theta)$ defined in equation (3.2.2) in a way that it accounts for singularity problems arising from redundant restrictions under the null hypothesis which causes the rank condition between the gradient of the population binding function and its functional estimator to fail. To do so, we shall exploit two general regularization techniques among which the Moore-Penrose generalized inverse of the corresponding matrix. The idea comes from LB (1997) to use the principal components associated with the largest eigenvalues of the estimated covariance matrix.

To do so, let $\hat{\Sigma}$ be a suitable reduced rank consistent estimator of a covariance matrix Σ with eigenvalues $\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_J$, and $\hat{V}$ an orthogonal matrix consisted of the associated eigenvectors, such that

$$\hat{\Sigma} = \hat{V}\hat{\Lambda}\hat{V}' ,$$

where $\hat{\Lambda} = diag(\hat{\lambda}_1, \ldots, \hat{\lambda}_J)$. For some $c > 0$, define $\hat{J}_c$ to be the number of $\hat{\lambda}_j > c$ and let $\hat{\Lambda}_c = diag(\hat{\lambda}_1, \ldots, \hat{\lambda}_{\hat{J}_c}, 0, \ldots, 0)$. Moreover, define

$$\hat{\Lambda}_c^+ = diag(\hat{\lambda}_1^{-1}, \ldots, \hat{\lambda}_{\hat{J}_c}^{-1}, 0, \ldots, 0) .$$

Then, the Moore-Penrose generalized inverse of $\hat{\Sigma}$ denoted by $\hat{\Sigma}^+$ is obtained as:

$$\hat{\Sigma}^+ = \hat{V}\hat{\Lambda}_c^+\hat{V}' . \tag{3.4.24}$$

Thus, we will denote

$$\hat{I}^{+} = I(\hat{\theta})^{+},$$

$$\hat{\Sigma}_{p}^{+} = [\hat{P}(\hat{J}'\hat{I}^{-1}\hat{J})^{-1}\hat{P}']^{+}$$

and $\hat{\Sigma}_{S}^{+}$, the Moore-Penrose generalized inverse of $I(\hat{\theta})$, $\hat{\Sigma}_{p} = [\hat{P}(\hat{J}'\hat{I}^{-1}\hat{J})^{-1}\hat{P}']$ and $\hat{\Sigma}_{S}$, respectively. When regularizing the estimated covariance matrices by taking their Moore-Penrose generalized inverse as proposed by LB (1997), the modified statistics will be referred to as W^{+} for the modified Wald statistic, LR^{+} for the modified LR statistic and S^{+} for the modified LM statistic.

Alternately, to regularize the estimated covariance matrix

$$\hat{\Sigma} = \hat{V}\hat{\Lambda}\hat{V}'$$

we propose instead to keep the estimated eigenvalues $\hat{\lambda}_j > c$ and set $\hat{\lambda}_j = c$ whenever they drop below the threshold c. For $c > 0$, let $\tilde{J}_c$ be the number of eigenvalues for which $\hat{\lambda}_j > c$. Let

$$\tilde{\Lambda}_c = diag(\hat{\lambda}_1, \ldots, \hat{\lambda}_{\tilde{J}_c}, c, \ldots, c) \ .$$

Thus, the second regularized covariance matrix is obtained as:

$$\hat{\Sigma}^{-} = \hat{V}\tilde{\Lambda}_c\hat{V}' \ . \tag{3.4.25}$$

Finally, a third regularized covariance matrix is obtained as the sum of the non-regularized initial matrix and a regularizing matrix such that:

$$\tilde{\Sigma} = \frac{1}{2}[\hat{V}\hat{\Lambda}\hat{V}' + \hat{V}\tilde{\Lambda}_c\hat{V}'] \ . \tag{3.4.26}$$

Note that $\tilde{\Sigma}$ is always less than or equal to $\hat{\Sigma}^{-}$ by construction and therefore its inverse will always be equal or greater than that of $\hat{\Sigma}^{-}$ and may induce some gains in power. Finally, the inverses of $\hat{\Sigma}^{-}$ and $\tilde{\Sigma}$ are obtained by taking a usual inverse defined for positive definite matrices. In particular, when all eigenvalues are greater than the

threshold, the regularized matrices coincide with the original matrices, since $\tilde{\Lambda}_c = \hat{\Lambda}$. Thus,

$$(\hat{I}^-)^{-1}\,, \qquad (\hat{\Sigma}_p^-)^{-1}\,, \qquad (\hat{\Sigma}_S^-)^{-1}$$

will correspond to the inverses of $I(\hat{\theta})$, $\hat{\Sigma}_p = [\hat{P}(\hat{J}'\hat{I}^{-1}\hat{J})^{-1}\hat{P}']$ and $\hat{\Sigma}_S$ respectively, regularized according to equation (3.4.25). Likewise, we will denote

$$\tilde{I}^{-1} = I(\hat{\theta})^{-1}\,, \quad \tilde{\Sigma}_p^{-1} = [\hat{P}(\hat{J}'\hat{I}^{-1}\hat{J})^{-1}\hat{P}']^{-1}\,, \text{and} \quad \tilde{\Sigma}_S^{-1}$$

the inverse matrices of $I(\hat{\theta})$, $\hat{\Sigma}_p = [\hat{P}(\hat{J}'\hat{I}^{-1}\hat{J})^{-1}\hat{P}']$ and $\hat{\Sigma}_S$ respectively, regularized according to equation (3.4.26). Thus, when using this two regularization techniques, the modified Wald statistics will be referred to as W^-, $\tilde{W}$, the modified LR statistics LR^-, $\tilde{LR}$ and the modified LM statistics as S^-, $\tilde{S}$ accordingly. These modified inverses will be built sequentially if necessary. The first one will help in regularizing the indirect criterion to account for singularity issues, and thereby will benefit to the statistics altogether whereas at the opposite the Wald and score statistics will take advantage of the two inverses jointly when the covariance matrices become singular. In the remaining of the paper, we will compare the modified statistics with the original statistics proposed by GMR (1993). However, although the regularization techniques help in keeping the test statistics computable despite some underidentified parameters, they do not ensure a χ^2 distribution for the modified statistics anymore. As a result, the distributional results developed by GMR (1993) become useless under nonregular conditions (*i.e.* the rank condition of the gradient of the binding function does not hold anymore). One way to overcome this difficulty and still provide valid critical points and p-values, is to resort on simulation-based inference techniques such as Monte Carlo tests whose *maximized* version achieves in controlling for size distortions irrespective of nuisance parameters in the distribution of the test statistic.

3.5. Monte Carlo testing

The technique of Monte Carlo tests has originally been proposed by Dwass (1957) for implementing permutation tests and did not involve nuisance parameters. This technique has been extended by Barnard (1963) and Birnbaum (1974). It has the great attraction of providing *exact* (randomized) tests based on any statistic whose finite sample distribution may be intractable but can be simulated. We briefly review the methodology of Monte Carlo tests covering both cases, first without nuisance parameters and then with nuisance parameters as it is proposed in Dufour (2002). The technique of Monte Carlo tests provides a simple method allowing one to replace the unknown or untractable theoretical distribution $F(x|\theta)$ by its sample analogue based on the statistics $S_1(\theta), \ldots, S_N(\theta)$ simulated under the null hypothesis. The procedure can be designed as follows.

First we present the case without nuisance parameters which provides an *exact* test.

- STEP 1: Using the observed sample, we calculate the relevant statistic denoted by S_0.
- STEP 2: Using draws under H_0, we generate N simulated samples: $S_1, \ldots, S_N$.
- STEP 3: Then we compute the estimated survival function:

$$\hat{G}_N[x; S(N)] = \frac{1}{N} \sum_{i=1}^{N} s(S_i - x).$$

and the associated p-value function

$$\hat{p}_N(x) = \frac{N\hat{G}_N(x) + 1}{N + 1} .$$

If N is chosen so that $\alpha(N + 1)$ is an integer, under H_0:

$$P(\hat{p}_N[S_0] \leq \alpha) = \alpha ,$$

yielding an *exact* test.

Second, in presence of nuisance parameters, Dufour (2002) proposes to maximize the nuisance parameters over the parameter space conformable with the null hypothesis. In this case the procedure is the following.

- STEP 1: To test the null hypothesis

$$H_0 : \bar{\theta} \in \Omega_0 ,$$

  we use first the observed sample to calculate the relevant statistic denoted by S_0.

- STEP 2: For each $\theta \in \Omega_0$, we generate N replications of S: $S_1(\theta), \ldots, S_N(\theta)$.

- STEP 3: Using these simulations we compute the corresponding simulated p-value function:

$$\hat{p}_N[x|\theta] = \frac{N\hat{G}_N[x|\theta] + 1}{N+1} .$$

Finally the p-value function $\hat{p}_N[S_0|\theta]$ as a function of θ is maximized over the parameter space. If the number of simulated statistics N is chosen so that $\alpha(N+1)$ is an integer, then we have under H_0:

$$P[\sup\{\hat{p}_N(S_0|\theta) : \theta \in \Omega_0\} \leq \alpha] \leq \alpha ,$$

that is we control for the size. Such a technique which provides an *exact* test irrespective of the presence of nuisance parameters under the null hypothesis is called a *Maximized Monte Carlo* test (henceforth MMC) by Dufour (2002). A proof of this assertion can be found in Dufour (2002). In the simulation exercises below we will implement the test in two forms, one in a local maximized version we call (MMC) and another one when the nuisance parameters are evaluated at a consistent point estimate yielding a form of parametric bootstrap we shall call (MC) tests. For the MMC version the nuisance parameters are maximized over a fine grid since there are only two nuisance parameters. When the nuisance parameters are numerous one can use simulated annealing [see Goffe, Ferrier and Rogers (1994)] an appropriate optimization algorithm which does not require differentiability. Indeed $\hat{G}_N[S_0|\theta]$ is step-type function which

typically has zero derivatives almost everywhere, except on isolated points where it is not differentiable.

3.6. Simulation results

In this section, we implement the Wald test (W), the Likelihood ratio test (LR) and the score-type test (S) for testing the null hypothesis of homoskedasticity in the volatility process, say, $H_0 : a = 0, r_w = 0$. The tests are performed in three ways. The first one uses the asymptotic χ^2 critical point ($\chi^2_{1-\alpha}(1) = 3.84$) for $\alpha = 5\%$ determined by the rank of the gradient of the population binding function which is equal to 1 under H_0, while the other ones are based on the simulated p-values. For the Monte Carlo test (hereafter, MC), the p-value is evaluated at a consistent restricted point estimate of the nuisance parameters. Concerning the maximized Monte Carlo test (hereafter, MMC), the p-value function is maximized over a neighborhood of the restricted estimate of the nuisance parameters using a grid with increments equal to 0.1. The null hypothesis is rejected each time the maximized p-value is less than the nominal level fixed at 5%. We assess the actual sizes of the tests averaged on 100 replications. The Monte Carlo tests are performed with $N = 19$ statistics simulated under the null hypothesis. Under the alternative, the length of the simulated statistics is increased to $N = 99$ to gain in power. The nuisance parameters have been set to $r_y = 0.4$ and $c = 0.95$ to produce a high level of persistence in the mean equation. In the simulations the drift parameter α has been fixed at 0.5 throughout the experiment. The simulations are run on the GAUSS software (3.2.37 version). Concerning the regularization techniques, in order to facilitate comparisons across methods, the thresholds have been set to $c = 0.01$ for $T = 2000, 1000$ and $c = 0.1$ for $T = 500, 200$ for both methods. We need to increase the threshold when the sample size decreases due to the "poor" performance of the indirect estimator in small samples. Indeed, we have to impose stronger regularization in small samples to force convergence otherwise the algorithm breaks down quite often. In this respect the rate at which the threshold should tend to zero with the sample size has to be investigated. LB (1997) gives some device on this issue suggesting to use the convergence rate

of the estimated eigenvalues toward the true ones. In practice, this device does not work that well and has to be investigated empirically. The LR statistic modified according to LB regularization and denoted by LR^+, will only benefit from the first regularization for computing $\hat{I}^+ = I(\hat{\theta})^+ = \hat{V}\hat{\Lambda}_{c_1}^+\hat{V}'$, with $\hat{\Lambda}_{c_1}^+ = diag(\hat{\lambda}_1^{-1}, \ldots, \hat{\lambda}_{\hat{J}_{c_1}}^{-1})$ and $\hat{J}_{c_1}$ denotes the number of $\hat{\lambda}_j$ of the estimated covariance matrix $I(\hat{\theta})$ which satisfy: $\hat{\lambda}_j > c_1$. The modified Wald statistic (denoted by W^+), will benefit not only from $\hat{I}^+$, but also from $\hat{\Sigma}_p^+ = [\hat{P}(\hat{J}'\hat{I}^{-1}\hat{J})^{-1}\hat{P}']^+ = \hat{V}\hat{\Lambda}_{c_2}^+\hat{V}'$. Similarly, the modified score statistic denoted by S^+, will benefit from $\hat{I}^+$, but also from $\hat{\Sigma}_S^+$. Likewise, the statistics modified by the second-type regularization based on the inverse of $\hat{\Sigma}^-$ defined in equation (3.4.25) will be referred to as W^-, LR^- and S^-. Finally, the modified statistics built on the inverse of $\tilde{\Sigma}$ will be referred to as $\tilde{W}$, $\tilde{LR}$ and $\tilde{S}$ in the simulation experiments.

3.6.1. Size analysis

First of all, we can see in Table 3.1 that the frequency at which the non-regularized Wald statistic becomes non computable is around 10% in small sample and it diminishes when the sample size increases. As for the score statistic, the frequency at which it fails varies between 4% and 10%. Concerning the non-regularized LR statistic, the frequency at which it fails is around 3%. The rejection frequencies for the non-regularized procedures have been computed after excluding the cases when the usual inverses crash. We can see in Table 3.1 that the regularization techniques work extremely well for the LM statistic which tends to systematically reject the null hypothesis when performing the test with the non-regularized LM statistic. But once they are regularized, the LM statistics are controlled for the size fixed at $\alpha = 5\%$. On the other hand, the size distortions displayed by the non-regularized Wald statistic is not severe and varies between 7 and 9% but do not diminish when the sample size increases. However, the attempts of regularization of the covariance matrices performed at two levels, at the estimation step when regularizing the indirect criterion and at the testing step when regularizing the covariance matrix in the Wald statistic seem to help the latter. We can see that regularizing the covariance matrices by modifying the inverses prevents the statistic from

breaking down but also help in reducing the large standard errors. The two regularization techniques we propose better control for the size of the Wald test when compared with that of LB for all sample sizes. Indeed, the Wald statistic regularized with the technique proposed by LB still slightly overrejects in small samples and more than the other statistics. However, when the sample size increases, its overrejection is getting less severe. In such situations, simulation-based inference techniques such as Monte Carlo tests [see Dufour (2002)] help controlling for the size especially for the modified Wald statistic W^+ in small samples. At the opposite, the non-regularized LR statistic tends to underreject. For $T = 1000$ for instance, the non-regularized LR statistic never rejects the null hypothesis. The results support earlier works that both finite sample and asymptotic distributions of the LR test may also be modified when identifiability conditions are not satisfied [see Sargan (1983), Phillips (1989), Staiger and Stock (1997), and other references in Dufour (1997)]. However, Dufour (1997) shows that LR statistics have null distributions which can be bounded by a nuisance-parameter-free distribution (possibly derived from the Wilks Λ distribution), hence inference methods based on such statistics are more reliable. Further, the LR statistic is known to be robust to non-invariance problems unlike the Wald statistics [see Breusch and Schmidt (1988), Nelson and Savin (1990), Dagenais and Dufour (1991)]. Concerning the regularized LR statistics, the size performances are quite similar and help correcting for the underrejection for LR^+ and $\tilde{LR}$ but LR^- still suffers from underrejecting. In consequence, one can expect LR^- to loose power under the alternative. . Moreover, we observe for the LR tests (LR^+, $\tilde{LR}$) at $T = 500$ and $T = 2000$ in Table 3.1 that when MC tests whose distribution also depends on strong regularity conditions [see Dufour (2002)], cannot achieve in correcting for some over-rejections, the *maximized* MC test usually solves over-size problems.

3.6.2. Power analysis

We also study in Table 3.2 the power properties of the tests for an alternative hypothesis of stochastic volatility with a quite high persistence feature in the volatility process,

namely $H_1 : a = 0.9, r_w = 0.9$. The asymptotic tests suffering from oversize problems, have been corrected for size distortions. Monte Carlo tests are implemented with $N = 99$ simulated statistics since for power considerations, the number of simulated statistics may have an impact on gains in power. As expected, the Wald test is not consistent at all. When the sample size increases, the gains in power for the Wald statistic for the three procedures (Asy, MC, MMC), are not significant and are even diminishing for $\tilde{W}$. This observation carries out a crucial message concerning the behavior of the Wald statistic in a context of (almost) unidentified parameters. It is impossible to build a valid test based on the Wald statistic despite the various technical tools in hand, such as regularization techniques which may also contributes in correcting for "poor" standard errors. The Wald statistic is not *reformable* in situations close to non-identification [see Dufour (1997)]. Indeed, Dufour (1997) shows that the distribution of the Wald test cannot be bounded by any finite set of distribution functions under nonregular conditions. Under $H_0 : a = 0, r_w = 0$, the moment conditions defining the auxiliary estimator become nonlinearly redundant arising some singularity issue. On the other hand, we can note the very erratic behavior of the non-regularized LR statistic which support the fact that the regularization techniques help increasing the power performances significantly. This observation is particularly outstanding in large samples for $T = 1000, 2000$ where for instance $P[LR > \chi^2_{1-\alpha}(1)|H_1] = 0.17$ compared with $P[LR^+ > \chi^2_{1-\alpha}(1)|H_1] = 0.57$, $P[\tilde{LR} > \chi^2_{1-\alpha}(1)|H_1] = 0.50$ or $P[LR^- > \chi^2_{1-\alpha}(1)|H_1] = 0.44$. We further observe that LR^+ outperform the other statistics in small samples but in large samples, especially for $T = 2000$, $\tilde{LR}$ demonstrates equivalent power which reaches 68%. As expected, LR^- underperforms in term of power LR^+ and $\tilde{LR}$ for all sample sizes. We further observe in Table 3.2, a loss in power for both versions of Monte Carlo tests w.r.t. their asymptotic counterparts. Indeed, there is always a loss of "power" of the simulated tests compared with the asymptotic ones due to the noise introduced by the simulations. In this respect, one has to be aware that the asymptotic tests remain infeasible and are considered as a benchmark useful for comparisons purposes. Indeed, implementing the asymptotic

tests requires the prior knowledge of the nuisance parameters which is not available in practice. By contrast, *maximized* Monte Carlo tests provide *provably* exact tests irrespective of presence of nuisance parameters in the distribution of the test statistic. The only requirement of the procedure is that the test statistic can be simulated. On the other hand, once the score-type statistic has been corrected for overrejecting, the non-regularized statistic fails to reject the null hypothesis and gives a strong evidence on the fact that its asymptotic distribution is clearly modified in presence of singularity issue. All regularization technique indistinctively improve power significantly in such a situation. However, we observe that S^{+} less power than $\tilde{S}$ and S^{-}. On the other hand, as predicted by constructing $\tilde{\Sigma}$ which is always less than or equal to Σ^{-}, the modified score statistic $\tilde{S}$ based on the former demonstrates more power than S^{-}, namely $P[\tilde{S} > \chi^2_{1-\alpha}(1)|H_1] = 0.73$ whereas $P[S^{-} > \chi^2_{1-\alpha}(1)|H_1] = 0.68$. Clearly, the third-type regularization technique outperform in term of power the other ones for the score statistic whereas the LB technique built on the Moore-Penrose inverse seems to work better for the LR statistic in presence of weak identification.

3.7. Empirical application

In this section we test the null hypothesis of homoskedasticity in the volatility process from real data on the Standard and Poor's Composite Price Index (SP), 1928-87.

3.7.1. Data

The data have been provided by Georges Tauchen where Efficient Method of Moments have been used by Gallant, Hsieh and Tauchen (1997) to fit a standard stochastic volatility model. The data to which we fit the univariate stochastic volatility model is a long time series comprised of 16,127 daily observations, $\{\tilde{y}_t\}_{t=1}^{16,127}$, on adjusted movements of the Standard and poor's Composite Price Index, 1928-87. The raw series is the Standard and Poor's Composite Price Index (SP),daily, 1928-87. The raw series is converted to a price movements series, $100[\log(SP_t) - \log(SP_{t-1})]$, and then adjusted for

systematic calendar effects, that is, systematic shifts in location and scale due to different trading patterns across days of the week, holidays, and year-end tax trading. This yields a variable we shall denote y_t.

3.7.2. Results

To conduct the asymptotic tests, we use the asymptotic critical value of a $\chi^2_{1-\alpha}(1) = 3.84$ for a $\alpha = 5\%$ significance level. In Table 3.3, we observe that W^+ and $\tilde{W}$ reject the null hypothesis $H_0 : \ a = 0, \ \ r_w = 0$ of homoskedasticity in the volatility process whereas the other ones, which are W^- and W do not reject the null hypothesis. The same observation holds for simulated tests where this time W and W^+ cannot reject H_0 at both level whereas Monte Carlo tests based on $\tilde{W}$ and W^- statistics do reject H_0 at $\alpha = 5\%$ and $\alpha = 1\%$. Once again, these controversial results obtained with the Wald statistic highlight the unreliable feature of the latter when making inference under non regular conditions. As predicted by Dufour (1997), whatever powerful tools in hand, the Wald statistic is not reformable. Such a statistic cannot produce valid inference in nonstandard situations. By contrast, the LR statistic still provides reliable inference under nonregular conditions, even though its finite and asymptotic distribution may be modified. Our results reported in Table 3.3 for the LR statistics give evidence on this statement. The LR statistic did not need to be regularized since its estimated eigenvalues were greater than the thresholds, $c = 0.1$ and $c = 0.01$. Based on the LR statistic, asymptotic and simulated tests do reject H_0 at $\alpha = 5\%$ and $\alpha = 1\%$. Concerning the score statistic, the non-regularized statistic is not computable due to its covariance matrix Σ_S which is singular and not invertible. Therefore, we need to resort to its g-inverted covariance matrices and implement S^+, S^- and $\tilde{S}$. The results obtained with real data support those obtained with artificial data since the asymptotic test based on $\tilde{S}$ is the most powerful one over S^- and far beyond S^+ which cannot reject the null hypothesis. However, although the asymptotic test built on S^- fails to reject H_0 at the specified levels, the simulated tests do achieve in rejecting the null. Thus, simulations may provide more accurate critical points compared with asymptotic approximation

and produce more reliable inferences as observed here with Monte Carlo tests. In sum, we can formulate three types of recommendations: the practitioner should use the LR statistic which is the more robust test statistic to any data generating process, avoid the Wald statistic which tends to provide unreliable inference under nonregular conditions; finally if computing the restricted estimate of the parameter of interest is easier for the practitioner for the kind of null hypothesis under investigation, then he should use the score-type statistic modified according to the third regularization technique ($\tilde{S}$) to maximize power when conducting inference with singular covariance matrices.

Finally, based on these results, we can infer that the null hypothesis of homoskedasticity in the volatility observed on the Standard and Poor's Composite Price Index (SP),daily, 1928-87 can be rejected at both level of significance. However, although it is well-known that high-frequency financial data are time-varying and displays strong volatility clustering effects [see Engle (1982)], it is not clear that such a rejection may be attributed to volatility persistence effects but to tail thickness. In this vain, some researchers [see Chernov, Gallant, Ghysels and Tauchen (2003)] try to incorporate this aspect of asset returns distribution (*tail thickness*) by extending the single SV model by adding additional SV factors, thus breaking the link between tail thickness and volatility persistence.

3.8. Concluding remarks

To summarize, we provide regularization techniques of covariance matrices when these ones become singular and non invertible under nonregular conditions by resorting to some specific generalized inverses. However, although the regularization techniques help in keeping the test statistics computable under nonregular conditions, they do not ensure a χ^2 distribution for the modified statistics anymore. As a result, the distributional results developed by GMR (1993) become useless under nonregular conditions. One way to overcome this difficulty and still provide valid critical points and p-values, is to resort on simulation-based inference techniques such as Monte Carlo tests whose *maximized* version achieves in controlling for size distortions irrespective of nuisance

parameters in the distribution of the test statistic. The modified tests further demonstrate more power than their nonregularized counterparts. However, despite the attempts to regularize the covariance matrix of the Wald statistic, it still provides invalid inference in nonstandard problems. Indeed, the distribution of the Wald statistic cannot be bounded by any finite set of distribution functions under nonregular conditions. In such situations, *maximized* Monte Carlo tests can control for the size but at the cost of no power at all under the alternative. By contrast, the likelihood ratio test behave much better (both in size and power) in such situations even though its finite and asymptotic distributions may be modified. Concerning the performance of the score statistic under singularity issues, the nonregularized statistic behave very poorly but once regularized, especially according to the third technique based on $\tilde{\Sigma}$, it provides a powerful test statistic [see Hansen (1996)]. Finally, it is worth noting that the regularization techniques implemented here in the context of a stochastic volatility model estimated by indirect inference is not restricted to this particular framework but could be employed in more general models to handle singular weighting matrices as encountered for instance in GMM contexts or also in nonlinear models [see Gallant (1977), Gallant and Tauchen (1989)].

Table 3.1. Size

LEVEL in % ($H_0 : a = 0, r_w = 0$)								
$c = 0.95, r_y = 0.4$								
	T=200				*T=500*			
	Asy NON reg.	**Asy**	**MC**	**MMC**	**Asy NON reg.**	**Asy**	**MC**	**MMC**
failure	10	-	-	-	11	-	-	-
W	8.8	-	-	-	7.9	-	-	-
W^+	-	11	6	3	-	11	6	1
W^-	-	3	2	1	-	1	1	0
$\tilde{W}$	-	3	3	3	-	5	7	4
failure	3	-	-	-	1	-	-	-
LR	1	-	-	-	4	-	-	-
LR^+	-	2	5	2	-	4	7	1
LR^-	-	1	1	1	-	1	1	0
$\tilde{LR}$	-	2	3	2	-	5	6	4
failure	4	-	-	-	5	-	-	-
S	100	-	-	-	100	-	-	-
S^+	-	3	3	1	-	4	3	2
S^-	-	2	2	1	-	1	2	1
$\tilde{S}$	-	2	5	1	-	1	1	1
	T=1000				*T=2000*			
	Asy NON reg.	**Asy**	**MC**	**MMC**	**Asy NON reg.**	**Asy**	**MC**	**MMC**
failure	6	-	-	-	4	-	-	-
W	7.4	-	-	-	7.9	-	-	-
W^+	-	8	2	1	-	7	2	1
W^-	-	1	2	1	-	3	1	0
$\tilde{W}$	-	6	2	2	-	6	5	4
failure	3	-	-	-	0	-	-	-
LR	0	-	-	-	2	-	-	-
LR^+	-	2	1	1	-	4	6	1
LR^-	-	1	1	0	-	0	1	0
$\tilde{LR}$	-	1	1	1	-	4	7	4
failure	9	-	-	-	10	-	-	-
S	95.6	-	-	-	100	-	-	-
S^+	-	3	3	1	-	3	2	2
S^-	-	1	1	1	-	1	1	1
$\tilde{S}$	-	2	2	1	-	5	2	1

Table 3.2. Power

Power in % ($H_1 : a = 0.9, r_w = 0.9$)								
$c = 0.95, r_y = 0.4$								
	T=200				*T=500*			
	Asy NON reg.	**Asy**	**MC**	**MMC**	**Asy NON reg.**	**Asy**	**MC**	**MMC**
W	10	-	-	-	1	-	-	-
W^+	-	9	8	5	-	9	6	3
W^-	-	8	7	1	-	8	4	2
$\tilde{W}$	-	7	8	3	-	4	5	4
LR	44	-	-	-	39	-	-	-
LR^+	-	45	26	20	-	48	28	21
LR^-	-	30	26	16	-	33	31	19
$\tilde{LR}$	-	32	25	20	-	38	30	28
S	0	-	-	-	0	-	-	-
S^+	-	39	18	13	-	42	20	15
S^-	-	44	24	13	-	50	41	28
$\tilde{S}$	-	49	46	22	-	59	43	33
	T=1000				*T=2000*			
	Asy NON reg.	**Asy**	**MC**	**MMC**	**Asy NON reg.**	**Asy**	**MC**	**MMC**
W	0	-	-	-	0	-	-	-
W^+	-	6	5	2	-	5	4	1
W^-	-	2	6	3	-	5	5	3
$\tilde{W}$	-	0	2	2	-	0	1	0
LR	17	-	-	-	42	-	-	-
LR^+	-	57	51	46	-	69	63	51
LR^-	-	44	35	22	-	49	53	39
$\tilde{LR}$	-	50	49	40	-	68	61	50
S	0	-	-	-	0	-	-	-
S^+	-	45	30	18	-	51	38	31
S^-	-	63	58	41	-	68	65	59
$\tilde{S}$	-	69	68	51	-	73	67	62

Table 3.3. Empirical application

Standard and Poor's Composite Price index			
$H_0: \ a = 0 \ \ r_w = 0$			
	Asymptotic tests	*Monte Carlo tests*	
	S_0	N=19	N=99
W	0.000772	0.249	0.23
$W+$	6.70	0.30	0.30
$W-$	3.46	0.05	0.01
$\tilde{W}$	5.20	0.05	0.01
LR	111.91	0.05	0.01
LR^+	111.91	0.05	0.01
LR^-	111.91	0.05	0.01
$\tilde{LR}$	111.91	0.05	0.01
S	failure	-	-
S^+	0.0015	0.05	0.01
S^-	6.67	0.05	0.01
$\tilde{S}$	13.35	0.05	0.01

Part II

Nonlinear Canonical Analysis

Chapitre 4

Diffusion Processes with Polynomial Eigenfunctions [1]

[1] This paper is co-authored with Christian Gouriéroux and Eric Renault.

4.1. Introduction

One dimensional stochastic differential equations (s.d.e), such as:

$$dy_t = \mu(y_t)dt + \sigma(y_t)dW_t \ , \tag{4.1.1}$$

where (W_t) is a brownian motion, μ and σ the drift and volatility functions are basic specifications for describing the evolution of financial returns [see e.g. Black and Scholes (1973)] interest rates [see e.g. Vasicek (1977), Cox, Ingersoll, and Ross (1985b), Cox, Ingersoll, and Ross (1985a)], or macroeconomic series [see e.g. Chen and Epstein (1999), Anderson, Hansen, and Sargent (2003), Cagetti, Hansen, Sargent, and Williams (2002)] in continuous time. A recent literature points out the importance of the spectral analysis of the associated infinitesimal generator for the analysis of a s.d.e. On the one hand, the knowledge of the spectral decomposition simplifies the computation of nonlinear predictions at any horizon. This feature is used for instance to determine the pattern of the term structure of interest rates when the short term interest rate follows an equation like (4.1.1) [see e.g. Pagan, Hall, and Martin (1996)]. On the second hand the spectral analysis underlies nonparametric estimation methods of the drift and volatility functions. The basic idea is to estimate the infinitesimal generator either by kernel approach [see e.g. Darolles, Florens, and Gouriéroux (2000)] or by projecting on a basis of polynomials [see e.g. Darolles, Florens, and Renault (1997), Hansen, Scheinkman, and Touzi (1998), Chen, Hansen, and Scheinkman (1998), Florens, Renault, and Touzi (1998), Darolles and Gouriéroux (2001)], to perform the spectral decomposition of this estimated generator, and then to deduce from the first and second eigenfunctions the drift and volatility functions [see Demoura (1993)].

The aim of this paper is to fully characterize the one-dimensional stochastic differential equations, for which the eigenfunctions of the infinitesimal generator are polynomials in y. For these s.d.e., it can be expected that the estimation method by projection will be accurate, even in finite sample.

The characterization of the diffusion processes with polynomial eigenfunctions is

given in section 2. We also provide in this section the eigenvalues, the expressions of the eigenfunctions, the stationarity conditions and the density of the marginal distribution of the processes. The proofs of the main results are gathered in section 3.

4.2. Characterization

Let us consider a one dimensional stationary diffusion process:

$$dy_t = \mu(y_t)dt + \sigma(y_t)dW_t \ , \tag{4.2.2}$$

with drift and volatility functions denoted by μ and σ, respectively. Its infinitesimal generator A is defined by:

$$A\psi(y) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} E[\psi(Y_{t+\Delta t}) - \psi(Y_t)|Y_t = y] \ , \tag{4.2.3}$$

and explains how to compute the infinitesimal drift of the transformed series $(\psi(Y_t))$. By applying Ito's lemma, it is easily seen that the generator A corresponds to the differential operator for C^2 functions ψ:

$$A\psi(y) = \mu(y)\frac{\partial \psi}{\partial y}(y) + \frac{1}{2}\sigma(y)^2\frac{\partial^2 \psi}{\partial y^2}(y) \ . \tag{4.2.4}$$

It is known that this operator is generally self-adjoint, and in particular admits a spectral decomposition with real eigenvalues [see e.g. Hansen and Scheinkman (1995)]. We assume [see Florens, Renault, and Touzi (1998) for a discussion] that :

Assumption 4.2.1 COMPACTNESS OF THE INFINITESIMAL OPERATOR. *A is a compact operator, with distinct negative eigenvalues λ_n, $n \in N$, say, and eigenfunctions ψ_n, $n \in N$.*

Proposition 4.2.2 CHARACTERIZATION PROPERTY. *Under Assumption* **4.2.1** *the diffusion process admits polynomial eigenfunctions ψ_n with increasing degree n if and*

only if one of the following conditions is satisfied:

i) $\mu(y) = b(y - \beta), \quad \sigma^2(y) = c_0$, *where* $b < 0$, *and* y *is defined on* R;

ii) $\mu(y) = b(y - \beta), \quad \sigma^2(y) = c_1 y + c_0$, *where* $b < 0$, *and* y *is defined on the semi-interval* $[-c_0/c_1, +\infty[$, *if* $c_1 > 0$, *or on the semi-interval* $]-\infty, -c_0/c_1]$, *if* $c_1 < 0$.

iii) $\mu(y) = b(y - \beta), \quad \sigma^2(y) = c(y - \gamma_1)(y - \gamma_2)$, *where* $b < 0$, $c < 0$, $\gamma_1 < \beta < \gamma_2$, *and* y *is defined in the interval* (γ_1, γ_2).

In any case the eigenvalues are: $\lambda_n = bn + \frac{1}{2}cn(n-1), \quad n \geq 1$ *where* $c = 0$ *for cases* $i)$ *and* $ii)$.

Thus we get three types of processes which can be distinguished by the restrictions on the domain of admissible values. They are affine transformations of the Ornstein-Uhlenbeck process, the Cox-Ingersoll-Ross process and the Jacobi process, respectively. The Ornstein-Uhlenbeck process, or mean-reverting process [see the negativity condition imposed on parameter b in i)] underlies the Vasicek model [see Vasicek (1977)]. The Cox-Ingersoll-Ross process when $c_0 = 0$ in ii), and more generally the square root processes are used for describing the evolution of interest rates [see Pagan, Hall, and Martin (1996)], or for defining time deformation [see Conley, Hansen, Luttmer, and Scheinkman (1997), Carrasco, Hansen, and Chen (1999), Ghysels, Gouriéroux, and Jasiak (1995), Ghysels, Gouriéroux, and Jasiak (1998)]. Finally the Jacobi process is appropriate for the evolution of a probability or a default rate, which are between 0 and 1 [see Nielsen, Saa Requeja, and Santa Clara (1993), Lando (1998), Cagetti, Hansen, Sargent, and Williams (2002)].

The corollary below provides different properties of these processes concerning the stationary distribution and the expressions of the eigenfunctions. The stationary distribution belongs to the Pearson family, that is their density f satisfy $\frac{d \log f(y)}{dy} = \frac{ay+b}{cy^2+dy+c}$, say. Thus, the class of diffusion processes with polynomial eigenfunctions coincides with the class of stationary markov processes with marginal distribution in the Pearson family [see Wong (1964), Wong and Thomas (1962)]. Then, the expressions of

the eigenfunctions are deduced from standard results on orthogonal polynomials [see e.g. Abramowitz and Stegun (1965)]. It is important to note that the eigenfunctions are not uniquely defined. The eigenfunctions given below ψ_n, say, are standardized with respect to the marginal distribution of the process, that is they satisfy $\int \psi_n(y)\psi_m(y)f(y)\,dy = 0$, if $n \neq m$, $= 1$, if $n = m$, where f is the p.d.f of the marginal distribution.

Corollary 4.2.3 .

i) The eigenfunctions of the Ornstein-Uhlenbeck process solution of the s.d.e.

$$dy_t = b(y_t - \beta)dt + \sqrt{c_0}dW_t \ ,$$

are the Hermite polynomials given by:

$$\tilde{H}e_n(y_t) = (n!)^{1/2} \sum_{m=0}^{[\frac{n}{2}]} (-1)^m \frac{1}{m!2^m(n-2m)!} \left(\frac{-2b}{c_0}\right)^{\frac{n-2m}{2}} (y_t - \beta)^{n-2m} \ . \tag{4.2.5}$$

They are standardized with respect to the marginal gaussian distribution $N(\beta, -\frac{c_0}{2b})$ *of* (y_t).

ii) The eigenfunctions of the square root process (y_t)*, solution of the s.d.e.:*

$$dy_t = b(y_t - \beta)dt + \sqrt{c_1 y_t + c_0}dW_t \ ,$$

are the Generalized Laguerre polynomials:

$$\tilde{L}_n^{(\alpha)}(z_t) = \sum_{j=0}^{n} (-1)^j \frac{\Gamma(\alpha+1+n)^{1/2}}{\Gamma(\alpha+1+j)} \frac{[n!\Gamma(\alpha+1)]^{1/2}}{j!(n-j)!} z_t^j \ , \tag{4.2.6}$$

where $\alpha = -\frac{2b}{c_1^2}(c_1\beta + c_0) - 1$, $z_t = -\frac{2b}{c_1^2}(c_1 y_t + c_0)$. *The polynomials are standardized with respect to the marginal distribution of* (y_t)*, which corresponds to the gamma distribution* $\gamma[-\frac{2b}{c_1^2}(c_1\beta + c_0), -\frac{2b}{c_1}]$ *shifted from* $-\frac{c_0}{c_1}$. *The special case* $c_0 = 0$ *yields the Cox-Ingersoll-Ross process.*

iii) The eigenfunctions of the Jacobi process (y_t), solution of the following s.d.e.:

$$dy_t = b(y_t - \beta)dt + \sqrt{c(y_t - \gamma_1)(y_t - \gamma_2)}dW_t \ ,$$

are the Jacobi polynomials given by:

$$\begin{aligned} \tilde{P}_n^{(\tilde{\alpha},\tilde{\beta})}(y_t) &= \left[\frac{\Gamma(\tilde{\alpha}+n+1)(2n+\tilde{\alpha}+\tilde{\beta}+1)\Gamma(\tilde{\alpha}+1)\Gamma(\tilde{\beta}+1)}{n!\Gamma(\tilde{\alpha}+\tilde{\beta}+n+1)\Gamma(\tilde{\alpha}+\tilde{\beta}+2)\Gamma(\tilde{\beta}+n+1)}\right]^{1/2} \\ &\quad \sum_{m=0}^{n}\binom{n}{m}\frac{\Gamma(\tilde{\alpha}+\tilde{\beta}+n+m+1)}{\Gamma(\tilde{\alpha}+m+1)}\frac{(y_t-\gamma_2)^m}{(\gamma_2-\gamma_1)^m} \ , \end{aligned} \tag{4.2.7}$$

with $\tilde{\alpha} = \frac{2b}{c}\frac{\gamma_2-\beta}{\gamma_2-\gamma_1} - 1$ and $\tilde{\beta} = \frac{2b}{c}\frac{\beta-\gamma_1}{\gamma_2-\gamma_1} - 1$. They are standardized with respect to the marginal distribution of (y_t), which corresponds to an affine transformation of the Beta distribution $B(\tilde{\beta}+1, \tilde{\alpha}+1)$.

4.3. Proof of the properties

The proof involves five steps. We first establish the necessary patterns of the drift and volatility functions, then the necessary expressions of the eigenvalues. In the third step, we discuss the constraints to be introduced on the parameters to ensure a nonnegative volatility and a stationary solution. In the fourth step, we establish the marginal distributions of the processes. Finally, we determine the standardized polynomial eigenfunctions.

4.3.1. The pattern of the drift and volatility functions

Assumption 4.3.1 *If the eigenfunctions are polynomials, the drift is a polynomial of degree one:*

$$\mu(y) = b(y - \beta), \quad \text{with } b < 0,$$

whereas the volatility is a polynomial of degree at most 2:

$$\sigma^2(y) = cy^2 + c_1 y + c_0 \ .$$

Proof: It is known from (a generalized version of) Sturm-Liouville theory, that the eigenfunctions satisfy the following shape restrictions [see Chen, Hansen, and Scheinkman (1998)]:

- ψ_j crosses the zero axis precisely j times;
- ψ_j' has precisely $j-1$ interior zeros (same sign between any two consecutive zeros).

In accordance with these shape restrictions, the first eigenfunction crosses the zero axis once, the second one twice, and so forth... As a consequence, the first two eigenfunctions are of the form:

$$\psi_1(y) = y + a_{10}, \quad \psi_2(y) = y^2 + a_{21}y + a_{20}, \text{ say.}$$

They satisfy the condition:

$$A\psi_n(y) = \lambda_n \psi_n(y), \quad n = 1,2 \quad \text{with } \lambda_n < 0$$

$$\Leftrightarrow \mu(y)\frac{\partial \psi_n}{\partial y}(y) + \frac{1}{2}\sigma^2(y)\frac{\partial^2 \psi_n}{\partial y^2}(y) = \lambda_n \psi_n(y) \quad n = 1,2$$

$$\Leftrightarrow \begin{cases} \mu(y) = \lambda_1(y + a_{10}) \\ \mu(y)(2y + a_{21}) + \sigma^2(y) = \lambda_2(y^2 + a_{21}y + a_{20}). \end{cases}$$

By solving this system we deduce the result of Lemma 4.3.1. Q.E.D

4.3.2. Expression of the eigenvalues

Lemma 4.3.1 *If the eigenfunctions are polynomials, the eigenvalues are:*

$\lambda_n = nb + \frac{1}{2}cn(n-1)$, *where* $c \leq 0$, *and* $b < 0$, *if* $c = 0$.

Proof: After replacing μ and σ^2 by their expressions, the condition $A\psi_n(y) = \lambda_n \psi_n(y)$ becomes:

$$b(y - \beta)\frac{\partial \psi_n}{\partial y}(y) + \frac{1}{2}(cy^2 + c_1 y + c_0)\frac{\partial^2 \psi_n}{\partial y^2}(y) = \lambda_n \psi_n(y).$$

When $\psi_n(y) = y^n + a_{n,n-1}y^{n-1} + ... + a_{n0}$ is a polynomial, we get by identifying the coefficients of the terms of degree n:

$$nb + \frac{1}{2}cn(n-1) = \lambda_n.$$

For large n, λ_n is equivalent to either $\frac{1}{2}cn(n-1)$, if $c \neq 0$, or nb, if $c = 0$. We deduce the constraints on parameters b and c to ensure that λ_n is negative. Q.E.D

4.3.3. The constraints on the parameters

case i) : Constant volatility.

The volatility is $\sigma^2(y) = c_0 > 0$ and the eigenvalues are $\lambda_n = nb$, with $b < 0$. These constraints are sufficient to characterize affine transformations of the Ornstein-Uhlenbeck process.

case ii) : Affine volatility.

The volatility is $\sigma^2(y) = c_1 y + c_0$, with $c_1 \neq 0$ and the eigenvalues are $\lambda_n = nb$, with $b < 0$. The positivity of the volatility is ensured if the domain of admissible values of y is restricted:

$$y \in]-c_0/c_1, +\infty[, \ if \ c_1 > 0 \quad , y \in]-\infty, -c_0/c_1[, \ if \ c_1 < 0 \ .$$

These constraints are sufficient to characterize affine transformations of the square root process.

case iii) : Quadratic volatility.

The volatility is $\sigma^2(y) = cy^2 + c_1 y + c_0$ and the eigenvalues are $\lambda_n = nb + \frac{1}{2}cn(n-1)$ with $c < 0$. Since $c < 0$, the volatility function can take positive values if and only if the polynomial $\sigma^2(y) = cy^2 + c_1 y + c_0$ has two distinct real roots $\gamma_1 < \gamma_2$.

Lemma 4.3.2 *β is between the roots γ_1 and γ_2.*

Proof: The strict positivity of the volatility $\sigma^2(y) = cy^2 + c_1 y + c_0$ implies that $y \in (\gamma_1, \gamma_2)$, which implies that $E(y) = \beta$ belongs to the $(\gamma_1, \gamma_2) = (0,1)$ interval. Q.E.D

Then it can be checked that the process is well defined, stationary, with range (γ_1, γ_2).

4.3.4. Stationary distributions

It is known [see Hansen, Scheinkman, and Touzi (1998)] that the density function of the stationary distribution of a diffusion process is proportional to:

$$\frac{1}{\sigma^2(x)} \exp\left[2\int_a^x \frac{\mu(y)}{\sigma^2(y)}\, dy\right], \tag{4.3.8}$$

where a is an arbitrary interior point of the state space.

case i) : Ornstein-Uhlenbeck process.

The drift and volatility functions are $\mu(y) = b(y-\beta)$ and $\sigma^2(y) = c_0$ respectively, which yields that the p.d.f of the stationary distribution is proportional to:

$$\frac{1}{c_0} \exp[\frac{b}{c_0}(y-\beta)^2]\,. \tag{4.3.9}$$

Therefore, we recognize a $N(\beta, -\frac{c_0}{2b})$ distribution for the Ornstein-Uhlenbeck process and

$$f(y) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{-\frac{c_0}{2b}}} \exp\left[-\frac{1}{2}\frac{(y-\beta)^2}{-\frac{c_0}{2b}}\right]. \tag{4.3.10}$$

case ii) : Square root process.

For the square root process, we have $\mu(y) = b(y-\beta)$ and $\sigma^2(y) = c_1 y + c_0$ which yields that the p.d.f of the stationary distribution is proportional to:

$$\frac{1}{c_1}(x + \frac{c_0}{c_1})^{-\frac{2b}{c_1}(\beta + \frac{c_0}{c_1})-1} \exp(\frac{2b}{c_1}x)\,. \tag{4.3.11}$$

Let us consider the case $c_1 > 0$, which corresponds to the domain $[-\frac{c_0}{c_1}, +\infty]$ of

the square root process. Thus the p.d.f of the stationary distribution corresponds to a gamma distribution with drift. We get:

$$f(y_t) = \frac{(-\frac{2b}{c_1})^{-\frac{2b}{c_1}(\beta+\frac{c_0}{c_1})}}{\Gamma(-\frac{2b}{c_1}(\beta+\frac{c_0}{c_1}))}\exp\left[-(-\frac{2b}{c_1})(y_t+\frac{c_0}{c_1})\right](y_t+\frac{c_0}{c_1})^{-\frac{2b}{c_1}(\beta+\frac{c_0}{c_1})-1} I_{(-\frac{c_0}{c_1},+\infty)}(y_t)\ . \tag{4.3.12}$$

Thus, $y_t = \theta_0 z_t + \theta_1$, where $\theta_0 = -\frac{c_1}{2b}$, $\theta_1 = -\frac{c_0}{c_1}$ and z_t follows the gamma distribution with parameter $-\frac{2b}{c_1}(\beta+\frac{c_0}{c_1})$.

case iii) : Jacobi process.

The drift and volatility functions are defined by $\mu(y) = b(y-\beta)$ and $\sigma^2(y) = c(y-\gamma_1)(y-\gamma_2)$ with $c<0$, which yields that the p.d.f of the stationary distribution is proportional to:

$$\frac{1}{-c}(y-\gamma_1)^{\frac{2b}{c}\frac{\beta-\gamma_1}{\gamma_2-\gamma_1}-1}(\gamma_2-y)^{\frac{2b}{c}\frac{\gamma_2-\beta}{\gamma_2-\gamma_1}-1}\ . \tag{4.3.13}$$

We deduce that the p.d.f

$$f(y) = \frac{(y-\gamma_1)^{\frac{2b}{c}\frac{\beta-\gamma_1}{\gamma_2-\gamma_1}-1}(\gamma_2-y)^{\frac{2b}{c}\frac{\gamma_2-\beta}{\gamma_2-\gamma_1}-1}}{(\gamma_2-\gamma_1)^{\frac{2b}{c}-1}B\left(\frac{2b}{c}\frac{\beta-\gamma_1}{\gamma_2-\gamma_1},\frac{2b}{c}\frac{\gamma_2-\beta}{\gamma_2-\gamma_1}\right)}1_{(\gamma_1,\gamma_2)}(y) \tag{4.3.14}$$

corresponds to a Beta distribution defined on $[\gamma_1,\gamma_2]$.

4.3.5. Polynomial eigenfunctions

It is easily checked that the differential equation

$$b(y-\beta)\frac{\partial\psi_n}{\partial y}(y) + \frac{1}{2}(cy^2+c_1y+c_0)\frac{\partial^2\psi_n}{\partial y^2}(y) = [nb+\frac{1}{2}cn(n-1)]\psi_n(y) \tag{4.3.15}$$

admits a polynomial solution of degree n. Therefore there is a basis of canonical eigenfunctions corresponding to polynomials of increasing degrees. Then we have just to

give the solutions for the three cases described in Proposition **4.2.2**.

case i) : Ornstein-Uhlenbeck process.

The differential equation (4.3.15) with $c = 0$ and $c_1 = 0$ is directly related to the Hermite equation:

$$\frac{\partial^2 \Phi}{z^2} + (-z)\frac{\partial \Phi}{z} + n\Phi(z) = 0$$

after an appropriate change of variable. More precisely, starting from the condition $A\psi_n(y) = \lambda_n \psi_n(y)$ with $\lambda_n = bn$, we get:

$$b(y-\beta)\frac{\partial \psi_n}{\partial y} + \frac{1}{2}c_0 \frac{\partial^2 \psi_n}{\partial y^2} = bn\psi_n(y) \; . \tag{4.3.16}$$

Considering an affine transform of the form: $y = \alpha z + \gamma$ such that: $\Phi_n(z) = \psi_n(\alpha z + \gamma)$ we can rewrite equation (4.3.16) as:

$$\frac{\sigma^2}{2\alpha^2}\Phi_n''(z) + \frac{b}{\alpha}(\alpha z + \gamma - \beta)\Phi_n'(z) - bn\Phi_n(z) = 0 \; . \tag{4.3.17}$$

Equating $\frac{\sigma^2}{2\alpha^2} = -b$, we get after a few manipulations:

$$\Phi_n''(z) + (-z + \frac{\beta - \gamma}{\alpha})\Phi_n'(z) + n\Phi_n(z) = 0 \tag{4.3.18}$$

which yields $\gamma = \beta$ [see Abramowitz and Stegun (1965), p.781 formula 22.6.21]. It is known [see Abramowitz and Stegun (1965), p.775 formula 22.3.11] that:

$$\begin{aligned} \Phi_n(z) &= He_n(z) \\ &= n! \sum_{m=0}^{[\frac{n}{2}]} (-1)^m \frac{1}{m! 2^m (n-2m)!} z^{n-2m} \; . \end{aligned} \tag{4.3.19}$$

Thus,the transformed variable $z_t = \sqrt{\frac{-2b}{c_0}}(y_t - \beta)$ satisfies

$$b(y-\beta)\frac{\partial \psi_n}{\partial y}(y) + \frac{1}{2}(c_0)\frac{\partial^2 \psi_n}{\partial y^2}(y) = nb\psi_n(y)$$

whose solutions are the Hermite polynomials:

$$He_n(y_t) = n! \sum_{m=0}^{[\frac{n}{2}]} (-1)^m \frac{1}{m! 2^m (n-2m)!} \left(\frac{-2b}{c_0}\right)^{\frac{n-2m}{2}} (y_t - \beta)^{n-2m} \ . \quad (4.3.20)$$

Given that $y_t \sim N(\beta, -\frac{c_0}{2b})$ [see paragraph 4.3.4], $z_t \sim N(0,1)$. We shall standardize the Hermite polynomials in the sequel. We can state [see Abramowitz and Stegun (1965), p.775 formula 22.2.15] that

$$\int_{-\infty}^{+\infty} \exp(-\frac{z_t^2}{2} He_n(z_t)^2 \, dz_t = \sqrt{2\pi} n!$$

or either,

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp(-\frac{z_t^2}{2} He_n(z_t)^2 \, dz_t = n! \, .$$

Hence,

$$\tilde{He}_n(z_t) = \frac{He_n(z_t)}{(n!)^{1/2}} \ ,$$

that is

$$\tilde{He}_n(z_t) = (n!)^{1/2} \sum_{m=0}^{[\frac{n}{2}]} (-1)^m \frac{1}{m! 2^m (n-2m)!} z_t^{n-2m} \ , \quad (4.3.21)$$

and therefore

$$\tilde{He}_n(y_t) = (n!)^{1/2} \sum_{m=0}^{[\frac{n}{2}]} (-1)^m \frac{1}{m! 2^m (n-2m)!} \left(\frac{-2b}{c_0}\right)^{\frac{n-2m}{2}} (y_t - \beta)^{n-2m} \ . \quad (4.3.22)$$

case ii) : Square root process.

The differential equation (4.3.15) with $c = 0$ corresponds to

$$b(y-\beta)\frac{\partial \psi_n}{\partial y}(y) + \frac{1}{2}(c_1 y + c_0)\frac{\partial^2 \psi_n}{\partial y^2}(y) = nb\psi_n(y) \ , \quad (4.3.23)$$

and is directly related to the Kummer's equation [see Abramowitz and Stegun

(1965), p.504 formula 13.1.1]:

$$z\frac{\partial^2\tilde{\psi}}{\partial z^2}(z)+\left(-\frac{2b}{c_1^2}(c_1\beta+c_0)-z\right)\frac{\partial\tilde{\psi}}{\partial z}(z)+n\tilde{\psi}(z)=0, \qquad (4.3.24)$$

through the following change of variable: $z_t=\frac{-2b}{c_1^2}(c_1y_t+c_0)$ with $\tilde{\psi}(\frac{-2b}{c_1^2}(c_1y_t+c_0))=\psi(y)$. The solution to (4.3.24) is a polynomial of order n in z corresponding to the Kummer's function up to order n, i.e.

$$M(a_K,b_K,z_t)=1+\frac{a_K}{b_K}z_t+\frac{(a_K)_2}{(b_K)_2}\frac{z_t^2}{2!}+...+\frac{(a_K)_n}{(b_K)_n}\frac{z_t^n}{n!},$$

where the coefficients are given by $a_K=-n$, $b_K=-\frac{2b}{c_1^2}(c_1\beta+c_0)$,$z_t=-\frac{2b}{c_1^2}(c_1y_t+c_0)$, and the following Pochammer's symbols $(a_K)_n=a_K(a_K+1)(a_K+2)...(a_K+n-1)$, $(a_K)_0=1$ [see Abramowitz and Stegun (1965), p.504 formulas 13.1.2 and 13.1.3]. It is known that the Kummer function $M(a_K,b_K,z)$ with the Kummer coefficients of the form $a_K=-n$ and $b_K=-\frac{2b}{c_1^2}(c_1\beta+c_0)=\alpha+1$ (set $\alpha=b_K-1$), corresponds to a Generalized Laguerre polynomial of the form [see Abramowitz and Stegun (1965), p.509 formula 13.6.9]:

$$\frac{n!}{(\alpha+1)_n}L_n^{(\alpha)}(z)\ ,$$

with $\alpha=-\frac{2b}{c_1^2}(c_1\beta+c_0)-1$. The Generalized Laguerre polynomials [see Abramowitz and Stegun (1965), p.775 formula 22.3.9] can be standardized as follows [see Abramowitz and Stegun (1965), p.775 formula 22.2.12].

$$\int_0^\infty \exp(-z_t)z_t^\alpha L_n^{(\alpha)}(z_t)^2\,dz_t=\frac{\Gamma(\alpha+1+n)}{n!}$$

Besides we know that

$$\int_0^\infty \exp(-z_t)z_t^\alpha\,dz_t=\Gamma(\alpha+1)\ ,$$

hence,

$$\frac{1}{\Gamma(\alpha+1)}\int_0^\infty \exp(-z_t) z_t^\alpha L_n^{(\alpha)}(z_t)^2\, dz_t = \frac{\Gamma(\alpha+1+n)}{n!\Gamma(\alpha+1)}\ ,$$

or either, that

$$\frac{1}{\Gamma(\alpha+1)}\int_0^\infty \exp(-z_t) z_t^\alpha \left[\frac{n!}{(\alpha+1)_n}\right]^2 L_n^{(\alpha)}(z_t)^2\, dz_t = \left[\frac{n!}{(\alpha+1)_n}\right]^2 \frac{\Gamma(\alpha+1+n)}{n!\Gamma(\alpha+1)}\ ,$$

Therefore, the standardized polynomials correspond to:

$$\tilde{L}_n^{(\alpha)}(z_t) = \frac{L_n^{(\alpha)}(z_t)}{\left(\frac{\Gamma(\alpha+1+n)}{n!\Gamma(\alpha+1)}\right)^{1/2}}$$

which yields with $z_t = \frac{-2b}{c_1^2}(c_1 y_t + c_0)$ and $\alpha = -\frac{2b}{c_1^2}(c_1\beta + c_0) - 1$:

$$\tilde{L}_n^{(\alpha)}(z_t) = \sum_{j=0}^{n}(-1)^j \frac{\Gamma(\alpha+1+n)^{1/2}}{\Gamma(\alpha+1+j)} \frac{[n!\Gamma(\alpha+1)]^{1/2}}{j!(n-j)!} z_t^j\ .$$

case iii) : Jacobi process.

After introducing the roots γ_1, γ_2, equation (4.3.15) can be written as:

$$[-y^2+(\gamma_1+\gamma_2)y-\gamma_1\gamma_2]\frac{\partial^2\psi}{\partial y^2}(y)-\frac{2b}{c}(y-\beta)\frac{\partial\psi}{\partial y}(y)+\frac{2}{c}[nb+\frac{1}{2}cn(n-1)]\psi(y)=0\ . \tag{4.3.25}$$

Without loss of generality, we can focus on the case $[\gamma_1, \gamma_2] = [-1, 1]$ where equation (4.3.25) becomes:

$$(1-y^2)\frac{\partial^2\psi}{\partial y^2}(y)+(\frac{2b}{c}\beta-\frac{2b}{c}y)\frac{\partial\psi}{\partial y}(y)+\frac{2}{c}[nb+\frac{1}{2}cn(n-1)]\psi(y)=0\ . \tag{4.3.26}$$

It is known [see Abramowitz and Stegun (1965), p.781 formula 22.6.1] that this differential equation admits as solutions the Jacobi polynomials [see Abramowitz

and Stegun (1965), p.775 formula 22.3.2]:

$$P_n^{(\tilde{\alpha}^*,\tilde{\beta}^*)}(y_t^*) = \frac{\Gamma(\tilde{\alpha}^*+n+1)}{n!\Gamma(\tilde{\alpha}^*+\tilde{\beta}^*+n+1)} \sum_{m=0}^{n} \binom{n}{m} \frac{\Gamma(\tilde{\alpha}^*+\tilde{\beta}^*+n+m+1)}{2^m\Gamma(\tilde{\alpha}^*+m+1)} (y_t^*-1)^m \tag{4.3.27}$$

with $\tilde{\alpha}^* = \frac{b^*}{c^*}(1-\beta^*) - 1$ and $\tilde{\beta}^* = \frac{b^*}{c^*}(\beta^*+1) - 1$. We know [see Abramowitz and Stegun (1965), p.773 formulas 22.1.1, 22.1.2, and 22.2.1] that

$$\int_{-1}^{1} (1-y_t^*)^{\tilde{\alpha}^*}(1+y_t^*)^{\tilde{\beta}^*} P_n(y_t^*)P_m(y_t^*)dy_t^* = 0\,, \quad n \neq m$$

and

$$\int_{-1}^{1} (1-y_t^*)^{\tilde{\alpha}^*}(1+y_t^*)^{\tilde{\beta}^*} P_n^2(y_t^*)dy_t^* = \frac{2^{\tilde{\alpha}^*+\tilde{\beta}^*+1}}{2n+\tilde{\alpha}^*+\tilde{\beta}^*+1} \frac{\Gamma(\tilde{\alpha}^*+n+1)\Gamma(\tilde{\beta}^*+n+1)}{n!\Gamma(\tilde{\alpha}^*+\tilde{\beta}^*+n+1)}\,. \tag{4.3.28}$$

But we need to standardize the weight function:

$$\begin{aligned} \int_{-1}^{1} (1-y_t^*)^{\tilde{\alpha}^*}(1+y_t^*)^{\tilde{\beta}^*} dy_t^* &= \int_0^1 (-2v_t+2)^{\tilde{\alpha}^*}(2v_t)^{\tilde{\beta}^*} 2dv_t \\ &= \int_0^1 2^{\tilde{\alpha}^*+\tilde{\beta}^*+1}(1-v_t)^{\tilde{\alpha}^*} v_t^{\tilde{\beta}^*} dv_t \\ &= 2^{\tilde{\alpha}^*+\tilde{\beta}^*+1} \frac{\Gamma(\tilde{\alpha}^*+1)\Gamma(\tilde{\beta}^*+1)}{\Gamma(\tilde{\alpha}^*+\tilde{\beta}^*+2)} \end{aligned} \tag{4.3.29}$$

using the transformation $y_t^* = 2v_t - 1$ in order to have a distribution function. Using this standardization, eq.4.3.28 becomes

$$\begin{aligned} &\frac{\Gamma(\tilde{\alpha}^*+\tilde{\beta}^*+2)}{2^{\tilde{\alpha}^*+\tilde{\beta}^*+1}\Gamma(\tilde{\alpha}^*+1)\Gamma(\tilde{\beta}^*+1)} \int_{-1}^{1} (1-y_t^*)^{\tilde{\alpha}^*}(1+y_t^*)^{\tilde{\beta}^*} P_n^2(y_t^*)dy_t^* \\ &= \frac{\Gamma(\tilde{\alpha}^*+\tilde{\beta}^*+2)\Gamma(\tilde{\alpha}^*+n+1)\Gamma(\tilde{\beta}^*+n+1)}{(2n+\tilde{\alpha}^*+\tilde{\beta}^*+1)\Gamma(\tilde{\alpha}^*+1)\Gamma(\tilde{\beta}^*+1)\Gamma(\tilde{\alpha}^*+\tilde{\beta}^*+n+1)n!} \end{aligned} \tag{4.3.30}$$

which yields the standardized Jacobi polynomials

$$\tilde{P}_n^{(\tilde{\alpha}^*,\tilde{\beta}^*)}(y_t^*) = \left[\frac{\Gamma(\tilde{\alpha}^*+n+1)(2n+\tilde{\alpha}^*+\tilde{\beta}^*+1)\Gamma(\tilde{\alpha}^*+1)\Gamma(\tilde{\beta}^*+1)}{n!\Gamma(\tilde{\alpha}^*+\tilde{\beta}^*+n+1)\Gamma(\tilde{\alpha}^*+\tilde{\beta}^*+2)\Gamma(\tilde{\beta}^*+n+1)}\right]^{1/2}$$

$$\sum_{m=0}^{n} \binom{n}{m} \frac{\Gamma(\tilde{\alpha}^* + \tilde{\beta}^* + n + m + 1)}{2^m \Gamma(\tilde{\alpha}^* + m + 1)} (y_t^* - 1)^m \,. \qquad (4.3.31)$$

Finally from the stochastic differential equation on $(-1, 1)$:

$$dy_t^* = b^*(y_t^* - \beta^*)dt + \sqrt{c^*(y_t^* + 1)(y_t^* - 1)}dW_t^*$$

we can deduce the solution of the stochastic differential equation on (γ_1, γ_2):

$$dy_t = b(y_t - \beta)dt + \sqrt{c(y_t - \gamma_1)(y_t - \gamma_2)}dW_t$$

by applying the affine transform $y_t = \frac{\gamma_2 - \gamma_1}{2} y_t^* + \frac{\gamma_2 + \gamma_1}{2}$. We have $b = b^*$, $c = c^*$, and $\beta = (\gamma_1 + \gamma_2)/2 + \beta^*(\gamma_2 - \gamma_1)/2$. The polynomial eigenfunctions of the general s.d.e. are obtained by applying the same affine transformations to the Jacobi polynomials (4.3.31). We get:

$$\begin{aligned} \tilde{P}_n^{(\tilde{\alpha},\tilde{\beta})}(y_t) &= \left[\frac{\Gamma(\tilde{\alpha} + n + 1)(2n + \tilde{\alpha} + \tilde{\beta} + 1)\Gamma(\tilde{\alpha} + 1)\Gamma(\tilde{\beta} + 1)}{n!\Gamma(\tilde{\alpha} + \tilde{\beta} + n + 1)\Gamma(\tilde{\alpha} + \tilde{\beta} + 2)\Gamma(\tilde{\beta} + n + 1)} \right]^{1/2} \\ &\quad \sum_{m=0}^{n} \binom{n}{m} \frac{\Gamma(\tilde{\alpha} + \tilde{\beta} + n + m + 1)}{\Gamma(\tilde{\alpha} + m + 1)} \frac{(y_t - \gamma_2)^m}{(\gamma_2 - \gamma_1)^m} \,. \qquad (4.3.32) \end{aligned}$$

with $\tilde{\alpha} = \frac{2b}{c} \frac{\gamma_2 - \beta}{\gamma_2 - \gamma_1} - 1$ and $\tilde{\beta} = \frac{2b}{c} \frac{\beta - \gamma_1}{\gamma_2 - \gamma_1} - 1$.

Chapitre 5

A quasi-likelihood approach based on eigenfunctions for a Jacobi process [1]

[1] This paper is co-authored with Christian Gouriéroux.

5.1. Introduction

The Jacobi process is the solution of the stochastic differential equation:

$$dy_t = -b(y_t - \beta)dt + \sqrt{cy_t(1-y_t)}dW_t \,, \tag{5.1.1}$$

with $b > 0$, $c > 0$ and $0 < \beta < 1$. b represents the mean-reverting parameter, β the mean of the process and c the volatility coefficient. This process is stationary and takes values between 0 and 1. This is a continuously-valued process whose values are restricted to the finite interval $[0, 1]$. As a result, it is appropriate to model dynamic bounded variables such as a regime probability, or to capture the evolution of a state price. Such a process is particularly appealing since it allows for substantial improvements in various applications, among which Markov switching regimes first introduced by Hamilton (1988,1989,1990) and later extended to smooth transition autoregressive models by Terasvirta and Anderson (1992), Terasvirta (1994) are the most well-known applications. Indeed, by allowing for a stochastic specification for the regime shift probabilities, it relaxes the somehow "unrealistic" predetermined features of the regimes usually encountered in the Markov switching regimes literature.

More recently, the Jacobi process appears very useful in credit risk modelling by relaxing the assumptions of predetermined states which rules out the possibility of zero pricing of default risk in the short run. Furthermore, the Jacobi process which allows for smooth continuous regime shifting can smooth jump processes such as jumps in default intensity due to default correlation among firms [see Jarrow and Yu (2001), Schonbucher and Schubert (2001), Gagliardini and Gouriéroux (2003), Dai and Singleton (2003)] and jumps in the credit rating in a credit migration model [see Bielecki and Rutkowski (2000), Gagliardini and Gouriéroux (2004)].

Further, in a risk-neutral world the Jacobi process can also be used to model the dynamic of state prices of any derivative written on an underlying asset [see Clement, Gouriéroux and Monfort (2000)].

Besides that it naturally extends jump processes to smooth stochastic processes, it

is also better suited for modelling the dynamic of currency processes. In this regards, the Jacobi process finds an additional application field in target zone models through the work of Larsen and Sorensen (2003) in which they generalize the target zone model proposed by De Jong, Drost and Werker (2001) by allowing asymmetry between currencies, which is a crucial feature of data on exchange rates in a target zone. This asymmetry is achieved by allowing a specific dynamic near the boundaries of the target zone. More specifically, when the exchange rate modelled by a Jacobi process gets near the boundaries, the diffusion coefficient becomes small and the drift (which models the intervention of the central banks) drives the process away from the boundaries.

For a more exhaustive description of the potential application areas of the Jacobi process, the reader is referred to Gouriéroux and Jasiak (2003). Therefore, the issue of estimating a Jacobi process which is, either very convenient to model the dynamic of a probability process or flexible enough to fit data on exchange rates in a target zone, appears undeniably relevant. Moreover its simplicity makes it tractable enough for efficiently estimating its parameters. However, although the process is defined in continuous time, the data are available in discrete time. Unfortunately, the likelihood function for discrete observations generated by a Jacobi process does not admit a closed-form expression and therefore the maximum likelihood is not feasible. To remedy this problem, we propose a technique based on nonlinear canonical analysis to approximate the unknown discrete-time transition function of the continuous-time Jacobi process. The approximation technique consists in truncating the spectral decomposition of the transition density derived from the spectral decomposition of the infinitesimal generator associated with the diffusion process. Since the expression of the approximation to the likelihood function is explicit, the maximum likelihood approach becomes feasible. Our technique to approximate the transition function is much simpler than that proposed in Aït-Sahalia (2002) and is much closer tailored to the genuine diffusion process since we do not proceed with any preliminary transformation of the data. The maximization of this quasi-likelihood function over the parameter space yields the quasi-maximum likelihood estimator, denoted QML. It is worth noting that

this estimation technique can easily be extended to any diffusion process, either scalar or multivariate, since it involves only the spectral decomposition of the infinitesimal generator of the diffusion.

The finite sample properties of this estimator are then compared with the properties of some estimators existing in the literature such as the Kessler and Sorensen's estimator (EIG) [see Kessler and Sorensen (1999)], which is basically a method-of-moments which exploits the spectral decomposition of the infinitesimal generator to build some unbiased martingale estimating functions. The optimal estimating function thus obtained can be thought of as an approximation in terms of eigenfunctions to the unknown score function. While these authors try to approximate the unknown score function, the approximated QML estimator approximates the unknown transition density. We also compare the QML estimator with the generalized method of moments (GMM) estimator, and with simulation-based estimators such as the simulated method of moments (SMM) estimator, or indirect inference (II) estimator and also with an exact indirect (EI) estimator based on an identifying constraint. Compared with these methods, the QML estimation method is easy to implement, no computationally intensive (compared to SMM in particular) and demonstrates good statistical properties.

The paper is organized as follows. Section 2 exploits the subordination properties to induce a Jacobi process, and its distributional properties are analyzed by means of nonlinear canonical analysis. Based on the distributional results of Section 2, we introduce the quasi-maximum likelihood estimator (QML) in Section 3. We also review some alternative estimation methods (GMM, SMM, EI, II, EIG), which will serve as benchmarks in assessing the finite sample properties of the QML estimator. How to simulate the Jacobi process is detailed in Section 4 before presenting the finite sample Monte Carlo results in Section 5. Finally concluding remarks are given in Section 6. The proofs are gathered in appendices.

5.2. Distributional properties of the Jacobi process

We review in this section distributional properties of the Jacobi process, which are useful to interprete the parameters of interest and to define appropriate estimation methods [see Gouriéroux, Renault and Valéry (2002)].

5.2.1. Time deformation

The standard family of distributions used to specify the distribution of a random variable y with range $[0,1]$ is the beta family. It is well-known that the beta distribution can be deduced from gamma distributions. Typically, if x_1 and x_2 are two independent gamma variables, $y = x_1/(x_1 + x_2)$ follows a beta distribution. The first result extends this property to continuous time stochastic processes. Let us recall that a Cox-Ingersoll-Ross (CIR) process admits marginal (resp. conditional) distributions which are gamma (resp. noncentered gamma) distributions. A Jacobi process can be deduced from a bivariate Cox-Ingersoll-Ross process by a time deformation. Let us consider the bivariate stationary Cox-Ingersoll-Ross process:

$$\begin{cases} dx_{1t} = -b(x_{1t} - \beta_1)dt + \sqrt{cx_{1t}}dW_{1t} , \\ dx_{2t} = -b(x_{2t} - \beta_2)dt + \sqrt{cx_{2t}}dW_{2t} , \end{cases} \tag{5.2.2}$$

where (W_{1t}) and (W_{2t}) are mutually independent standard Brownian motions and the mean-reverting [resp. volatility] parameters b [resp. c] are identical. The parameters are constrained by $b > 0$, $\beta_1 > 0$, $\beta_2 > 0$, $c > 0$. The two CIR processes are independent. Let us now consider the transformations: $y_{1t} = \frac{x_{1t}}{x_{1t}+x_{2t}}$ and $y_{2t} = x_{1t} + x_{2t}$. They define a process with range $[0,1]$ and a positive process, respectively. By Ito's lemma, the bivariate process (y_{1t}, y_{2t}) satisfies the bivariate stochastic differential system:

$$\begin{cases} dy_{1t} = -(b/y_{2t})[y_{1t}(\beta_1 + \beta_2) - \beta_1]dt + [c(y_{1t}/y_{2t})(1 - y_{1t})]^{1/2}d\tilde{W}_{1t} , \\ dy_{2t} = -b[y_{2t} - (\beta_1 + \beta_2)]dt + \sqrt{cy_{2t}}d\tilde{W}_{2t} , \end{cases} \tag{5.2.3}$$

where $(\tilde{W}_{1t})$ and $(\tilde{W}_{2t})$ are independent standard Brownian motions. Therefore the process (y_{2t}) is a CIR process with parameters b, $\beta_1 + \beta_2$ and c, whereas (y_{1t}) is a Jacobi process after time deformation. Indeed, let us define the time deformed process:

$$y^*_{1t} = y_{1\tau_t} \ , \tag{5.2.4}$$

where the time deformation :

$$\tau_t = \int_0^t y_{2u} du \ , \tag{5.2.5}$$

has stationary increments $\{y_{2t}\}$. The process (y^*_{1t}) satisfies the stochastic differential equation:

$$dy^*_{1t} = -b(\beta_1 + \beta_2)[y^*_{1t} - \frac{\beta_1}{\beta_1 + \beta_2}]dt + [cy^*_{1t}(1 - y^*_{1t})]^{1/2} d\tilde{W}^*_{1t} \ , \tag{5.2.6}$$

and is a Jacobi process. To summarize, a Jacobi process can be deduced from independent CIR processes x_1, x_2 by first applying the transformation $y_1 = x_1/(x_1 + x_2)$, and then a time deformation with increments $y_2 = x_1 + x_2$. We see below how this property can be used to derive the marginal distribution of a Jacobi process, integral expressions of its transitions, and also of course for simulation purpose.

5.2.2. Canonical decomposition

5.2.2.1. Spectral decomposition of the infinitesimal generator

It is known that the dynamic properties of a diffusion process y are characterized by the infinitesimal generator, which explains how to compute the infinitesimal drift of any transformation $P(y)$ of process y. The infinitesimal generator $\mathcal{A}$ is defined by:

$$\mathcal{A}P(y) = \lim_{h \to 0} \frac{1}{h} E[P(y_{t+h}) - P(y_t) | y_t = y] \ . \tag{5.2.7}$$

By applying Ito's formula, it is easily checked that the restriction of $\mathcal{A}$ to the set of twice continuously differentiable functions P is the differential operator:

$$\mathcal{A}P(y) = \mu(y)\frac{\partial P}{\partial y}(y) + \frac{1}{2}\sigma^2(y)\frac{\partial^2 P}{\partial y^2}(y) \ , \tag{5.2.8}$$

where μ and σ^2 are the infinitesimal drift and volatility, respectively. Thus for a Jacobi process the differential operator becomes:

$$\mathcal{A}P(y) = -b(y-\beta)\frac{\partial P}{\partial y}(y) + \frac{1}{2}cy(1-y)\frac{\partial^2 P}{\partial y^2}(y) \ . \tag{5.2.9}$$

For a Jacobi process, the infinitesimal generator admits a spectral decomposition, that is there exists a set of eigenvalues $\lambda_n,\ n \in \mathbb{N}\backslash\{0\}$, and eigenfunctions $P_n,\ n \in \mathbb{N}\backslash\{0\}$, such that:

$$\mathcal{A}P_n = \lambda_n P_n \quad , \forall\ n, \tag{5.2.10}$$

and $(P_n,\ n \in \mathbb{N}\backslash\{0\})$ generates the set of square integrable functions P. The spectral decomposition has been initially given by Wong (1964), [see also Hansen, Scheinkman (1995)]. The eigenvalues are negative given by $\lambda_n = -bn - \frac{1}{2}cn(n-1)$, whereas the eigenfunctions are polynomials, called Jacobi polynomials [see Abramowitz, Stegun (1965)]. They are given by:

$$P_n(y_t) = \left[\frac{\Gamma(\tilde{\alpha}+n)(2n+\tilde{\alpha}+\tilde{\beta}-1)\Gamma(\tilde{\alpha})\Gamma(\tilde{\beta})}{n!\Gamma(\tilde{\alpha}+\tilde{\beta}+n-1)\Gamma(\tilde{\alpha}+\tilde{\beta})\Gamma(\tilde{\beta}+n)}\right]^{1/2} \sum_{m=0}^{n}(-1)^m\binom{n}{m} \frac{\Gamma(\tilde{\alpha}+\tilde{\beta}+n+m-1)}{\Gamma(\tilde{\alpha}+m)}y_t^m \ , \tag{5.2.11}$$

with $\tilde{\alpha} = \frac{2b}{c}\beta$ and $\tilde{\beta} = \frac{2b}{c}(1-\beta)$. These polynomials define an orthonormal basis with respect to the inner product $< P, P^* >= \int P(y)P^*(y)d\nu(y)$, where ν is the beta distribution ($\beta(\frac{2b\beta}{c}, \frac{2b(1-\beta)}{c})$). We will see below that this distribution is the marginal distribution of the Jacobi process. These polynomials are standardized with respect to the beta distribution $\beta(\frac{2b\beta}{c}, \frac{2b(1-\beta)}{c})$, that is they satisfy: $E[P_n(y)] = 0$ and $V[P_n(y)] = 1$, for any n.

5.2.2.2. The conditional expectation operator

The infinitesimal generator measures the drift at very short horizon. However in practice the observations are available in discrete time $t = 1, 2, ...$, say, and the drift of the transformed process is measured at a fixed horizon, by convention equal to 1. For this reason, it is useful to introduce the conditional expectation operator $\mathcal{T}$ which associates with any transformation P the new transformation $\mathcal{T}P$ defined by:

$$\mathcal{T}P(y) = E[P(y_{t+1})|y_t = y] \ . \tag{5.2.12}$$

The conditional expectation operator $\mathcal{T}$ is simply the exponential of the infinitesimal generator $\mathcal{A}$:

$$\mathcal{T} = \exp \mathcal{A} \ .$$

Therefore it admits the spectral decomposition with eigenvalues $\exp \lambda_n$ and eigenfunctions P_n, $n \in \mathbb{N}\backslash\{0\}$.

5.2.2.3. Moment conditions

The spectral decomposition can be used to derive moment conditions satisfied by a Jacobi process. Indeed we get:

$$E[P_n(y_t)|y_{t-1}] = \exp(\lambda_n)P_n(y_{t-1}) \ \ \forall \ n \in \mathbb{N}\backslash\{0\} \ ,$$

and, by iterated expectation theorem we deduce a similar relation at any horizon h, $h \in \mathbb{N}\backslash\{0\}$:

$$E[P_n(y_t)|y_{t-h}] = \exp(\lambda_n h)P_n(y_{t-h}) \ \ \forall h \ , \ n \in \mathbb{N}\backslash\{0\} \ . \tag{5.2.13}$$

This set of moment conditions corresponding to degree smaller than n can be written equivalently in terms of power moments. More precisely we get:

$$E\left[\begin{pmatrix} 1 \\ y_t \\ y_t^2 \\ \vdots \\ y_t^n \end{pmatrix} | y_{t-h}\right] = A^{-1} diag \begin{bmatrix} \exp(\lambda_0 h) \\ \exp(\lambda_1 h) \\ \exp(\lambda_2 h) \\ \vdots \\ \exp(\lambda_n h) \end{bmatrix} A \begin{bmatrix} 1 \\ y_{t-h} \\ y_{t-h}^2 \\ \vdots \\ y_{t-h}^n \end{bmatrix},$$

where A is the $(n+1) \times (n+1)$ matrix independent of the lag h, which describes the coefficients of the polynomial eigenfunctions:

$$\begin{bmatrix} P_0(y_t) \\ P_1(y_t) \\ P_2(y_t) \\ \vdots \\ P_n(y_t) \end{bmatrix} = A \begin{bmatrix} 1 \\ y_t \\ y_t^2 \\ \vdots \\ y_t^n \end{bmatrix}.$$

Matrix A is lower triangular:

$$A = \begin{pmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ a_{10} & a_{11} & 0 & \dots & \dots & 0 \\ a_{20} & a_{21} & a_{22} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & 0 & 0 \\ a_{n0} & a_{n1} & a_{n2} & \dots & \dots & a_{nn} \end{pmatrix},$$

with coefficients given by equation (5.2.11).

Explicit expressions of the conditional power moments can be found by solving recursively the system of moment conditions (5.2.13):

$$\sum_{m=0}^{n} (-1)^m \binom{n}{m} \frac{\Gamma(\tilde{\alpha} + \tilde{\beta} + n + m - 1)}{\Gamma(\tilde{\alpha} + m)} E[y_t^m | y_{t-h}] = \exp(\lambda_n h) \sum_{m=0}^{n} (-1)^m \binom{n}{m}$$

$$\frac{\Gamma(\tilde{\alpha}+\tilde{\beta}+n+m-1)}{\Gamma(\tilde{\alpha}+m)} y_{t-h}^{m} \cdot$$

(5.2.14)

For instance we have:

$$E[y_t|y_{t-h}] = [1-\exp(-bh)]\frac{\tilde{\alpha}}{\tilde{\alpha}+\tilde{\beta}} + \exp(-bh)y_{t-h} \ ,$$

$$\begin{aligned} E[y_t^2|y_{t-h}] \ = \ & -\frac{\tilde{\alpha}(\tilde{\alpha}+1)}{(\tilde{\alpha}+\tilde{\beta}+1)(\tilde{\alpha}+\tilde{\beta}+2)}[1-\exp[(-2b-c)h]] + \frac{2(\tilde{\alpha}+1)}{(\tilde{\alpha}+\tilde{\beta}+2)} \\ & [1-\exp(-bh)]\frac{\tilde{\alpha}}{\tilde{\alpha}+\tilde{\beta}} + \frac{2(\tilde{\alpha}+1)}{(\tilde{\alpha}+\tilde{\beta}+2)}\{\exp(-bh)-\exp[(-2b-c)h]\}y_{t-h} \\ & + \exp[(-2b-c)h]y_{t-h}^2 \ , \end{aligned}$$

and the conditional variance at horizon h, that is the volatility at term h, depends on the past by means of an affine function of y_{t-h}, y_{t-h}^2. And so forth.

Then cross moments of the type $E(y_t^m y_{t-h}^l)$ are easily derived from the conditional power moments since: $E[y_t^m y_{t-h}^l] = E[y_{t-h}^l E[y_t^m|y_{t-h}]]$, where $E[y_t^m|y_{t-h}]$ is given by equation (5.2.14). For instance, we get:

$$\begin{aligned} E[y_t^2 y_{t-h}^2] \ = \ & \frac{-\left[(\frac{2b}{c}\beta)^2+\frac{2b}{c}\beta\right]}{(\frac{2b}{c})^2+3(\frac{2b}{c})+2}(1-\exp[(-2b-c)h])k_2(\theta) \\ & +2\frac{\left[(\frac{2b}{c})\beta^2+\beta\right]}{\frac{2b}{c}+2}[1-\exp(-bh)]k_2(\theta) \\ & +2\frac{\left[(\frac{2b}{c})\beta+1\right]}{\frac{2b}{c}+2}\{\exp(-bh)-\exp[(-2b-c)h]\}k_3(\theta)+\exp[(-2b-c)h]k_4(\theta) \ , \end{aligned}$$

where $k_i(\theta) \equiv E_\theta(y_t^i)$ denotes the marginal power moment of degree i [see Appendix

for closed-form expressions of $k_i(\theta)$ and of the cross moments].

5.2.3. Marginal and conditional distributions of the Jacobi process

The results above can be used to get some insight on the transition density of a Jacobi process, even if this transition does not admit a closed form expression.

i) **The marginal distribution.**

The marginal distribution of the Jacobi process is the beta distribution $\beta(\frac{2b\beta}{c}, \frac{2b(1-\beta)}{c})$. This result can be immediately deduced from the interpretation of the Jacobi process in terms of time deformation [see Section 5.2.1]. Since $y^*_{1t} = y_{1,\tau_t}$ where y_1 and τ are independent processes, the marginal distribution of y^*_{1t} coincides with the marginal distribution of $y_{1t} = \frac{x_{1t}}{x_{1t}+x_{2t}}$. Therefore this is the distribution of $\frac{x_{1t}}{x_{1t}+x_{2t}}$, where $\frac{2b}{c}x_{1t}$ and $\frac{2b}{c}x_{2t}$ are independent with distributions $\gamma(\frac{2b\beta_1}{c})$, $\gamma(\frac{2b\beta_2}{c})$, respectively, corresponding to the marginal distributions of the CIR processes, that is the beta distribution. The result follows.

ii) **An expression of the transition based on nonlinear canonical analysis.**

From the spectral decomposition of the infinitesimal generator, it is possible to deduce a decomposition of the transition density at any horizon h [see Lancaster (1968)]:

$$f^{(h)}(y_t|y_{t-h};\theta) = f(y_t;\theta)\left\{1 + \sum_{n=1}^{\infty} \exp(\lambda_n h) P_n(y_t) P_n(y_{t-h})\right\}, \qquad (5.2.15)$$

where $\lambda_n = -bn - \frac{1}{2}cn(n-1)$, P_n are the orthonormal polynomials defined at equation (5.2.11) and $f(y_t;\theta)$ denotes the marginal distribution of y_t [see i)].

5.3. Estimation methods

In this section we describe the different estimation methods of the parameter $\theta = (b, \beta, c)'$ of the Jacobi process $\{y_t\}$ from discretely sampled data $\{y_1, y_2, \ldots, y_T\}$.

5.3.1. (Approximate) Quasi-maximum likelihood

The maximum likelihood estimator $\hat{\theta}_T^{ML}$ of θ is defined by:

$$\hat{\theta}_T^{ML} = \arg\max_{\theta} \sum_{t=1}^{T} \log f(y_t|y_{t-1};\theta) \ ,$$

where $f(y_t|y_{t-1};\theta)$ denotes the transition density at horizon 1. It is conditional on the initial value y_0 of the process. Since the transition density has no closed form expression for the Jacobi process, the exact maximum likelihood approach is infeasible. However, it is possible to approximate the likelihood along the lines described below.

An approximation to the true transition density $f(y_t|y_{t-1};\theta)$ based on its spectral decomposition (which depends on θ) can be obtained by:

$$\tilde{f}_N(y_t|y_{t-1};\theta) = f(y_t;\theta)\{1 + \sum_{n=1}^{N} \exp(\lambda_n(\theta))P_n(y_t;\theta)P_n(y_{t-1};\theta)\} \ ,$$

for a large value of N, since

$$\lim_{N\to\infty} \tilde{f}_N(y_t|y_{t-1};\theta) = f(y_t|y_{t-1};\theta)$$

for each $\theta \in \Theta$, where $f(y_t|y_{t-1};\theta)$ has been defined at equation (5.2.15) and $f(y_t;\theta)$ denotes the marginal p.d.f. of the process which corresponds to a Beta distribution for the Jacobi process.

Then we can define the (approximate) quasi-maximum likelihood estimator $\hat{\theta}_{TN}^{QML}$ of θ as:

$$\hat{\theta}_{T,N}^{QML} = \arg\max_{\theta} \sum_{t=1}^{T} \log \tilde{f}_N(y_t|y_{t-1};\theta) \ .$$

This estimator based on a truncated version of the likelihood is asymptotically equivalent to the ML estimator when N tends to infinity with T at an appropriate rate [see e.g. Carrasco, Florens (2000)]. This rate depends on the rate of decrease of nonlinear correlations. Finally note that the truncated canonical decomposition $f_N(y_t|y_{t-1};\theta)$ is not necessarily positive, when N is fixed. This can create numerical problems in the opti-

mization due to the logarithm. However this case occurs with probability tending to 0 when N tends to infinity, and equivalent asymptotic results are obtained after replacing f_N by its absolute value in the expression of the log-likelihood function.

5.3.2. Method of moments

The idea of the method is to calibrate the values of the parameters on well chosen conditional moments.

5.3.2.1. Selection of the moments

The basic moments selected for estimation purpose will be the first N conditional moments of the form $E[y_t|y_{t-1}], E[y_t^2|y_{t-1}], \ldots, E[y_t^N|y_{t-1}]$. When N is large, this set of conditional moments brings the same information as the score (due to the special canonical decomposition of the Jacobi process) and therefore the generalized method-of-moments (GMM) estimator [see Carrasco, Florens (2000)] becomes equivalent to the maximum likelihood estimator. Moreover, from Section 5.2.2, the conditional moments $E[y_t^N|y_{t-1}]$ are polynomials of order N; therefore it is equivalent to calibrate on marginal moments such as:

$$\begin{array}{llll} E_0 y_t = k_1(\theta_0), & E_0 y_t y_{t-1} = k_{11}(\theta_0), & & \\ E_0 y_t^2 = k_2(\theta_0), & E_0 y_t^2 y_{t-1} = k_{21}(\theta_0), & E_0 y_t^2 y_{t-1}^2 = k_{22}(\theta_0), & \\ E_0 y_t^3 = k_3(\theta_0), & E_0 y_t^3 y_{t-1} = k_{31}(\theta_0), & E_0 y_t^3 y_{t-1}^2 = k_{32}(\theta_0), & E_0 y_t^3 y_{t-1}^3 = k_{33}(\theta_0) \\ E_0 y_t^N = k_N(\theta_0), & E_0 y_t^N y_{t-1} = k_{N1}(\theta_0), & E_0 y_t^N y_{t-1}^2 = k_{N2}(\theta_0), & \ldots E_0 y_t^N y_{t-1}^N = k_{NN}(\theta_0) \, . \end{array} \tag{5.3.16}$$

In practice a finite number of relevant moments are selected. They will be chosen to be sufficiently informative, that is to provide insight on various features of the series such as skewness, kurtosis, volatility clustering, leverage effect, and to ensure the identification of the parameter of interest.

5.3.2.2. Identification issue

To determine how many conditional moments are required to identify the parameters of the Jacobi process, let us consider the first two conditional moments:

$$E[y_t|y_{t-1}] = [1 - \exp(-b)]\beta + \exp(-b)y_{t-1} \ ,$$

and

$$\begin{aligned} E[y_t^2|y_{t-1}] &= -\frac{[(\frac{2b}{c}\beta)^2 + \frac{2b}{c}\beta]}{[(\frac{2b}{c})^2 + 3(\frac{2b}{c}) + 2]}\Big(1 - \exp(-2b - c)\Big) + 2\frac{[\frac{2b}{c}\beta^2 + \beta]}{[\frac{2b}{c} + 2]}(1 - \exp(-b)) \\ &\quad +2\frac{\frac{2b}{c}\beta + 1}{\frac{2b}{c} + 2}\{\exp(-b) - \exp(-2b - c)\}y_{t-1} + \exp[(-2b - c)h]y_{t-1}^2 \ . \end{aligned}$$

Since the conditional moments are polynomials in y_{t-1}, they can be written as:

$$\begin{aligned} E[y_t|y_{t-1}] &= a_{11}y_{t-1} + a_{10} \\ E[y_t^2|y_{t-1}] &= a_{22}y_{t-1}^2 + a_{21}y_{t-1} + a_{20} \ . \end{aligned}$$

Thus the parameter of interest θ can be identified from these two conditional moments if the mapping $\theta \rightarrow (a_{11}, a_{10}, a_{22}, a_{21}, a_{20})$ is a one-to-one mapping. It is shown in Appendix that this identification condition is satisfied. More precisely the parameters can be identified from $E(y_t)$, $Var(y_t)$ and $Corr(y_t, y_{t-1})$. Typically, denoting by $\hat{m}_T$, $\hat{\sigma}_T^2$, $\hat{\rho}_T(1)$ the sample mean, sample variance and first order empirical correlation respectively, we can deduce by inverting this relation that $\hat{\beta}_T = \hat{m}_T$, $\hat{b}_T = -\ln(\hat{\rho}_T(1))$ and :

$$\begin{aligned} \hat{c}_T &= 2\ln(\hat{\rho}_T(1)) - \ln(1 - \frac{\hat{a}_{21}\hat{m}_T + \hat{a}_{20}}{\hat{\sigma}_T^2 + \hat{m}_T^2}) \\ &\equiv \hat{g}(\hat{m}_T, \hat{\sigma}_T^2, \hat{\rho}_T(1)) \ . \end{aligned}$$

This suggests a guideline for the choice of the moment conditions to include in the

indirect estimation procedure below[2].

5.3.2.3. An exact indirect estimator

An exact indirect estimator is a moment estimator based on a number of moments equal to the parameter size. A simple indirect estimator can be based on the identifying constraint which associates a unique parameter value to the summary statistics m, σ^2, $\rho(1)$. Let us denote by $\hat{a}_T = [\hat{m}_T, \hat{\sigma}^2_T, \hat{\rho}_T(1)]'$ the sample counterpart of these moments. $\hat{a}_T$ tends asymptotically to $a(\theta) = [k_1(\theta), k_2(\theta) - k_1(\theta)^2, \rho(1,\theta)]$, where $\rho(1,\theta) = \frac{k_{11}^{(1)}(\theta) - k_1(\theta)^2}{k_2(\theta) - k_1(\theta)^2}$. The exact indirect estimator denoted by $\hat{\theta}_T^{EI}$ is solution of:

$$\hat{a}_T = a[\hat{\theta}_T^{EI}] .$$

5.3.2.4. Generalized-method-of-moments estimator

The summary statistics m, σ^2, $\rho(1)$ are functions of first and second order moments of the pair (y_t, y_{t-1}). Thus we can expect an improvement of the estimator by considering a larger set of moments and applying GMM. We consider below a set of moments also including third and fourth marginal moments to account for skewness and kurtosis, as well as cross moments of the type $E(y_t y_{t-1}^2)$ (to capture the risk premium), $E(y_t^2 y_{t-1}^2)$ (to capture the possible volatility persistence). More precisely equation (5.3.16) can be rewritten under a vector form as:

$$E_0[K(y_t) - k(\theta_0)] = 0 .$$

[2]The identification issue can be considered for any lag h. We have just to replace $b \to bh$, $c \to ch$, $\hat{\rho}_T(1) \to \hat{\rho}_T(h)$. In particular another consistent estimator of b is $\hat{b}_T = -\ln(\hat{\rho}_T(h))/h$. The comparison of the estimated values $-\ln(\hat{\rho}_T(h))/h$, h varying, can be the basis of a specification test for the Jacobi hypothesis.

Typically, the set of moment conditions selected for implementing GMM is:

$$K(y_t) - k(\theta) = \begin{pmatrix} y_t - k_1(\theta) \\ y_t y_{t-1} - k_{11}^{(1)}(\theta) \\ y_t^2 - k_2(\theta) \\ y_t y_{t-1}^2 - k_{12}^{(1)}(\theta) \\ y_t^2 y_{t-1}^2 - k_{22}^{(1)}(\theta) \\ y_t^3 - k_3(\theta) \\ y_t^4 - k_4(\theta) \end{pmatrix} .$$

The GMM estimator is defined by:

$$\hat{\theta}_T^{GMM}(\Omega) = \arg\min_{\theta} \left(\sum_{t=1}^{T} [K(y_t) - k(\theta)] \right)' \hat{\Omega}^{-1} \left(\sum_{t=1}^{T} [K(y_t) - k(\theta)] \right) ,$$

where $\hat{\Omega}$ is a consistent estimator of the asymptotic variance covariance matrix of $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} [K(y_t) - k(\theta)]$. It can be obtained through a Bartlett kernel estimator [see Newey and West (1987)] as

$$\hat{\Omega} = \hat{\Gamma}_0 + \sum_{k=1}^{K} \left(1 - \frac{k}{K+1} \right) (\hat{\Gamma}_k + \hat{\Gamma}_k') ,$$

where:

$$\hat{\Gamma}_k = \frac{1}{T} \sum_{t=k+1}^{T} [K(y_{t-k}) - k(\tilde{\theta}_T)][K(y_t) - k(\tilde{\theta}_T)]' ,$$

where $\tilde{\theta}_T$ is any consistent estimator of θ and K tends to infinity with T at an appropriate rate.

5.3.2.5. Estimating equations based on eigenfunctions

As an alternative to the quasi-likelihood approach aforementioned, Kessler and Sorensen (1999) propose a moment estimator which also exploits the spectral decomposition of the infinitesimal generator to build unbiased martingale estimating functions.

The optimal estimating function thus obtained can be seen as an approximation in terms of eigenfunctions to the unknown score function. While these authors try to approximate the unknown score function, the (approximate) QML estimator approximates the unknown transition density itself. It is worth noting that the GMM methodology presented earlier in this section does not use optimal instruments. Therefore, we should expect the latter not to behave in finite samples as well as the GMM of Kessler and Sorensen based on optimal instruments. Indeed, Kessler and Sorensen exploits the spectral decomposition to derive optimal instruments in the sense of Hansen (1982), Godambe and Heyde (1987), [see also Heyde (1997)]. More precisely they note that the functions:

$$h_j(y_t, y_{t-1}; \theta) = P_j(y_t; \theta) - \exp[\lambda_j(\theta)] P_j(y_{t-1}; \theta) \quad , j = 1, \ldots, N, \tag{5.3.17}$$

where P_j, λ_j denote the eigenfunctions and eigenvalues, respectively, satisfy the conditional restrictions:

$$E[h_j(Y_t, Y_{t-1}; \theta) | Y_{t-1}] = 0 \quad , j = 1, \ldots, N \; . \tag{5.3.18}$$

These restrictions imply marginal restrictions of the type:

$$E[\sum_{j=1}^{N} \alpha_j(Y_{t-1}) h_j(Y_t, Y_{t-1}; \theta)] = E[\alpha(Y_{t-1}) h(Y_t, Y_{t-1}; \theta)] = 0 \; , \tag{5.3.19}$$

where the instruments $\alpha_j(Y_{t-1})$ have a dimension equal to the parameter size. The moment estimator deduced from these marginal restrictions is solution of:

$$\sum_{t=1}^{T} \alpha(y_{t-1}) h(y_t, y_{t-1}; \theta) = 0 \; . \tag{5.3.20}$$

As usual the asymptotic variance-covariance matrix of this estimator depends on the selected instruments. The optimal choice of the instruments is obtained by considering the best prediction of the conditional score on the estimating functions [see e.g. Go-

dambe and Heyde (1987)]. More precisely, let us denote by $f(y_t|y_{t-1};\theta_0)$ the transition density of the process. The score is $\frac{\partial \log f(y_t|y_{t-1};\theta_0)}{\partial\theta}$ and the best conditional prediction of the score based on the estimating functions $h_j(y_t, y_{t-1};\theta_0)$ is:

$$B(y_{t-1};\theta_0)C(y_{t-1};\theta_0)^{-1}h(y_t, y_{t-1};\theta_0) \ , \tag{5.3.21}$$

where:

$$h(y_t, y_{t-1};\theta_0) = (h_1(y_t, y_{t-1};\theta_0), \ldots, h_N(y_t, y_{t-1};\theta_0))' \ , \tag{5.3.22}$$

$$C(y_{t-1};\theta_0) = E_{\theta_0}[h(Y_t, y_{t-1};\theta_0)h(Y_t, y_{t-1};\theta_0)'|y_{t-1}] \ , \tag{5.3.23}$$

$$B(y_{t-1};\theta_0) = E_{\theta_0}[\frac{\partial \log f(Y_t|y_{t-1};\theta_0)}{\partial\theta}h(Y_t, y_{t-1};\theta_0)'|y_{t-1}] \ . \tag{5.3.24}$$

Thus the optimal moment estimator based on $h_1, \ldots, h_N$ is the solution of:

$$\sum_{t=1}^{T} \hat{\alpha}^*(y_{t-1})h(y_t, y_{t-1};\hat{\theta}^*) = 0 \ , \tag{5.3.25}$$

where $\hat{\alpha}^*(y_{t-1})$ is a consistent approximation of:

$$\alpha^*(y_{t-1}) = B(y_{t-1};\theta_0)C(y_{t-1};\theta_0)^{-1} \ .$$

To implement this estimator, we will use the optimal estimating function $G_T^*(\theta)$ with the above optimal instruments α^* evaluated at a preliminary consistent estimate denoted by $\tilde{\theta}_T$, (e.g. the GMM estimate). Let:

$$G_T^*(\theta) = \sum_{t=1}^{T} B(y_{t-1};\tilde{\theta}_T)C(y_{t-1};\tilde{\theta}_T)^{-1}h(y_t, y_{t-1};\theta) \ , \tag{5.3.26}$$

where $h = (h_1, \ldots, h_N)'$ with $h_j(y, x;\theta) \ , j = 1, \ldots, N$ such that:

$$h_j(y, x;\theta) = P_j(y;\theta) - \exp(\lambda_j(\theta))P_j(x;\theta) \ . \tag{5.3.27}$$

The eigenfunctions and eigenvalues are the Jacobi polynomials and the corresponding

eigenvalues, respectively. On the other hand, the estimating function is a one-to-one mapping, that is, when taking its expectation w.r.t. the true probability distribution, parameterized by the true parameter θ_0, then it has to satisfy $E_{\theta_0} G_T^*(\theta_0) = 0$ which identifies the true parameter θ_0 without ambiguity. Since $G_T^*(\theta)$ has a dimension equal to θ [$\dim(\theta) = p = 3$], the estimator $\hat{\theta}_T$ is easily obtained by solving an explicit system of $p = 3$ equations with 3 unknowns, namely, by solving the estimating equation $G_T^*(\theta) = 0$. Before solving the equations, we need to compute the matrices defining the optimal instruments. Given that the eigenfunctions are polynomials of the type: $P_j(y;\theta) = \sum_{k=0}^{j} a_{j,k}(\theta) y^k$, and further are also eigenfunctions for the conditional expectation operator associated with the eigenvalues $\exp(\lambda_j(\theta))$, the computation of these matrices involving θ can be simplified. Further, it is easily shown that the $p \times N-$matrix,

$$B(x;\theta_0) = E_{\theta_0}[\frac{\partial \log f(y|x;\theta_0)}{\partial \theta} h(y,x;\theta_0)'|x] = -E_{\theta_0}[\partial_{\theta_0} h(y,x;\theta_0)|x] \ .$$

As a result the matrix $B(x;\theta) = \{b_{ij}(x;\theta)\}$ has entries of the type:

$$b_{ij}(x;\theta) = \sum_{k=0}^{j} \partial_{\theta_i} a_{j,k}(\theta) \int_0^1 y^k f(y|x;\theta) dy - \partial_{\theta_i}(e^{\lambda_j(\theta)} P_j(x;\theta)) \ , \qquad (5.3.28)$$

and the $N \times N-$matrix $C(x;\theta) = \{c_{ij}(x;\theta)\}$ defined at equation (5.3.22) has entries of the type:

$$c_{ij}(x;\theta) = \sum_{r=0}^{i} \sum_{s=0}^{j} a_{i,r}(\theta) a_{j,s}(\theta) \int_0^1 y^{r+s} f(y,|x;\theta) dy - e^{\lambda_i(\theta)+\lambda_j(\theta)} P_i(x;\theta) P_j(x;\theta) \ . \qquad (5.3.29)$$

See Appendix for a proof. The weighting matrices $B(x;\theta)$ and $C(x;\theta)$ will be evaluated at a preliminary consistent estimate of θ available in practice. These coefficients require the computation of the integrals of the type $\int_0^1 y^k f(y|x;\theta) dy$, for $k = 0, 1, 2, \ldots, 2N$. To compute these integrals, we exploit again the fact that the Jacobi polynomials $P_j(y;\theta)$ are also eigenfunctions for the conditional expectation operator associated with the eigenvalues $\exp(\lambda_j(\theta))$, see Appendix 5.6 for a proof. These inte-

grals are computed recursively, $j = 1, 2, \ldots, 2N$ using the fact that $\int_0^1 f(y|x;\theta)dy = 1$ since this is the conditional density of y_t given that $y_{t-1} = x$, to start the recursion. Finally, the estimator of $\theta = (b, \beta, c)'$ denoted by $\hat{\theta}_N^{EIG}$ is obtained as the solution to the explicit system (5.3.26) of $p = 3$ equations. Kessler and Sorensen (1999) showed that for N going to infinity, the optimal estimating function of the type (5.3.26) will converge to the score function. Larsen and Sorensen (2003) applies this estimator for fitting a Jacobi process to exchange rates data.

5.3.3. Simulated methods

We consider two simulated methods that are the simulated method of moments (SMM) and the indirect inference. These approaches require artificial data sets simulated from the Jacobi dynamics. Let $(y_1^s(\theta), \ldots, y_T^s(\theta))$ with $s = 1, \ldots, S$ denote the simulated data sets, with parameter $\theta = (b, \beta, c)'$ [see Section 5.4.1 for the description of the simulation procedure].

5.3.3.1. The simulated method of moments

This method is essentially a moment method, in which the theoretical moments are approximated by simulation. The SMM estimator denoted by $\hat{\theta}_T^{SMM}$ is then defined as:

$$\hat{\theta}_T^{SMM} = \arg\min_{\theta} \Psi_{ST}(\theta) ,$$

where:

$$\Psi_{ST}(\theta) = \{\sum_{t=1}^{T}[K(y_t) - \frac{1}{S}\sum_{s=1}^{S} K(y_t^s(\theta))]\}'\hat{\Omega}^{-1}\{\sum_{t=1}^{T}[K(y_t) - \frac{1}{S}\sum_{s=1}^{S} K(y_t^s(\theta))]\} ,$$

and:

$$\hat{\Omega} = \hat{\Gamma}_0 + \sum_{k=1}^{K}\left(1 - \frac{k}{K+1}\right)(\hat{\Gamma}_k + \hat{\Gamma}_k') ,$$

where:

$$\hat{\Gamma}_k = \frac{1}{T}\sum_{t=k+1}^{T}[K(y_{t-k}) - \frac{1}{S}\sum_{s=1}^{S}K(y_t^s(\tilde{\theta}_T))][K(y_t) - \frac{1}{S}\sum_{s=1}^{S}K(y_t^s(\tilde{\theta}_T))]' ,$$

and $\tilde{\theta}_T$ is any consistent estimator of θ.

When the number of replications S tends to infinity, $\frac{1}{S}\sum_{s=1}^{S}K(y_t^s(\theta))$ tends to $k(\theta)$ and the estimator $\hat{\theta}_T^{SMM}$ coincides with the GMM estimator corresponding to the same moment conditions. Of course in our framework where closed-form expressions of the moments are available, GMM approach is preferred from the asymptotic point of view. However it can be informative to compare the finite sample properties of GMM and SMM. Indeed, some diminution of the finite sample bias is often observed with simulation based methods.

5.3.3.2. The indirect inference method

The indirect inference method (hereafter II) [see Gouriéroux, Monfort and Renault (1993)], is a calibrating method based on an instrumental model which approximates the true model, that is the Jacobi process, but is easier to estimate. The instrumental model is naturally deduced from the Euler discretization of the s.d.e. (5.1.1). The instrumental model corresponds to the autoregression with conditional heteroscedasticity:

$$y_{t+1} - y_t = -b^*(y_t - \beta^*) + \sqrt{c^* y_t(1-y_t)}\epsilon_{t+1} ,$$

where the errors ϵ_{t+1} are independent with standard normal distribution. After a change of parameters where $\alpha = 1 - b^*$, $\gamma = b^*\beta^*$ and $\delta = \sqrt{c^*}$, the instrumental model can be rewritten as an autoregressive model linear in the new parameters α, γ.

$$\frac{y_{t+1}}{\sqrt{y_t(1-y_t)}} = \alpha\frac{y_t}{\sqrt{y_t(1-y_t)}} + \frac{\gamma}{\sqrt{y_t(1-y_t)}} + \delta\epsilon_{t+1} .$$

The parameter $g = (\alpha, \gamma, \delta)$ can be estimated by OLS by regressing $\frac{y_{t+1}}{\sqrt{y_t(1-y_t)}}$ on $\frac{y_t}{\sqrt{y_t(1-y_t)}}$ and $\frac{1}{\sqrt{y_t(1-y_t)}}$. $\hat{g}_T$ denotes the associated estimator. We can also compute

$\hat{g}_T^s(\theta)$ the OLS estimate of the transformed parameter based on a data set $y_T^s(\theta) = (y_1^s(\theta), \ldots, y_T^s(\theta))$, for $s = 1, \ldots, S$, simulated under the structural model for a value θ of the parameter of interest.

The indirect inference estimator of θ is defined by choosing a value $\hat{\theta}_T^{II}$ for which $\hat{g}_T$ and $\hat{g}_T^s(\theta)$ are as close as possible:

$$\hat{\theta}_T^{II}(\Omega) = \arg\min_{\theta} \left[\hat{g}_T - \frac{1}{S}\sum_{s=1}^{S} \hat{g}_T^s(\theta)\right]' \Omega \left[\hat{g}_T - \frac{1}{S}\sum_{s=1}^{S} \hat{g}_T^s(\theta)\right],$$

where Ω is a symmetric nonnegative matrix defining the metric. Since the instrumental model has the same number of parameters as the initial Jacobi process, the estimator does not depend on the choice of Ω and we select $\Omega = I_3$ in the application.

5.4. Simulation of the Jacobi process

In this section we explain how the discrete time sampled Jacobi process can be simulated by means of a truncated Euler approximation or by using the time deformation interpretation. Then we check the accuracy of the simulated path for different sets of parameter value.

5.4.1. A truncated Euler scheme

The Jacobi process is simulated through an Euler discretization of the stochastic differential equation with a small time unit δ, where the values of the process y_t are truncated to restrict them to the range $[0, 1]$.

The Euler approximation of the equation is:

$$y_{(k+1)\delta} = y_{k\delta} - b(y_{k\delta} - \beta)\delta + \sqrt{c y_{k\delta}(1 - y_{k\delta})}\sqrt{\delta}\epsilon_k ,$$

where ϵ_k, k varying, are independent standard normal variables. However, this technique does not ensure values between 0 and 1 and therefore the positivity of the volatility term. To satisfy this restriction, we truncate the values out of the range $(0, 1)$.

Typically, when y takes a value greater than or equal to 1, we set $y = 0.99$ and, if y takes a value less than or equal to 0, we set $y = 0.01$. Thus the truncated Euler scheme is defined by:

$$y^s_{(k+1)\delta} = \begin{cases} 0.99\ , & if \quad y^{*s}_{(k+1)\delta} \geq 1\ , \\ y^{*s}_{(k+1)\delta}\ , & if \quad 0 \leq y^{*s}_{(k+1)\delta} \leq 1\ , \\ 0.01\ , & if \quad y^{*s}_{(k+1)\delta} < 0\ , \end{cases} \tag{5.4.30}$$

where:

$$y^{*s}_{(k+1)\delta} = y^s_{k\delta} - b(y^s_{k\delta} - \beta)\delta + \sqrt{c y^s_{k\delta}(1 - y^s_{k\delta})}\sqrt{\delta}\epsilon^s_k\ , \tag{5.4.31}$$

and ϵ^s_k are independent drawings in the standard normal distribution.

The truncated Euler discretization scheme has to be applied with a small time unit δ, to get a good approximation of the underlying continuous time path. For the illustration we select $\delta = 1/10$. Thus we first simulate by the truncated scheme (5.4.30-5.4.31) the underlying values corresponding to dates $1/10, 2/10, 3/10, ...$ The simulated discrete time path is deduced by considering only the integer time indexes that are $y^s_t = y^s_{k\delta}$, with $k = t/\delta$. They correspond to $k = 10$ for $t = 1$, $k = 20$ for $t = 2$ and so forth.

5.4.2. Simulation scheme based on time deformation

A simulation scheme can also be designed from the interpretation of a Jacobi process in term of time deformation as follows.

- **step 1:** Simulate two very long sample paths from independent χ^2 distributions [or equivalently from gamma distributions] such that:

$$x^{(s)}_{i,k\delta}\ , \quad i = 1, 2\ ,$$

  where the time unit has been fixed to $\delta = 1/10$.

- **step 2:** To generate a variable $y_{1,k\delta}$ with a beta distribution, apply the transformation:

$$y^{(s)}_{1,k\delta} = \frac{x^{(s)}_{1,k\delta}}{x^{(s)}_{1,k\delta} + x^{(s)}_{2,k\delta}}\ ,$$

where the $x_{i,k\delta}^{(s)}$ come from step 1.

- **step 3:** To simulate the time deformation process $\tau_{k\delta} = \int_0^{k\delta}(x_{1u} + x_{2u})du$, sum up:

$$\tau_{k\delta}^{(s)} = \frac{1}{\delta}\sum_{l=1}^{k}(x_{1,l\delta}^{(s)} + x_{2,l\delta}^{(s)}) \ .$$

- **step 4:** To produce draws from a Jacobi process, we select the simulated values $y_{1,k\delta}^{(s)}$ for which the time deformation process reaches integer values, such that:

$$y_{1t}^{*(s)} = y_{1,k\delta}^{(s)} \ , when \ , \tau_{k\delta}^{(s)} \in \mathbb{N} \ ,$$

with k varying.

The random draws of the Jacobi process thus produced will be unequally spaced in time.

5.4.3. Simulated series

The approach above is followed to simulate paths of the Jacobi process. The length of the path is $T = 2000$, and the parameters are fixed at different values:

- parameter set I : $(0.43, 0.5, 0.8)$
- parameter set II : $(0.5, 0.5, 0.25)$

The different parameter sets have been selected to reproduce the two typical patterns of the marginal beta distribution. Sets I and II correspond to symmetric beta distribution, with more weights on boundary values 0-1 for set I. The dynamics also differ. The processes associated with parameter sets I and II admit rather high first order correlation, larger than 0.6. Simulated paths of the Jacobi process and of its transformations corresponding to the first three canonical polynomials are provided in Figures 1 and 2 for parameter sets I and II.

[Insert Figure 5.1: Simulated paths, set I]

[Insert Figure 5.2: Simulated paths, set II]

Due to the choice of parameter values, the process distribution can give more or less weight in a neighborhood of the limiting values 0 and 1. Larger weights on extremes can be immediately observed on simulated paths. Moreover since the autocorrelation is rather high ($\simeq 0.6$) for set I we observe also some extreme clustering. Indeed when y_t is close to 0 or 1, the random component in equation (5.4.31) is close to zero and the equation becomes almost deterministic.

The paths associated with the canonical directions are simply polynomial transformations of the initial path. Nonlinear features can be observed, such as skewed paths for the second degree polynomial, or more extreme phenomenon for polynomial of degree 3. Distributional properties of the paths can also be derived by replicating the simulations. The number of replications is $M = 1000$. We provide in Figures 5.3 and 5.4 the empirical marginal distributions of y_t, $P_2(y_t)$, $P_3(y_t)$ for the first two sets of parameter values.

[Insert Figure 5.3: Empirical marginal distributions, set I]

[Insert Figure 5.4: Empirical marginal distributions, set II]

The comparison between the sample distribution for y_t and the theoretical beta distribution (see the first row of Figures 5.3 and 5.4) gives some information on the accuracy of the simulations as well as Table 5.1 reporting summary statistics.

The skewness [resp. fat tail] effects are also clearly seen on the sample distribution of $P_2(y_t)$ [resp. $P_3(y_t)$] on Table 5.2. The sample means are close to zero whereas the variances close to 1 for the three polynomials. This corresponds to the normalization of polynomials in the canonical decomposition.

We have seen that the empirical results concerning marginal distributions coincide with the expected theoretical results. Let us now focus on dynamic features.

From the theoretical results we expect that the processes $P_1(y_t)$, $P_2(y_t)$, $P_3(y_t)$ are not correlated and are autoregressive of order one. Figures 5.5, 5.6 provide the joint autocorrelogramms of the three series for the two sets of parameter values. The dashed lines represent the confidence bands of plus or minus twice standard deviations computed under the i.i.d hypothesis. The absence of cross correlation is clearly seen on

Table 5.1. Summary statistics for y_t and beta distr.

set I		
	sample moments of y_t	theoretical beta
mean	0.4891	0.5
variance	0.1063	0.1204
skewness	0.056	0
kurtosis	1.6140	1.5276
set II		
mean	0.4960	0.5
variance	0.0493	0.05
skewness	0.0337	0
kurtosis	2.2051	2.1428

Table 5.2. Summary statistics for P_1,P_2,P_3.

set I				
	mean	variance	skewness	kurtosis
$P_1(y_t)$	0.0313	0.8826	-0.056	1.614
$P_2(y_t)$	-0.1608	0.9021	0.2477	1.622
$P_3(y_t)$	-0.0227	0.9275	0.0222	1.5339
set II				
	mean	variance	skewness	kurtosis
$P_1(y_t)$	0.016	0.9869	-0.0337	2.2051
$P_2(y_t)$	-0.0124	1.0265	1.2733	3.8740
$P_3(y_t)$	-0.0174	1.0161	0.1509	6.7443

Table 5.3. Sample and theoretical correlations for P_1,P_2,P_3.

set I				
	$\hat{\rho}(1)$	$\rho(1)$	$\hat{\rho}(2)$	$\rho(2)$
$P_1(y_t)$	0.587	0.650	0.334	0.4231
$P_2(y_t)$	0.136	0.1901	-0.007	0.0361
$P_3(y_t)$	0.064	0.025	0.017	6.23E-4
set II				
	$\hat{\rho}(1)$	$\rho(1)$	$\hat{\rho}(2)$	$\rho(2)$
$P_1(y_t)$	0.585	0.606	0.328	0.367
$P_2(y_t)$	0.274	0.286	0.041	0.082
$P_3(y_t)$	0.097	0.105	-0.018	0.011

the correlogramms, but the autoregressive dynamics is more difficult to detect on the autocorrelogramms shown in Figures 5.5, 5.6 except for the first polynomial. For this reason we also provide a plot for another set of parameter values $(0.1, 0.5, 0.03)$ corresponding to a very high correlation level ($\simeq 0.9$). For this set the typical exponential decay is clearly seen for the three polynomials. A complementary information is provided in Table 5.3 where are reported the sample and theoretical first and second order correlations.

[Insert Figure 5.5: Empirical correlations, set I]

[Insert Figure 5.6: Empirical correlations , set II]

[Insert Figure 5.7: Empirical correlations , set $(0.1, 0.5, 0.03)$]

To summarize, the comparison of the empirical and theoretical results concerning the Jacobi process and its transformations allows for the validation of the simulation scheme based on the Euler discretization, both for marginal and dynamic features.

5.5. Comparison of the estimators

5.5.1. The estimation methods

The aim of this section is to compare by simulations the finite sample properties of the quasi-maximum likelihood (QML) estimator with the moment estimator of Kessler and

Sorensen (EIG), which uses the spectral decomposition to approximate the unknown score function, together with the exact indirect estimator (EI), the GMM estimator, the SMM estimator and the indirect inference estimator (II). Different sample sizes are considered $T = 500, 1000, 2000, 3000$ and the number of simulations used in the Monte Carlo study is $M = 1000$.

The QML approach is applied with $N = 4$ terms in the canonical decomposition. This number has been chosen small and independent of the sample size to have an idea of the truncation bias. The Kessler and Sorensen estimator based on the eigenfunctions is implemented with the first two eigenfunctions as recommended by Larsen and Sorensen (2003). The weighting matrices defining the optimal instruments are evaluated at a preliminary consistent estimate of θ. In the subsequent Monte Carlo study, the GMM estimate of θ has been used to get a consistent estimate of the instruments. Finally, to solve the estimating function $5.3.26 = 0$, we use the nlsys library of the GAUSS sofware for solving a system of nonlinear equations in the unknown parameter θ. The EI approach calibrates the three parameters of the Jacobi process on the sample mean, variance and first order correlation. This set of moments is sufficient to identify the Jacobi parameters and can serve as a benchmark for other GMM estimation methods based on a larger set of moments. The GMM approach is applied with the seven moments described in Section 5.3.2.4. These moments include those used in the EI approach together with higher moments associated with skewness, kurtosis, and cross moments in order to capture more dynamic features. The GMM approach is performed in two steps. The first step estimator is obtained with the identity matrix replacing the weighting matrix in the GMM criterion. This preliminary estimator is then plugged into the Ω matrix to get a Newey-West estimator of the weighting matrix. The second step estimator is then obtained by minimizing the second step GMM criterion. Similarly, the SMM approach is the simulated version of the GMM approach, but, instead of comparing the sample moments to their theoretical analogs $k(\theta)$, we compare the sample moments to the simulated ones $K(y_t^s(\theta))$ averaged over $S = 10$ simulations. Since one expects simulated methods to correct for finite sample bias by means

of simulations, the SMM approach is meaningful to study the behavior of finite sample estimation bias. Similarly, the indirect inference approach can possibly diminish finite sample estimation bias [see Gouriéroux, Renault and Touzi (2000)]. It is important to compare the distributional properties of the estimators for different sample sizes. Such an analysis gives an idea of the number of observations necessary for the asymptotic theory to be valid and of how this number depends on parameter values. Moreover when the sample sizes are too small, we can detect the most important differences with asymptotic normality, such as skewness, fat tails, or multimodes. We first consider the comparison for each type of parameters, the mean reverting parameter b, the volatility coefficient c and the mean parameter β and in a second step the joint distribution of b and c .

5.5.2. Marginal properties of the estimated coefficients

i) Analysis of the bias.

Let us first consider the finite sample bias for the different estimation methods. In order to facilitate the comparison with respect to sample size and across experiment, we consider the bias standardized by T and divided by the true value of the parameter. Such a standardization will not change the interpretation of the bias since the parameter values are positive.

Due to the interpretation of the parameters we can expect less bias on the mean parameter β than on the mean reverting parameter b and on the volatility coefficient c. Moreover we expect a bias (resp. a standardized bias) tending to zero (resp. to a limit) when the sample size tends to infinity. Indeed,the standardized bias seem to stabilize towards a limit when the sample size increases which gives support for the fact that the bias seems to converge at a rate of $1/T$.

Let us consider the sign of the bias as reported in Table **??**. When examining parameter set II in Table **??**, the bias for the volatility coefficient c tends to be positive for the exact methods (QML,EIG,EI,GMM) in contrast to the simulated methods (SMM,II)

for which it is negative. A positive bias for c leads to overestimate the volatility. If one has in mind a credit barrier model which is mainly a derivative pricing model in which the underlying is a credit quality variable with the meaning of distance to default, then y may represent the dynamic of the default probability. In such a framework, a higher volatility increases the default probability.

In this sense, the negative bias is not a suitable property of the simulated methods since it may lead to underestimate the risk of default. However this observation is less clear in parameter set I. The same type of observation holds for the mean reverting parameter b which is biased upwards for the exact methods (QML,EIG,EI,GMM), while it tends to be negative for the simulated ones (SMM,II). It is known that the speed of reversion is relatively difficult to estimate without bias [see Larsen and Sorensen (2003)]. The exact methods seem to overestimate the speed of reversion and therefore might drive the manager to take more risk as the latter expect the probability of default occurrence to go back faster to its long-run equilibrium level modelled by the mean parameter β. Therefore, a risk-averse manager could be mistaken by an overestimation of the mean-reverting parameter b by getting a wrong appreciation of the evolution of the default probability. Concerning the sign of the bias for the mean parameter β, the sign is not constant and varies across experiments and across methods. Note that the sign of the bias for β is always positive for SMM. In particular, all the other methods exhibit a negative bias for β in parameter set II whereas SMM overestimates its bias. Paradoxically, SMM exhibits more bias for the mean parameter than the one observed for the other methods but the magnitude of the bias remains limited (less than 0.8).

Concerning the magnitude of the bias, we clearly see that SMM drastically reduces the finite sample bias of the mean reverting parameter b and to a lesser extent of the volatility coefficient c.The simulated moments seem to perform better than the indirect procedure in reducing the bias. Indeed, the indirect procedure suffers from singularity problems when trying to invert the $X'X$ matrix in the first step OLS estimator of the auxiliary parameter due to a lack of robustness of the latter in presence of outliers. This point is clearly highlighted in Genton and Ronchetti (2001) where they show that the

finite sample bias may not be necessarily negligible under a contaminated version of the DGP. Moreover, we can see that GMM exhibits slightly more bias than EI, which confirms the fact that including more moments in the estimation increases the magnitude of the bias [see Buse (1992), Andersen and Sorensen (1996), Chao and Swanson (2000)]. When the moment estimator of Kessler and Sorensen is not too affected by numerical instability, we note that the bias of the mean-reverting parameter b is less important than those observed for EI and GMM in parameter set I. This better performance of the EIG may be attributed to the selection of the optimal instruments of the latter. The same remark holds for the bias of the volatility coefficient c in parameter set II. The parameter with the strongest bias is the mean reverting parameter b mostly in set I, which corresponds to the marginal symmetric Beta distribution which puts more weight on the extreme values, whereas the mean parameter β is much less biased.

The mean reverting parameter b is less biased in set II which corresponds to the symmetric Beta distribution with more weight on the averaged values. But this is not as clear as for the volatility coefficient which seems to be less biased in set I. Note that the QML estimator tends to exhibit more bias (upwards) for the mean reverting parameter b than for the volatility coefficient c when compared with other methods. More specifically, QML demonstrates less bias for the volatility coefficient when compared with the estimator of Kessler and Sorensen based on eigenfunctions for all experiments. QML seems therefore to be potentially more accurate in evaluating the risk.

[Insert Table **??**: standardized biases, set I,II.]

ii) Analysis of the variance.

Let us now consider in Table 5.4, the variance of the estimators standardized by the sample size and divided by the square of the true value of the parameter in order to facilitate the comparison across sample size and experiment. Indeed, the variance of the estimators tends to diminish when the sample size increases whereas the standardized variance tends to a limit. Concerning the magnitude of the variance, QML is more accurate than the other methods. In particular, the estimator proposed by Kessler and

Sorensen exhibits a very large variance when the sample size is small. This large variability may be attributed to a lack of robustness of the $G_T^*(\theta)$ criterion defined at equation (5.3.26) to outliers. There is always a tradeoff between robustness and efficiency and the latter focuses on efficiency issues. But, at the same time, the $G_T^*(\theta)$ criterion is also sensitive to numerical instability, in particular in small samples. Indeed, the computation of the integrals as nonlinear functions of the parameters, fails a few times in small samples. Therefore, it is quite difficult to disentangle the effects stemming from essentially numerical instability from those due to robustness considerations. Fortunately, when the sample size grows, the situation is getting better. To get an insight on the relative efficiency of the QML estimator w.r.t. the EIG estimator, we compute some relative efficiency coefficients. In parameter set I when EIG is less affected by numerical instability, the relative efficiency of the estimator of b decreases from 0.7245 (6.8181/9.4099) for $T = 500$ toward 0.9884 (6.2937/6.3672) for $T = 3000$ in favor of QML. In other words, the QML estimator of the mean-reverting parameter is 27.5% more efficient than its EIG estimator in small sample; they become equivalent in large samples where for $T = 3000$ the QML estimator outperforms the EIG estimator of only 1.16% approximately. Concerning the diffusion parameter c in set I, the EIG estimator remains dominated by the QML estimator. Indeed, the relative efficiency for $T = 500$ is around 0.7117 (3.6209/5.0874) in favor of QML and this time the EIG estimator has some trouble to catch up with QML when the sample size increases, since the relative efficiency is still around 0.8652 (3.4640/4.0036) in favor of QML when $T = 3000$. Hence, despite the large sample size of $T = 3000$, the QML estimator remains more efficient than the EIG estimator of 13.48%. Note the extreme high variances displayed by II in parameter set I reported in Table 5.4 not only for the mean-reverting parameter but also for the volatility coefficient. They contrast sharply with the variances of its competitors. II is known to be a very general estimation method which can be used in estimating a large variety of models. However, the more general an estimation method is, the less precise it may be, and hence, the less efficient it appears. Although II is said to be suitable in estimating continuous-time diffusion processes since the crude dis-

cretization of the latter naturally provides an auxiliary model for estimation, it remains that its performance critically depends on the choice of the discretization scheme. As a result, what the indirect inference estimation method wins in generality it looses in efficiency. Further, when comparing EI and GMM variances, there is no obvious gain in efficiency in implementing the over-identified method which is more complicated and also computationally costly even though including moments of the form Ey_t^3, $Ey_t y_{t-1}^2$ in the GMM estimation may capture dynamic features such as skewness, leverage effect compared with the EI estimator. Moreover we may expect that GMM and SMM are equivalent at least asymptotically. The results seem to predict something in that sense mostly in set II with the sample sizes $T = 2000, 3000$. On the other hand, the QML estimator is quite appealing since it does not require estimating a weighting matrix as it is the case for EIG, GMM, SMM. Indeed, estimating a weighting matrix may arises numerical instability which may affect the estimation results. Moreover using a Bartlett estimator leads to a truncation bias in the estimation of Ω. Moreover, the more moments are included in the estimation procedure, the larger is the risk of colinearity and therefore the more trouble we have to invert the weighting matrix Ω. Again, we observe that the mean-reverting parameter exhibits more variability than the diffusion coefficient in both parameter sets.

[Insert Table 5.4: standardized variances, set I,II.]

iii) Analysis of skewness and kurtosis coefficients.

Globally the skewness and kurtosis coefficients reported in Tables 5.5 and 5.6 are getting closer to those characterizing normality(skew$=$ 0, and kurt. $=$ 3) when the sample size increases. QML suffers a little asymmetry around $T = 3000$ in set I mostly for the mean-reverting and the mean parameter, which may be attributed to the small number of polynomials in the expansion of the density function. Increasing the number N of polynomials in this expansion may take into account more dynamic features of the process such as heavy tails and asymmetries. Note again the extremely high kurtosis and skewness for EIG and II (greater than 10 in set I and 100 in set II)

in small samples due to some numerical instability. In this respect, QML behave better than the estimator of Kessler and Sorensen because of numerical instability arising from nonlinearities in the parameters for the latter. Note again some very similar results in terms of skewness and kurtosis for the just-identified moment estimator (EI) and the over-identified one (GMM). SMM also exhibits some good symmetric properties and does not appear leptokurtic. Unexpectedly, deviations from normality in terms of asymmetries and fat-tailed effects appears more severe in parameter set II than in set I and could say more in favor of the numerical instability hypothesis. Again the mean parameter demonstrates better distributional behavior than the mean-reverting parameter and the volatility coefficient.

[Insert Tables 5.5, 5.6: nonstandardized skewness and kurtosis coefficients, set I, II.]

iv) Analysis of marginal distributions.

The empirical marginal distributions of the parameters have been standardized as $\sqrt{T}(\hat{\theta}_T - \theta_0)$ where θ_0 denotes the true value of the parameter. The parameters are represented on the figures by column, (b, β, c) from the left to the right whereas the sample size increases from top to bottom ($T = 500, 1000, 2000, 3000$). We study different sample sizes to get an insight on the speed at which the asymptotic works. As expected, the asymptotic might work faster for SMM than displayed for GMM and QML. Increasing the number of polynomials in the expansion of the transition function might help QML for reaching the asymptotic behavior faster, that is in smaller sample sizes. The distributions are getting closer to the Gaussian distribution when the sample size increases and in particular more symmetric.Globally, the distribution of the mean-reverting parameter b exhibits more dispersion than the volatility coefficients c as already observed when analyzing the variances. In particular, we observe that the distributions of the mean-reverting parameter b are clearly biased to the right for the exact methods (QML, EIG, EI, GMM) in set I which correspond to the marginal Beta distribution which puts more weight on the extreme values. As for EI and GMM we should better choose the instruments to include in the procedure, ideally select the op-

timal instruments. By contrast, the simulated methods (SMM, II) display distributions which are more centered around 0 in set I and therefore achieves in correcting for the bias as usually expected for this kind of estimation methods. But this improvement has a cost in term of computation time since these methods are known to be very computation intensive. All the distributions of all standardized estimators of b are more centered around 0 for parameter set II and do not display any bias. The distribution of the mean parameter β and of the volatility coefficient c do not suffer from such bias and are globally centered around 0 for all parameter sets and all methods. They also converge faster to the Gaussian distribution than observed for the mean-reverting parameter b which appears undeniably as the most challenging parameter to estimate. Note also that the tails are getting thinner when the sample size grows, mostly for the EIG estimators of all parameters. We further observe that the distribution of the EIG estimator of the volatility coefficient is left-skewed for $T = 500, 1000$ in parameter set II. Indeed, we observe heavy tails and skewness features on the small sample distributions of the EIG estimators as already noticed when analyzing the kurtosis and skewness coefficients, respectively. Deviations from normality, in terms of skewness and fat-tailed effects, are more severe in small samples $T = 500, 1000$ for indirect inference, in particular in parameter set II with extremely high skewness and kurtosis coefficient as reported in Table 5.6.

[Insert Figures 5.8,5.9:Empirical marginal distributions of the EI estimates, set I,II] [Insert Figures 5.10,5.11:Empirical marginal distributions of the QML estimates, set I,II] [Insert Figures 5.12,5.13:Empirical marginal distributions of the EIG estimates, set I,II] [Insert Figures 5.14,5.15:Empirical marginal distributions of the GMM estimates, set I,II] [Insert Figures 5.16,5.17:Empirical marginal distributions of the SMM estimates, set I,II] [Insert Figure 5.18,5.19 :Empirical marginal distributions of the II estimates, set I,II]

5.5.3. Joint distributional properties of the estimators of b and c

In addition to marginal behaviors, studying the joint dynamic of the parameters could be informative concerning potential useful connections across them. Such an insight is particularly relevant if one wants to make derivative pricing. In this respect, understanding the joint dynamic of the parameters which drives the drift and diffusions functions respectively is essential for risk considerations.To do so, we have chosen to examine the joint behavior of the mean-reverting parameter b and of the volatility coefficient c since they are the parameters which drives the dynamic of the process, leaving aside the long-term mean parameter β which informs more on the long-run equilibrium level. The joint distributions of b and c have the typical ellipsoidal shape characterizing the bivariate Gaussian distribution for all estimators. The ellipsoids are fatter in set I than in set II where they are thinner illustrating that there is more dispersion in the estimated values obtained in parameter set I. We also observe for all estimators indistinctively, some outliers in small samples. The presence of outliers may be interpreted as a measure of occurrence of extreme events. Indeed, outliers are much numerous in parameter set I which correspond to the beta distribution which put more weight on the extreme values. In other words, the proportion of outliers could be interpreted as a measure of nonlinear dependence, that is as a measure of dependence in the tails whose the usual statistics of linear dependency are unable to capture. Thus, when the estimate of volatility parameterized by c tends to get too high, the estimate of the mean-reverting parameter b has to reach higher values in order drive the process back to the long-run equilibrium. The higher the volatility coefficient c is, the higher the probability for the process to hit the boundaries will be, hence the higher the mean-reverting parameter should be in order to be sufficiently forceful to keep the process stationary. Again we observe that the ellipsoids are biased to the right in set I for the exact methods (QML, EIG, EI, GMM), as already observed on the marginal distribution of the mean-reverting parameter b, whereas they are more centered around 0 in set II. Once again, the simulated estimation methods (SMM, II) correct for the bias observed in set I for the exact methods. Indeed, all the ellipsoids of SMM and II are centered around 0 whatever

sample size considered. Note also that SMM performs much better than II which unfortunately suffers from numerical instability which may produce some multimodes. Such a deviation from normality is unfortunate for II, since in such situations, we do not know which optimum to choose. Indeed for II, the procedure fails quite often due to some singularity issue stemming from the fact that the procedure gets very unstable numerically when the process hits the boundaries 0 and 1. As a result, the $X'X$ matrix, defining the OLS estimator of the auxiliary parameter, becomes non-invertible. As for the estimator EIG of Kessler and Sorensen, the procedure fails a few times as well, but in a much lesser extent than for II. These computational crashes arise from the computation of the integrals defining the weighting matrices B and C which are highly nonlinear in the parameters and creates some trouble to invert the weighting matrix C. The joint distribution of b and c are well-behaved for the QML estimator and seem to be more robust to outliers than the EIG estimator. In this view the QML estimator has an advantage over the EIG estimator and the II estimator since in the QML procedure we do not have to invert any matrix which may create some numerical instability of any kind. Further, as it is not computational intensive as II, it does not have to handle with some trouble arising at the boundaries. Concerning the EI and the GMM estimators, there is not obvious benefit to implement the over-identified method which is also more complicated, when one compares their joint distributions. This remark gives support once again to the detractors of over-identified methods. In general the optimal number of instruments allows for the just-identified case. Intuitively, to maximize efficiency, one should not include irrelevant instruments. This assertion is largely documented in the literature on asymptotic theory [see Buse (1992), Chao and Swanson(200)]. Dufour and Taamouti (2003) provides evidence on this issue through Monte Carlo methods.

[Insert Figures 5.20,5.21 :Empirical joint distributions of the EI estimates of b and c, set I,II.] [Insert Figures 5.22,5.23 :Empirical joint distributions of the QML estimates of b and c, set I,II.] [Insert Figures 5.24,5.25 :Empirical joint distributions of the EIG estimates of b and c, set I,II.] [Insert Figures 5.26,5.27:Empirical joint distributions of the GMM estimates of b and c, set I,II.] [Insert Figures 5.28,5.29:Empirical joint

distributions of the SMM estimates of b and c, set I,II.] [Insert Figures 5.30,5.31:Empirical joint distributions of the II estimates of b and c, set I,II.]

5.6. Concluding remarks

To summarize, the quasi-maximum likelihood estimator QML exhibits globally one of the best behaviors with respect to bias and variance, while being one of the easiest estimation method to implement, at the exception of the just-identified moment estimator EI. It is further no computation intensive and so very fast. Indeed, it rules out some difficulty like estimating a weighting matrix in the GMM, SMM, EIG procedure which may create some numerical instability due to some difficulties to invert the weighting matrix. These numerical difficulties can be avoided by resorting to the quasi-maximum likelihood. Further, the quasi-maximum likelihood estimator appears more robust to outliers than the estimator based on the eigenfunctions of Kessler and Sorensen or than the indirect inference procedure. The empirical joint distribution of the QML estimators of the volatility coefficient and of the mean-reverting parameter looks similar to the bivariate gaussian distribution for all sets.

Table 5.4. Standardized variance

set I						
T	*b*					
	QML	EIG	EI	GMM	SMM	II
500	6.8181	9.4099	8.8657	9.6319	7.6623	76.9341
1000	7.0358	7.2155	8.8517	9.2338	7.6680	61.8202
2000	6.5798	6.5723	7.9215	8.2802	7.3122	53.1076
3000	6.2937	6.3672	7.6626	8.0455	7.2385	73.1982
T	*β*					
	QML	EIG	EI	GMM	SMM	II
500	0.7927	6.8107	1.8001	1.7622	2.3917	1.1550
1000	0.7468	4.8142	1.7248	1.7030	2.4774	1.8748
2000	0.7768	4.8381	1.7412	1.7846	2.3639	1.0854
3000	0.7015	4.5428	1.5722	1.5919	2.3274	1.7815
T	*c*					
	QML	EIG	EI	GMM	SMM	II
500	3.6209	5.0874	4.3377	4.8351	5.2806	63.6620
1000	3.6777	4.1932	4.6238	4.8790	4.6170	31.0570
2000	3.5836	4.1786	4.2579	4.4773	5.2834	26.0690
3000	3.4640	4.0036	4.2487	4.5058	5.1112	34.3427

set II						
T	*b*					
	QML	EIG	EI	GMM	SMM	II
500	5.6704	302.43	6.6339	7.1301	7.0027	6.4383
1000	5.2873	31.69	6.7363	7.1621	7.3463	7.7470
2000	5.4264	5.9182	6.1352	6.4619	6.5078	11.9585
3000	5.2865	5.7946	5.7885	6.0650	5.9193	13.5154
T	*β*					
	QML	EIG	EI	GMM	SMM	II
500	0.4986	17.99	0.8456	0.8548	0.8527	0.4139
1000	0.3085	5.9503	0.8221	0.8018	0.8212	0.2368
2000	0.3276	2.2619	0.8423	0.8425	0.8352	0.0554
3000	0.1514	2.2685	0.7724	0.7632	0.7479	0.0424
T	*c*					
	QML	EIG	EI	GMM	SMM	II
500	3.1459	185.12	3.9844	4.5354	3.8282	3.9735
1000	2.2798	21.12	4.2944	4.6027	3.9695	5.4284
2000	2.2827	3.6963	3.9792	4.3394	3.8796	4.7032
3000	1.6182	3.7815	3.9571	4.3110	3.7541	4.8279

Table 5.5. Non standardized skewness and kurtosis

set I												
T	*b*											
	QML		EIG		EI		GMM		SMM		II	
	skew.	kurt.	skew.	kurt.	skew.	kurt.	skew.	kurt.	skew.	kurt.	skew.	kurt.
500	0.3465	3.5802	-0.7566	7.3796	0.2147	2.8252	0.3719	3.0200	0.3748	3.2863	2.8427	20.6607
1000	0.2548	3.1806	0.0830	4.1183	0.2647	2.9780	0.2480	3.2413	0.2895	3.3721	0.5174	4.6998
2000	0.0715	3.2265	0.1106	3.1990	0.0713	3.1637	0.0476	3.0739	0.1538	3.1416	0.3329	4.3811
3000	0.1448	3.2234	0.1616	3.2545	0.0152	2.9392	0.0838	3.1394	0.1344	3.0809	-0.0973	4.8069
T	β											
	QML		EIG		EI		GMM		SMM		II	
	skew.	kurt.	skew.	kurt.	skew.	kurt.	skew.	kurt.	skew.	kurt.	skew.	kurt.
500	0.0624	3.5140	-0.5371	10.2772	0.1139	2.9487	0.0242	2.7876	-0.0881	3.0596	2.2202	14.8308
1000	0.0256	3.2220	0.4258	4.1006	0.0704	3.4036	0.1865	3.4670	0.1060	3.3814	0.7026	3.5598
2000	-0.0086	2.8767	0.1240	3.0830	-0.0722	2.8219	-0.0905	2.7367	-0.0229	2.9750	0.8557	4.2123
3000	0.1393	3.1348	0.1444	3.1173	0.0402	2.9953	0.0168	2.9524	0.0057	2.7883	0.6320	3.1607
T	*c*											
	QML		EIG		EI		GMM		SMM		II	
	skew.	kurt.	skew.	kurt.	skew.	kurt.	skew.	kurt.	skew.	kurt.	skew.	kurt.
500	0.2598	3.2251	-1.3701	17.1722	0.2253	3.0160	0.3400	3.2727	0.4091	3.8291	2.6233	14.7022
1000	0.2400	3.1541	0.2048	3.1322	0.2019	2.9381	0.2271	3.1478	0.1873	3.2272	0.7470	3.7854
2000	0.1755	3.0187	0.1670	3.0708	0.2253	3.0725	0.1243	2.8337	0.3796	3.3155	0.8552	3.3384
3000	0.1728	3.0097	0.1483	2.9511	0.2059	2.9493	0.2345	2.9351	0.2135	2.9556	0.3105	3.0196

Table 5.6. Non standardized skewness and kurtosis

set II												
T	*b*											
	QML		EIG		EI		GMM		SMM		II	
	skew.	kurt.	skew.	kurt.	skew.	kurt.	skew.	kurt.	skew.	kurt.	skew.	kurt.
500	0.4521	3.6627	11.3972	172.65	0.2553	3.0948	0.3269	3.4787	0.4191	3.6903	7.6061	63.4115
1000	0.1820	3.4299	10.0832	136.91	0.2376	2.9473	0.4165	3.4960	0.4181	3.5107	8.1775	73.7320
2000	0.1261	2.9229	0.3255	4.3387	0.0570	2.9709	0.0698	3.0920	0.1644	3.1731	-1.0319	7.5950
3000	0.1858	2.9212	0.2205	2.9731	0.0519	2.8956	0.1461	3.0909	0.1735	3.2699	-0.9570	6.6982
T	*β*											
	QML		EIG		EI		GMM		SMM		II	
	skew.	kurt.	skew.	kurt.	skew.	kurt.	skew.	kurt.	skew.	kurt.	skew.	kurt.
500	0.4204	6.6211	-1.7893	16.0653	0.1078	3.1191	0.0986	3.2682	0.0867	3.3120	7.6484	190.3159
1000	-0.1736	4.3923	-6.3475	89.0878	-0.0021	3.3494	0.0993	3.2597	0.0661	3.2407	-12.4716	210.0755
2000	-0.0502	4.5017	0.0046	3.1486	-0.0558	2.7132	-0.0366	2.7267	-0.0452	2.7383	-5.6920	50.6449
3000	-0.0395	4.4534	0.1028	3.0807	0.0742	3.1357	0.0270	3.0358	0.0187	3.1030	-0.7343	7.7828
T	*c*											
	QML		EIG		EI		GMM		SMM		II	
	skew.	kurt.	skew.	kurt.	skew.	kurt.	skew.	kurt.	skew.	kurt.	skew.	kurt.
500	-0.7187	10.7779	20.1151	469.86	0.1990	2.8839	0.2317	2.9061	0.2306	3.1285	19.2643	508.4167
1000	0.1539	3.2236	10.1673	210.49	0.1741	2.8828	0.2384	3.0355	0.2166	2.9532	9.0121	85.8473
2000	0.1431	2.9885	0.3269	4.0870	0.1970	3.0407	0.2302	3.0878	0.2668	3.0564	0.5013	4.0201
3000	0.0600	3.8633	0.1566	3.0003	0.1786	3.0355	0.1857	3.0330	0.1578	2.9759	-0.0347	4.5803

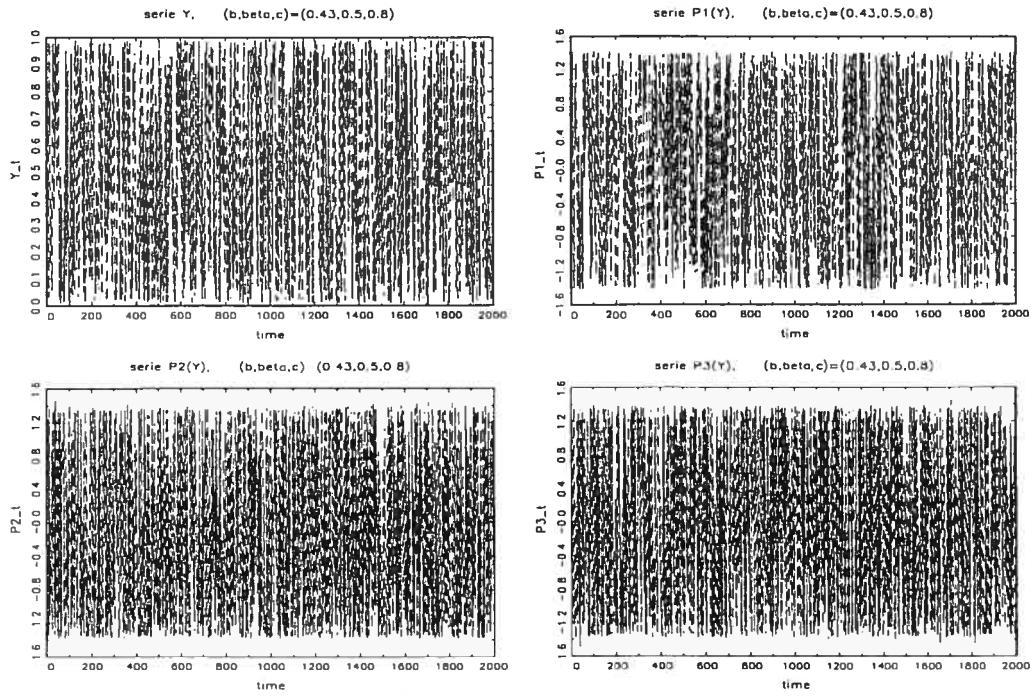


Figure 5.0: Simulated paths, set I.

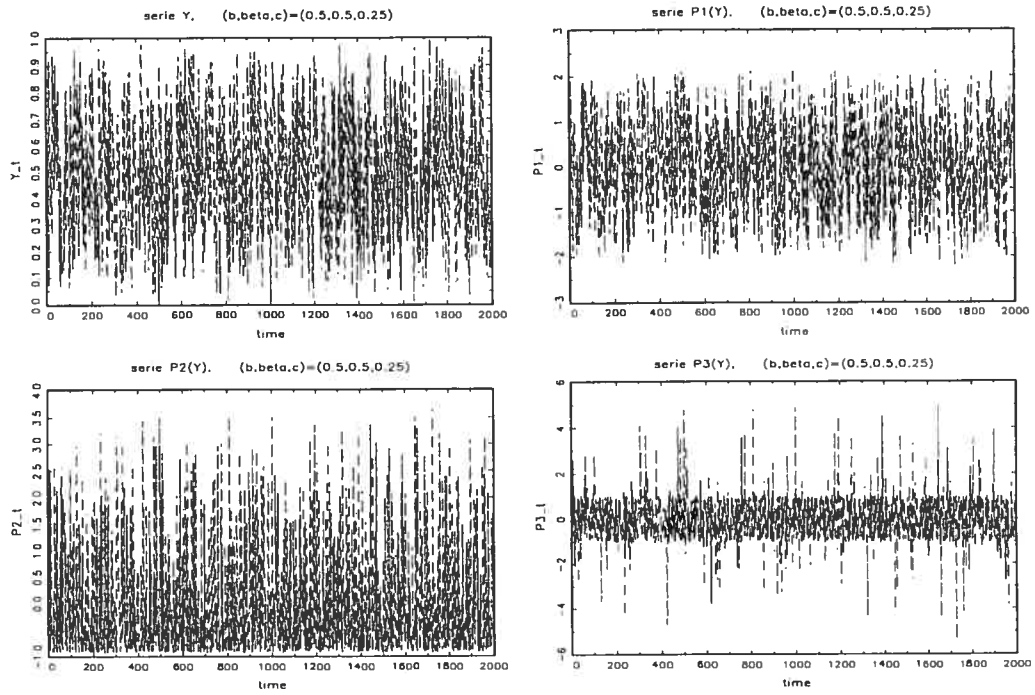


Figure 5.1: Simulated paths, set II.

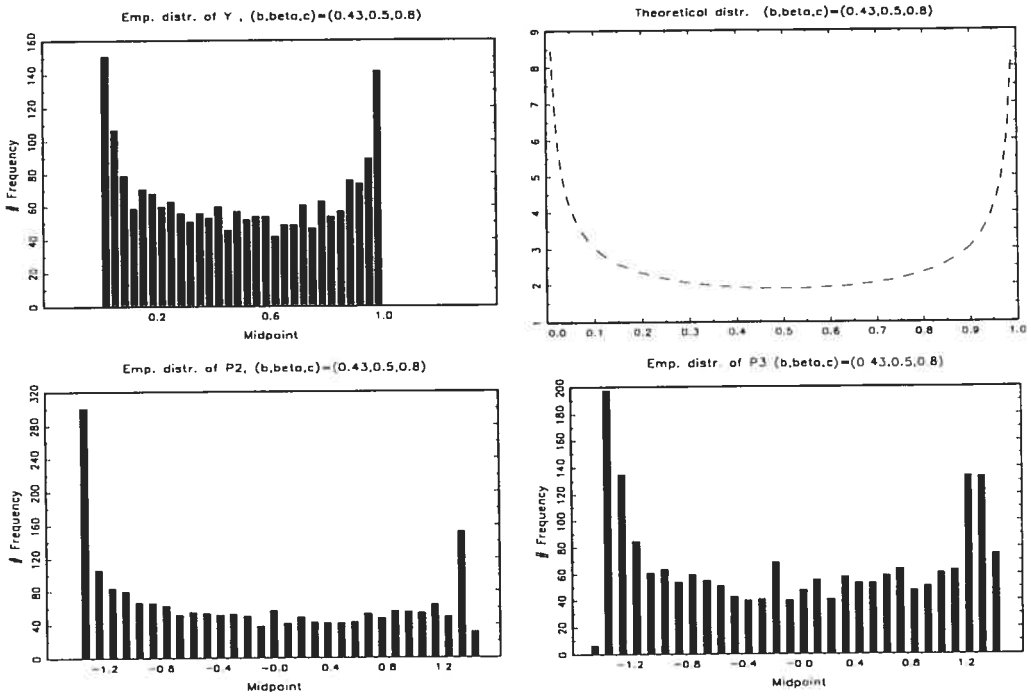


Figure 5.2: Empirical marginal distributions, set I.

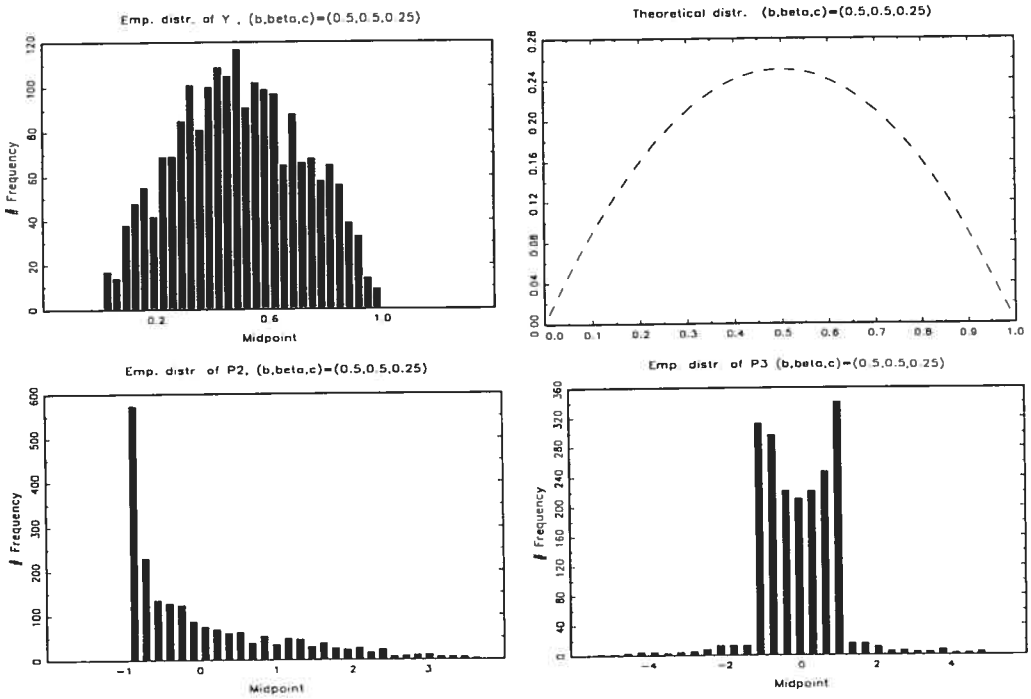


Figure 5.3: Empirical marginal distributions, set II.

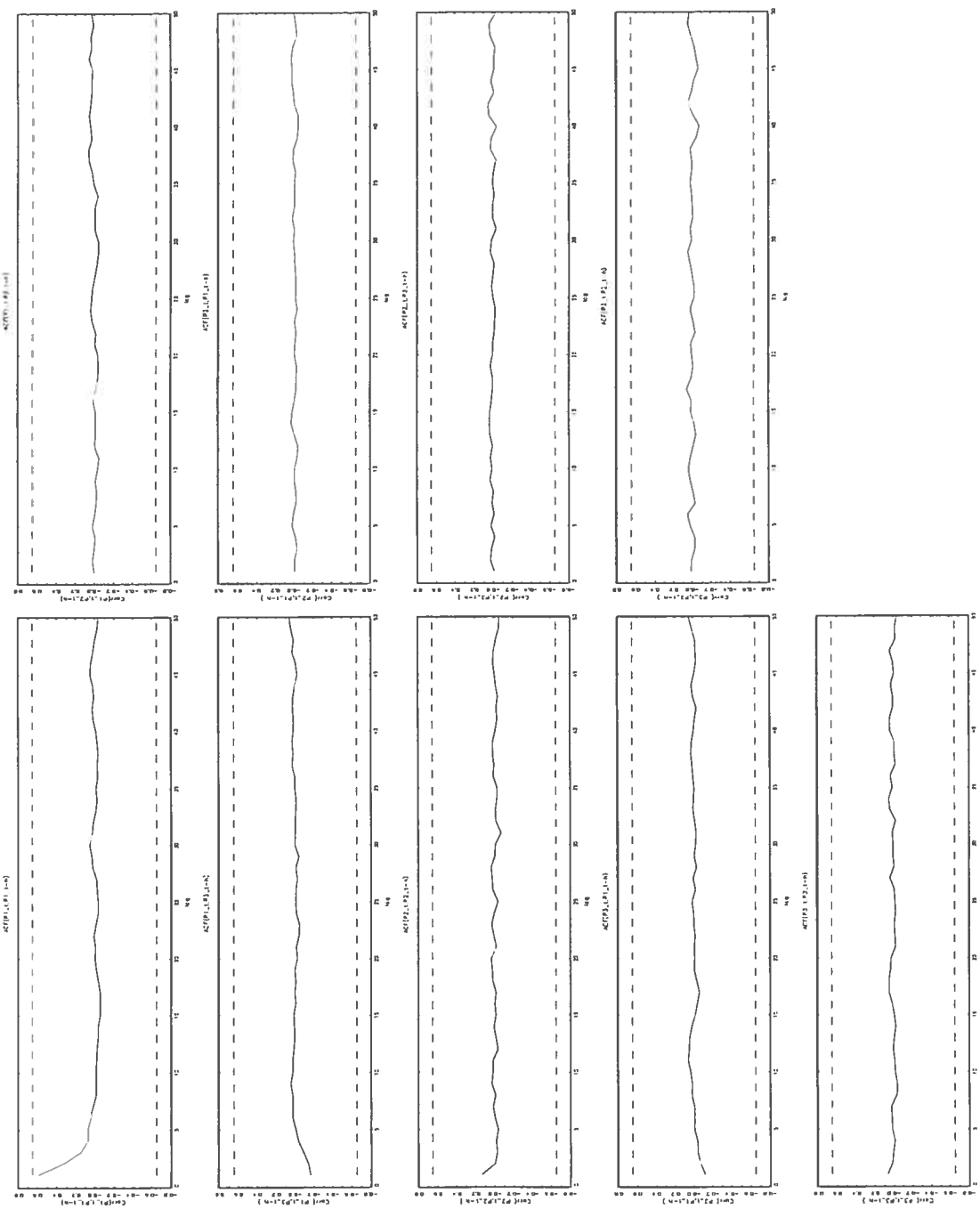


Figure 5.4: Empirical correlations, set I.

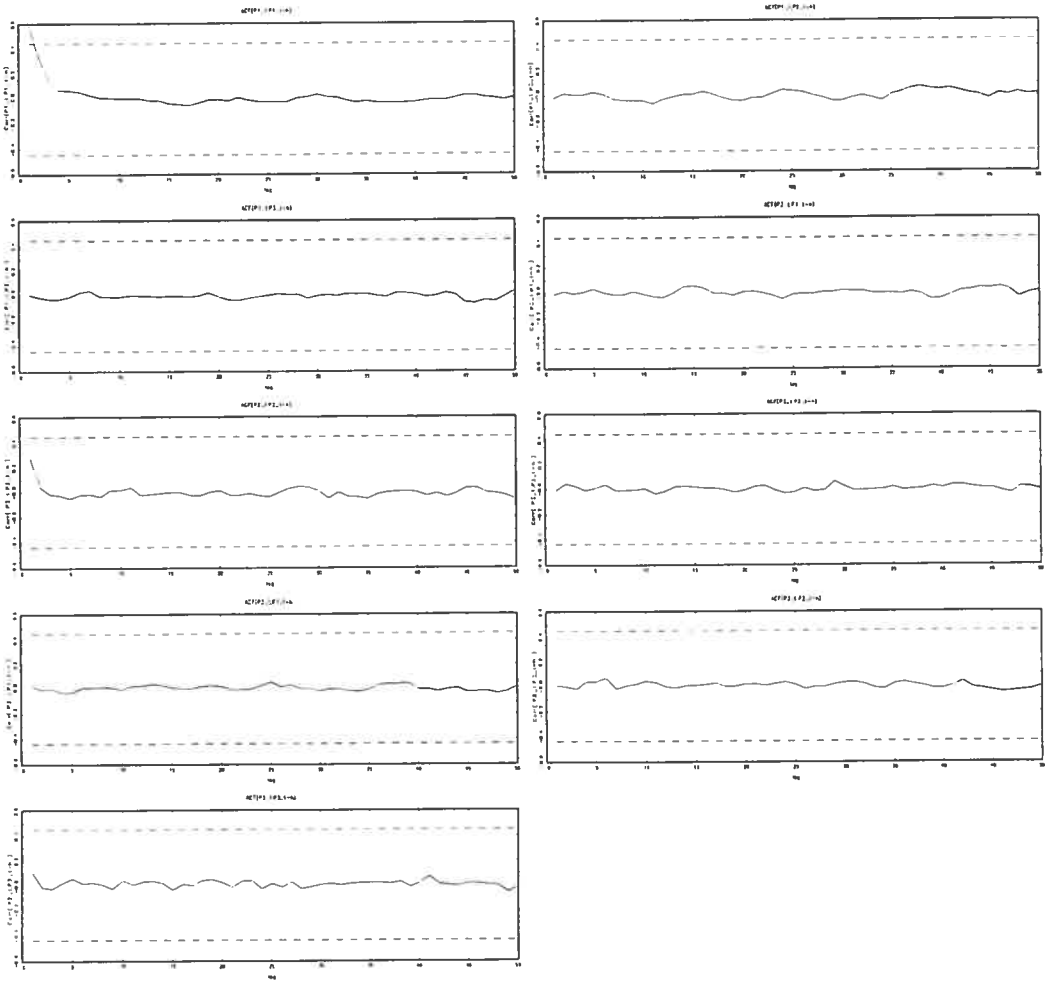


Figure 5.5: Cross autocorrelograms, set II.

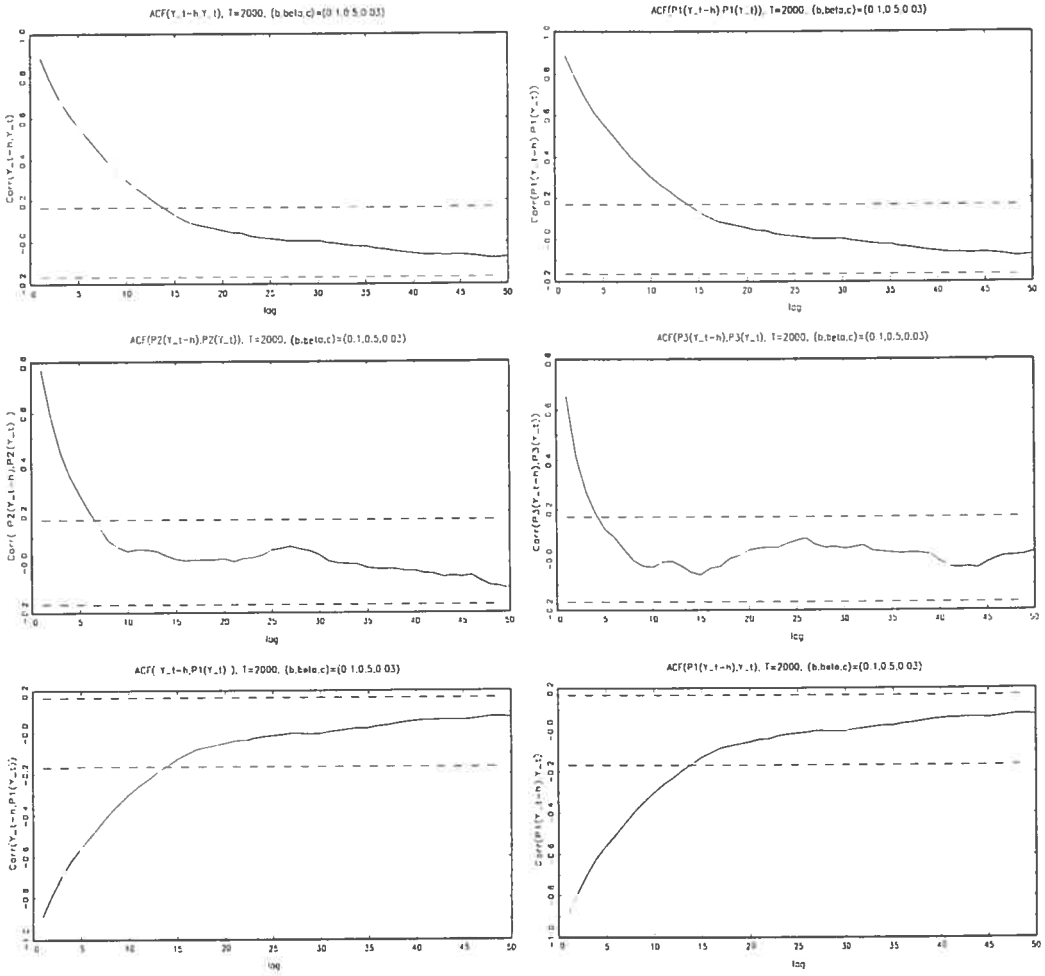


Figure 5.6: Cross autocorrelograms, set $(0.1, 0.5, 0.03)$.

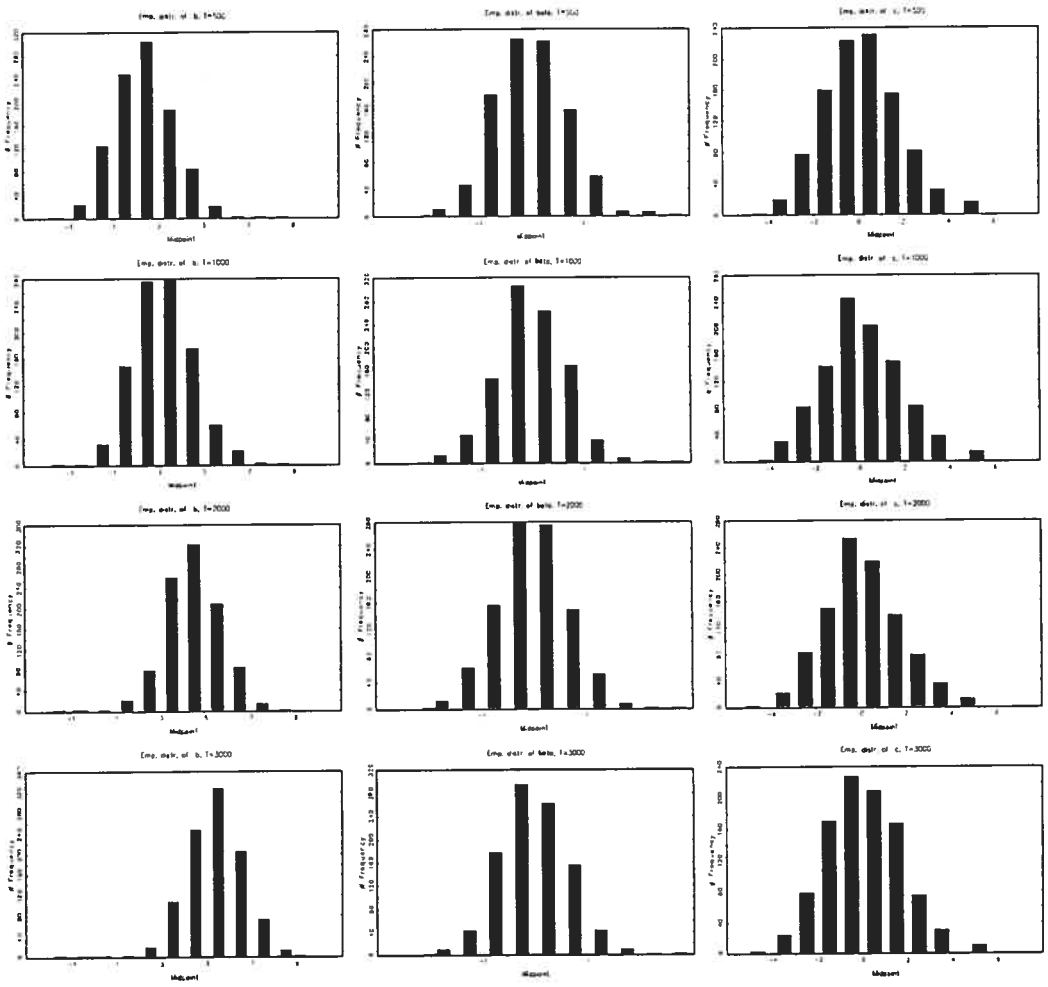


Figure 5.7: Standardized marginal sample distribution of EI, set I: (parameter per column, size per row:500,1000,2000,3000).

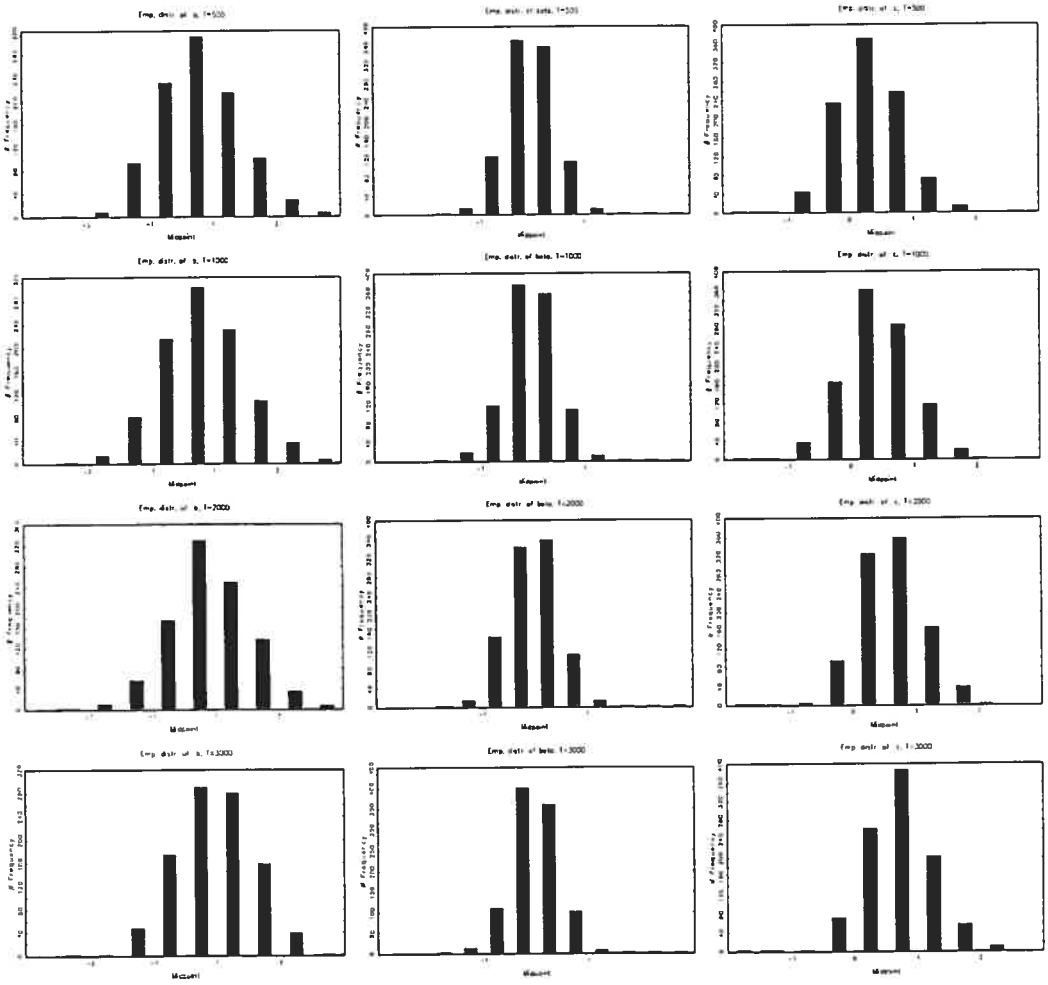


Figure 5.8: Standardized marginal sample distribution of EI, set II: (parameter per column, size per row:500,1000,2000,3000).

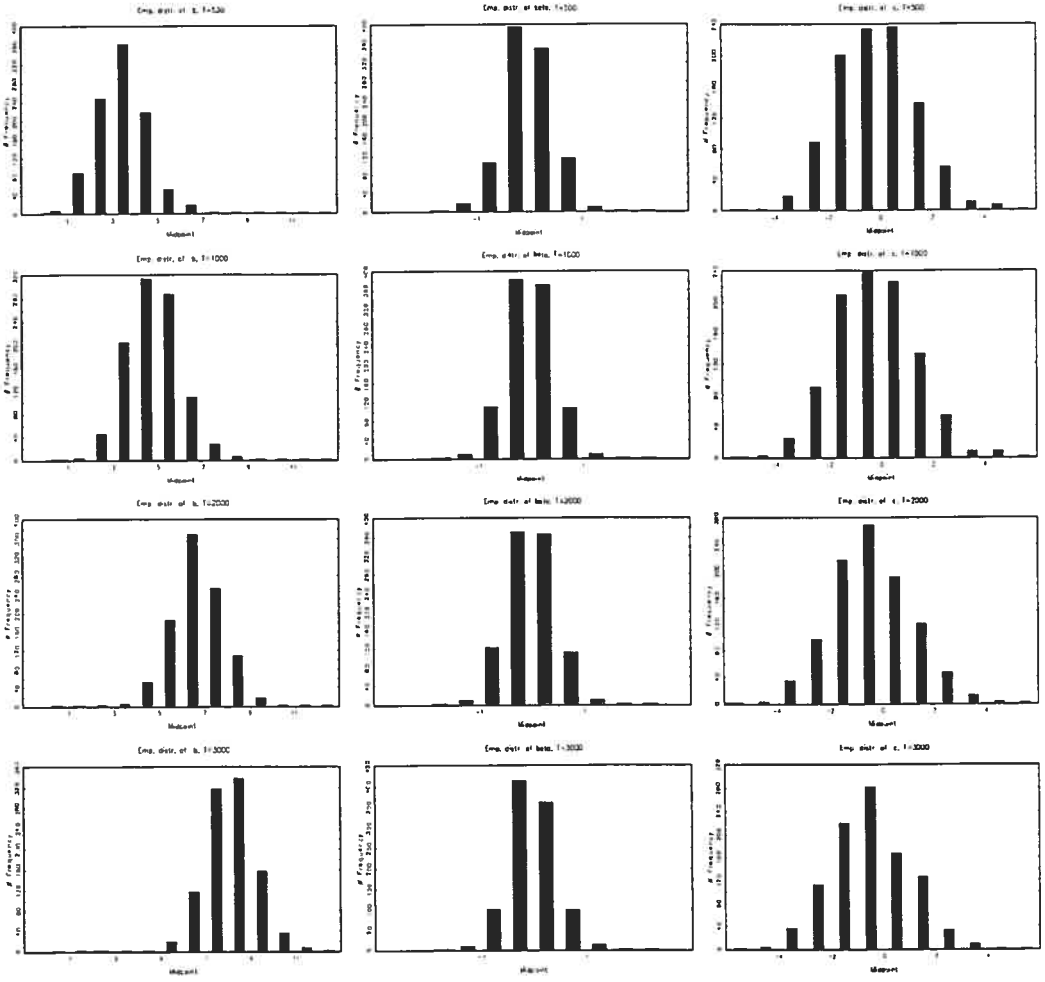


Figure 5.9: Standardized marginal sample distribution of QML, set I: (parameter per column, size per row:500,1000,2000,3000).

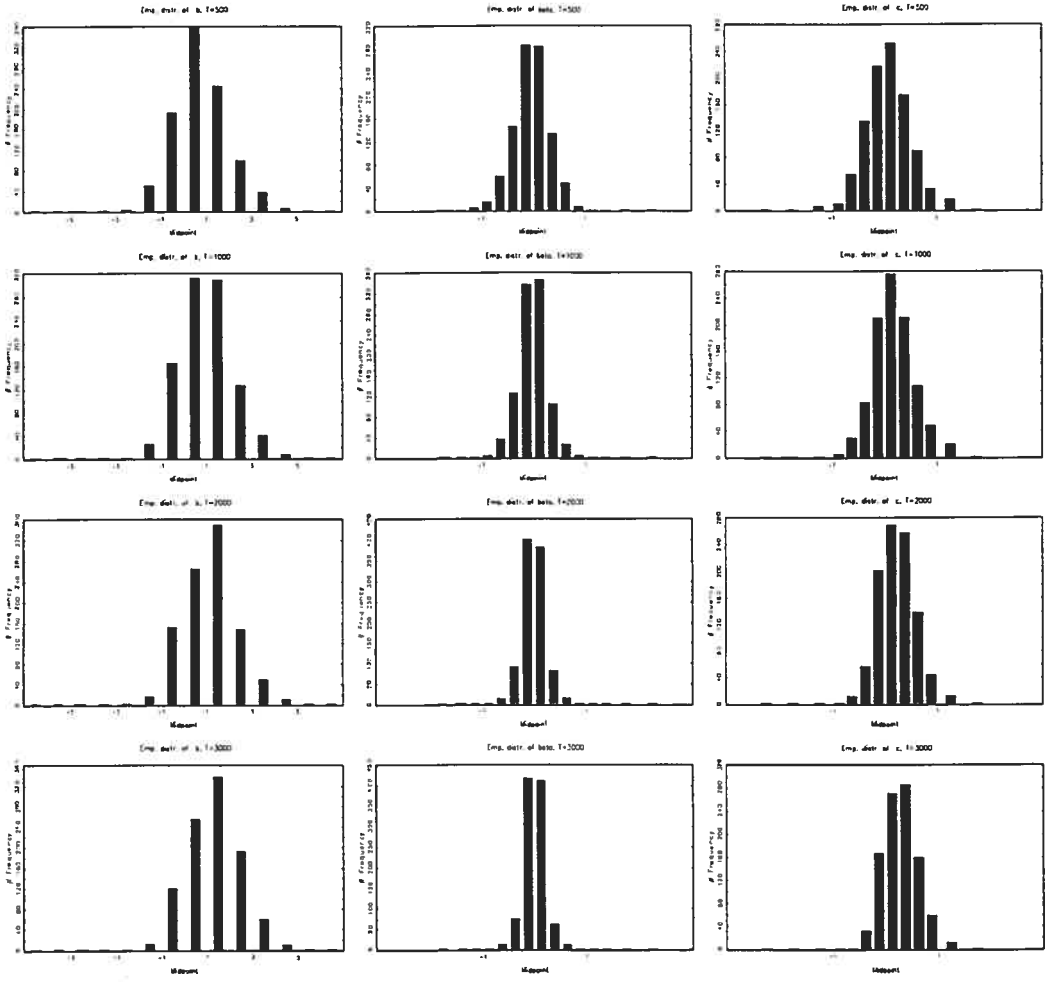


Figure 5.10: Standardized marginal sample distribution of QML, set II: (parameter per column, size per row:500,1000,2000,3000).

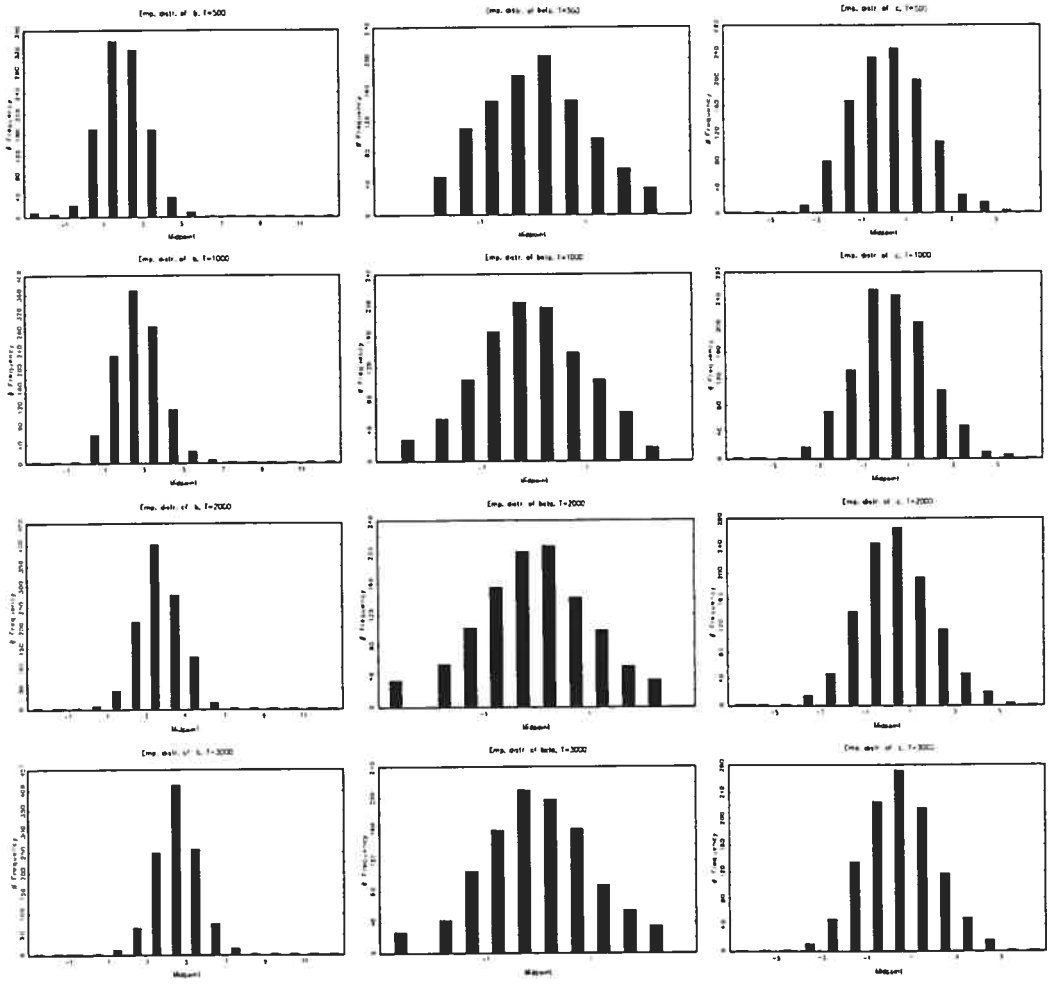


Figure 5.11: Standardized marginal sample distribution of EIG, set I: (parameter per column, size per row:500,1000,2000,3000).

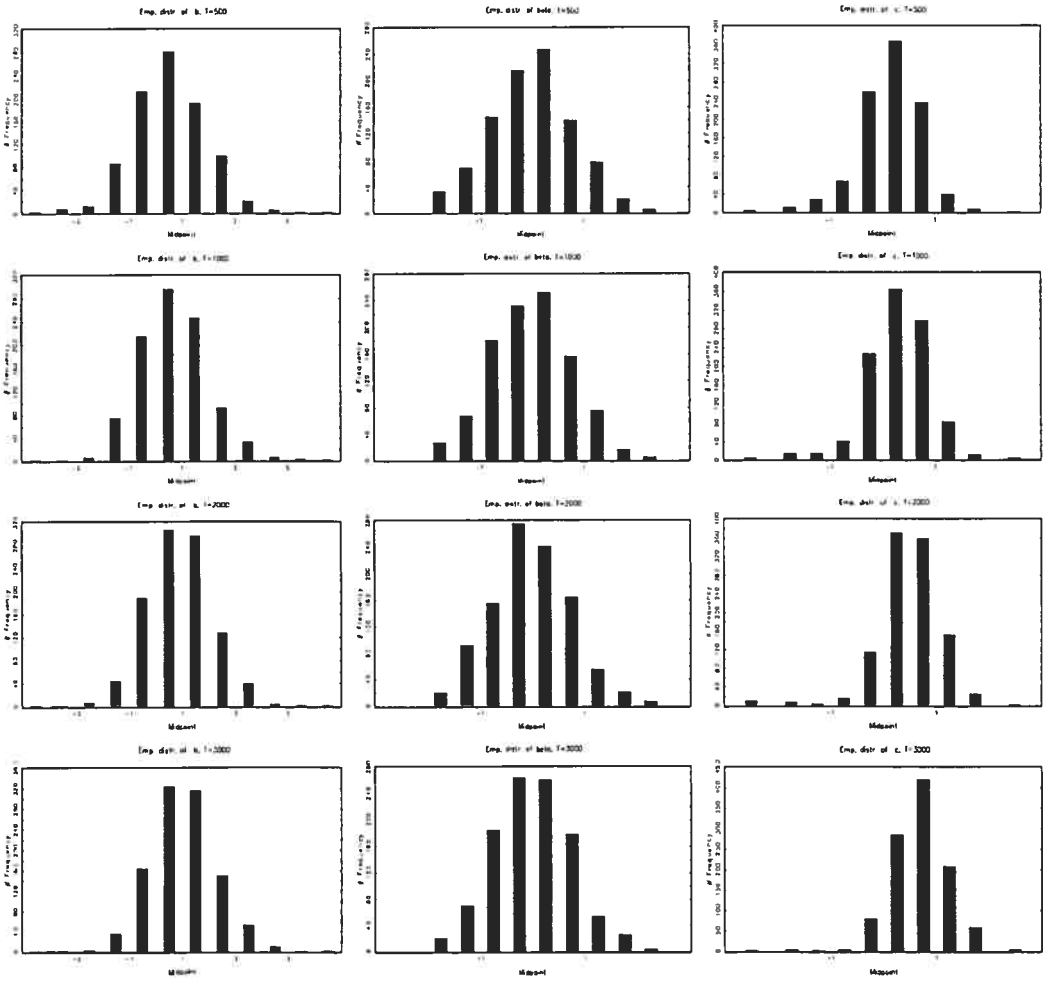


Figure 5.12: Standardized marginal sample distribution of EIG, set II: (parameter per column, size per row:500,1000,2000,3000).

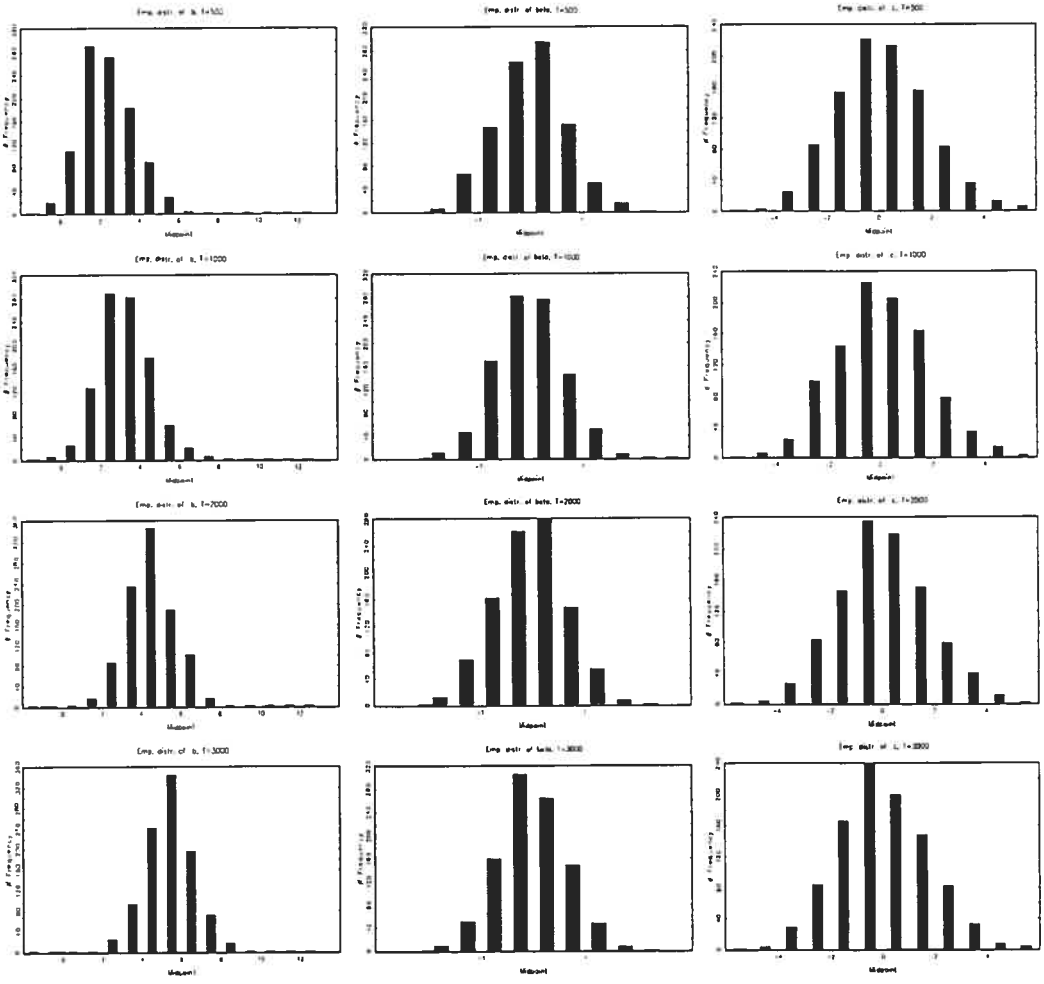


Figure 5.13: Standardized marginal sample distribution of GMM, set I: (parameter per column, size per row:500,1000,2000,3000).

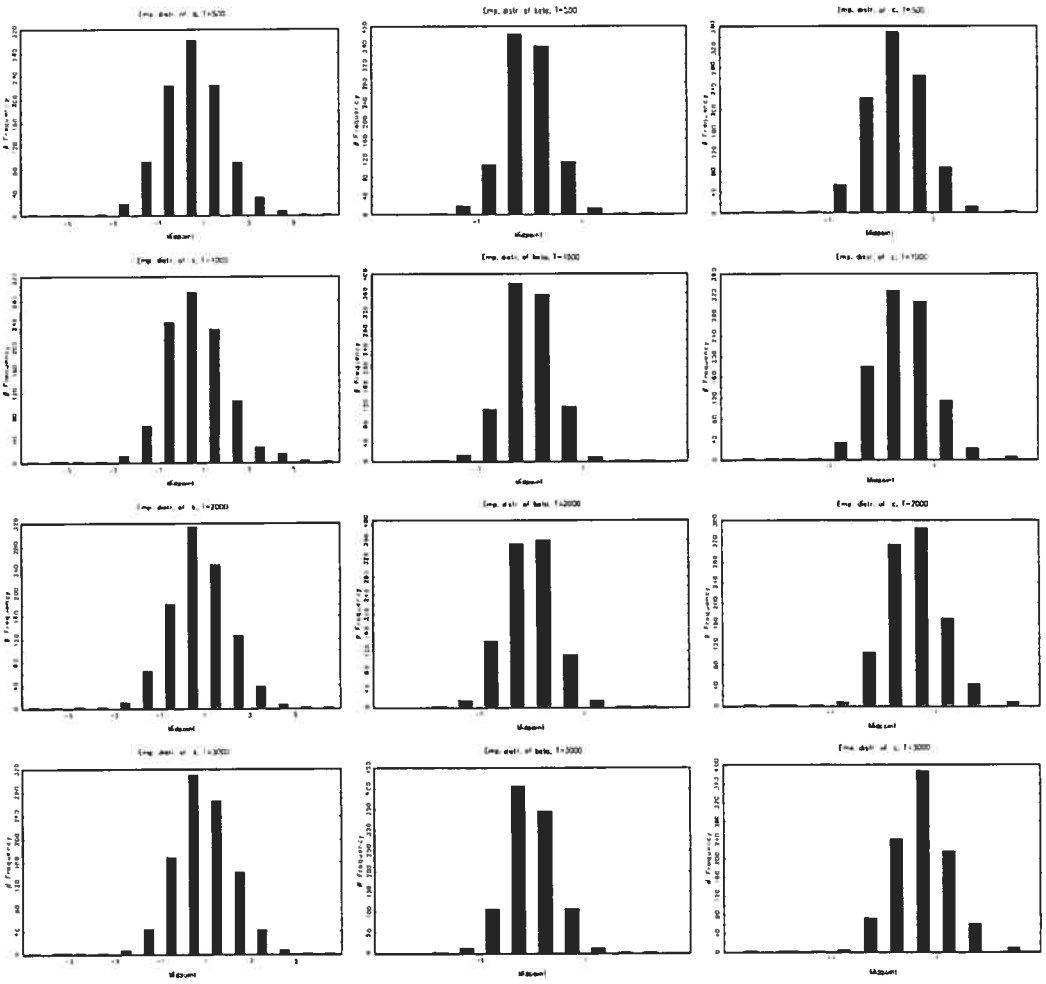


Figure 5.14: Standardized marginal sample distribution of GMM, set II: (parameter per column, size per row:500,1000,2000,3000).

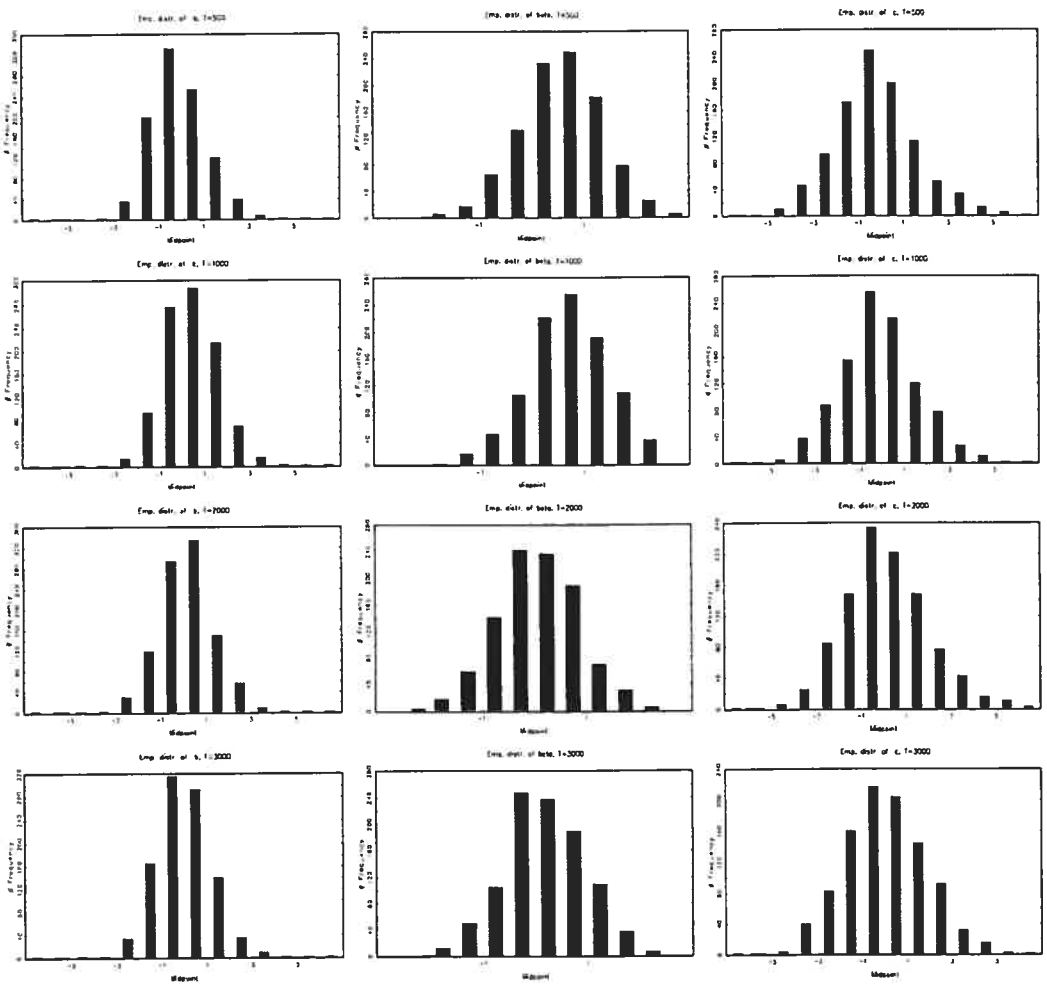


Figure 5.15: Standardized marginal sample distribution of SMM, set I: (parameter per column, size per row:500,1000,2000,3000).

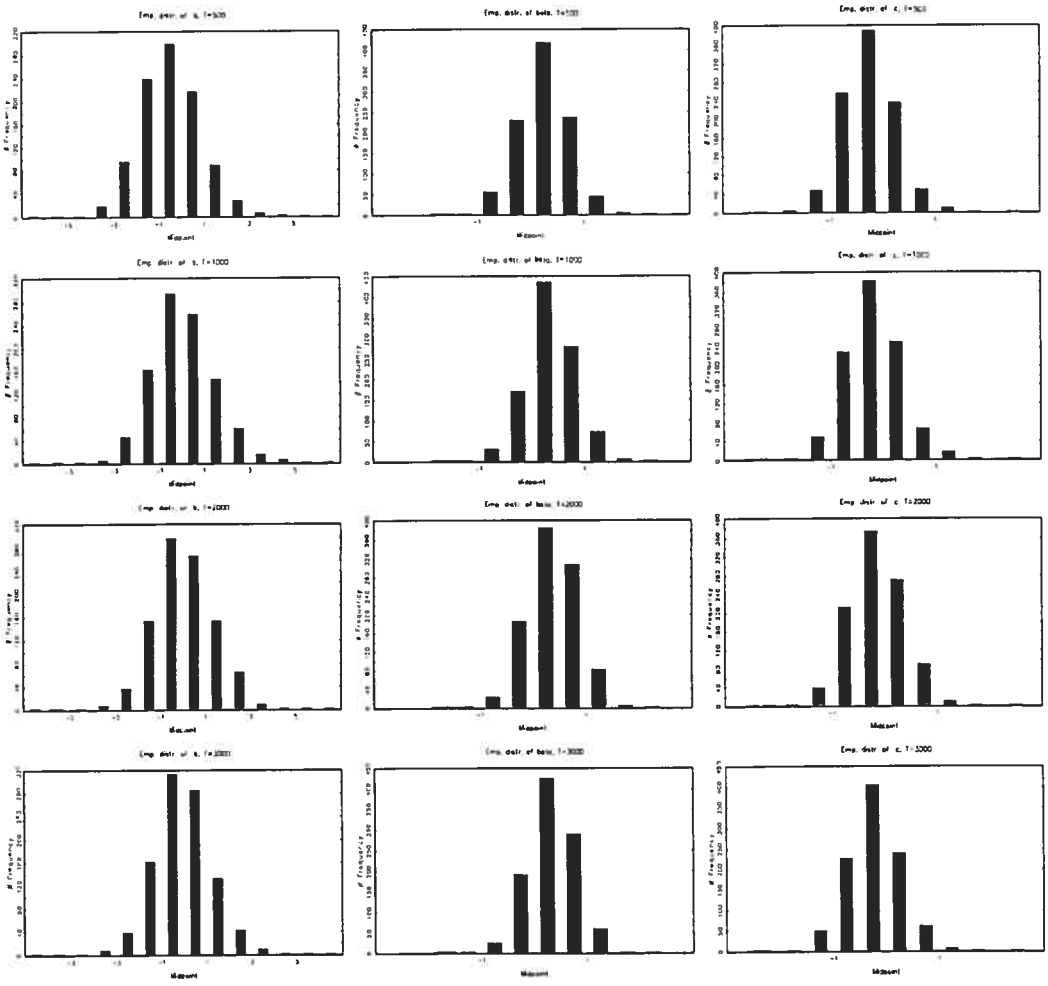


Figure 5.16: Standardized marginal sample distribution of SMM, set II: (parameter per column, size per row:500,1000,2000,3000).

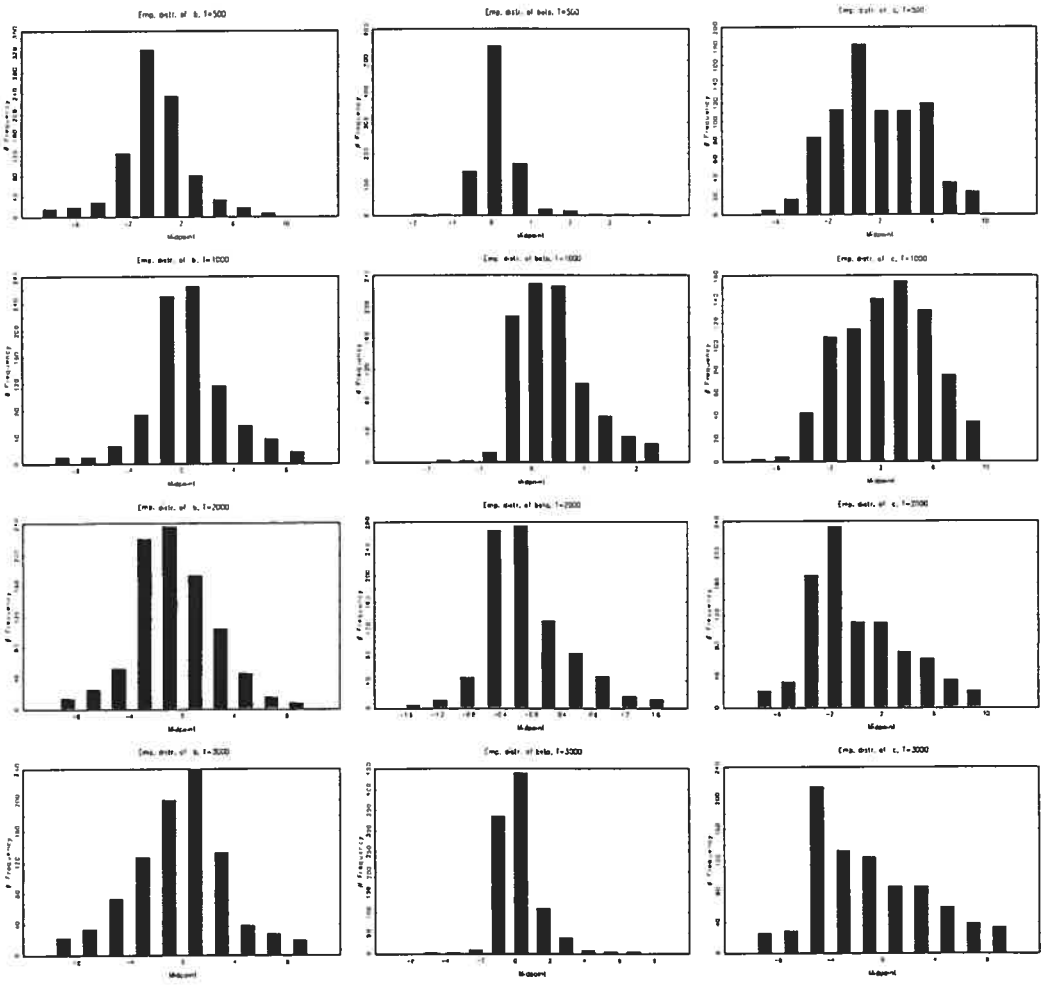


Figure 5.17: Standardized marginal sample distribution of II, set I: (parameter per column, size per row:500,1000,2000,3000).

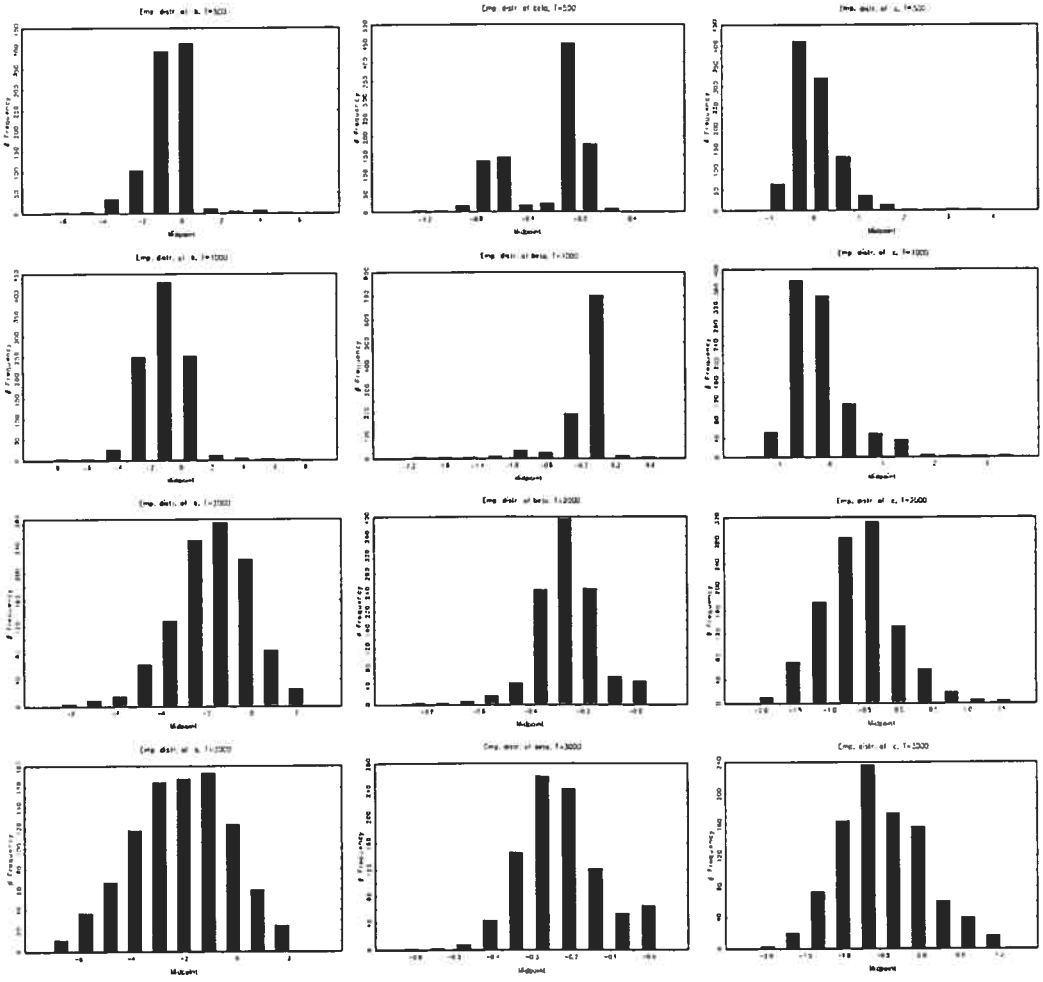


Figure 5.18: Standardized marginal sample distribution of II, set II: (parameter per column, size per row:500,1000,2000,3000).

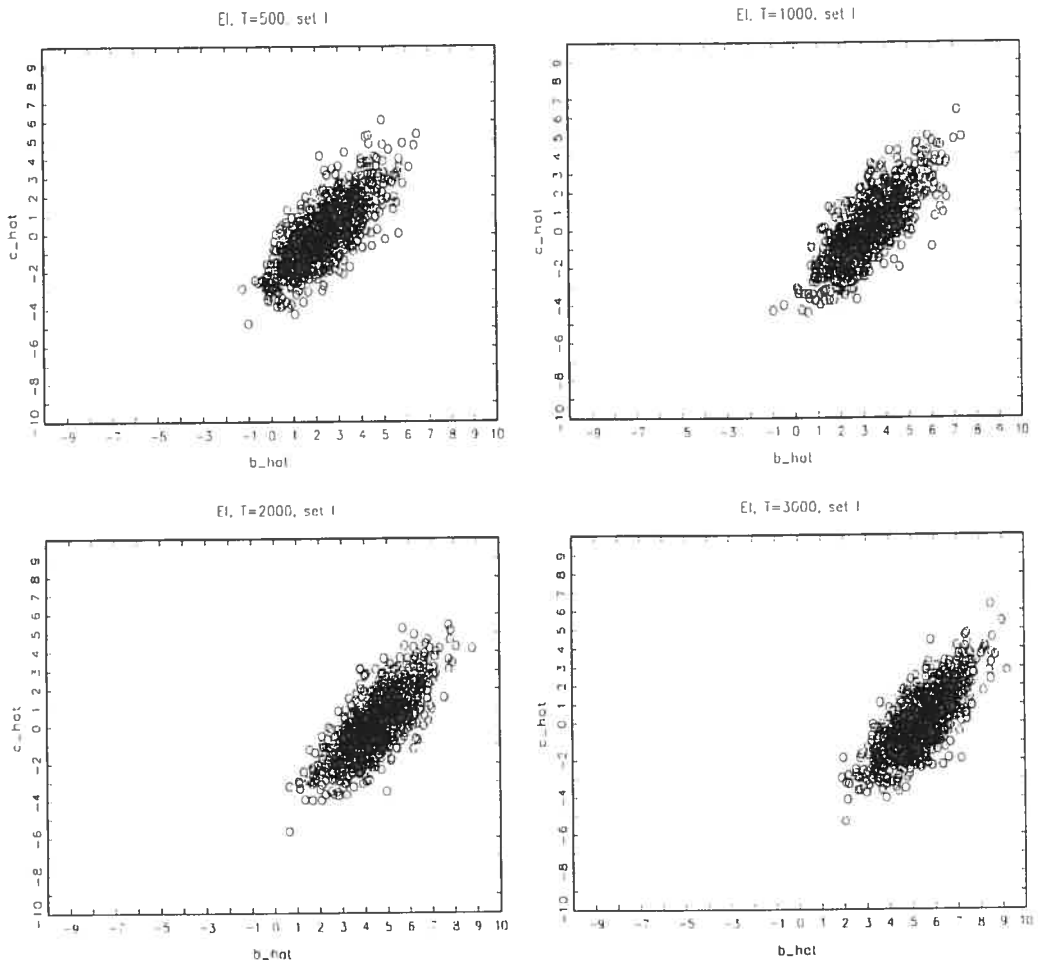


Figure 5.19: Standardized joint sample distribution of b and c, EI, set I: (parameter per column, size per row:500,1000,2000,3000).

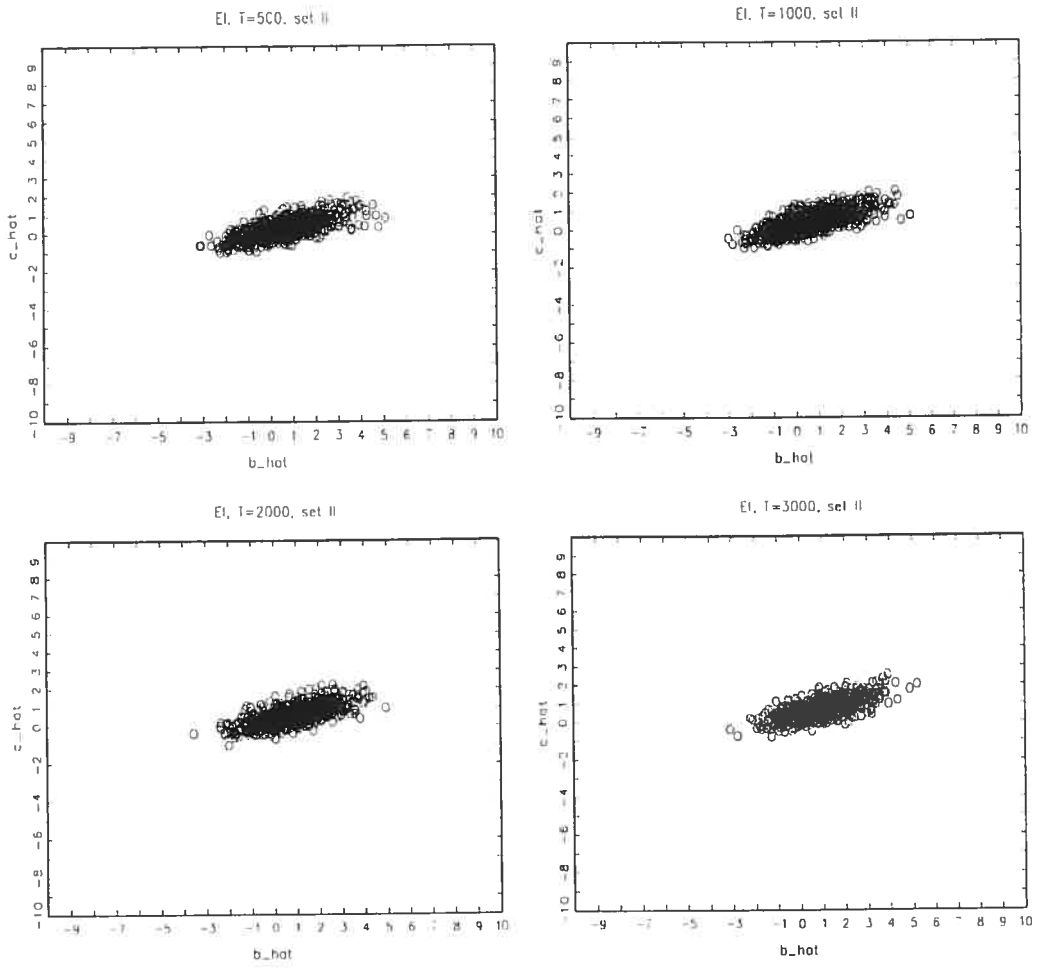


Figure 5.20: Standardized joint sample distribution of b and c, EI, set II: (parameter per column, size per row:500,1000,2000,3000).

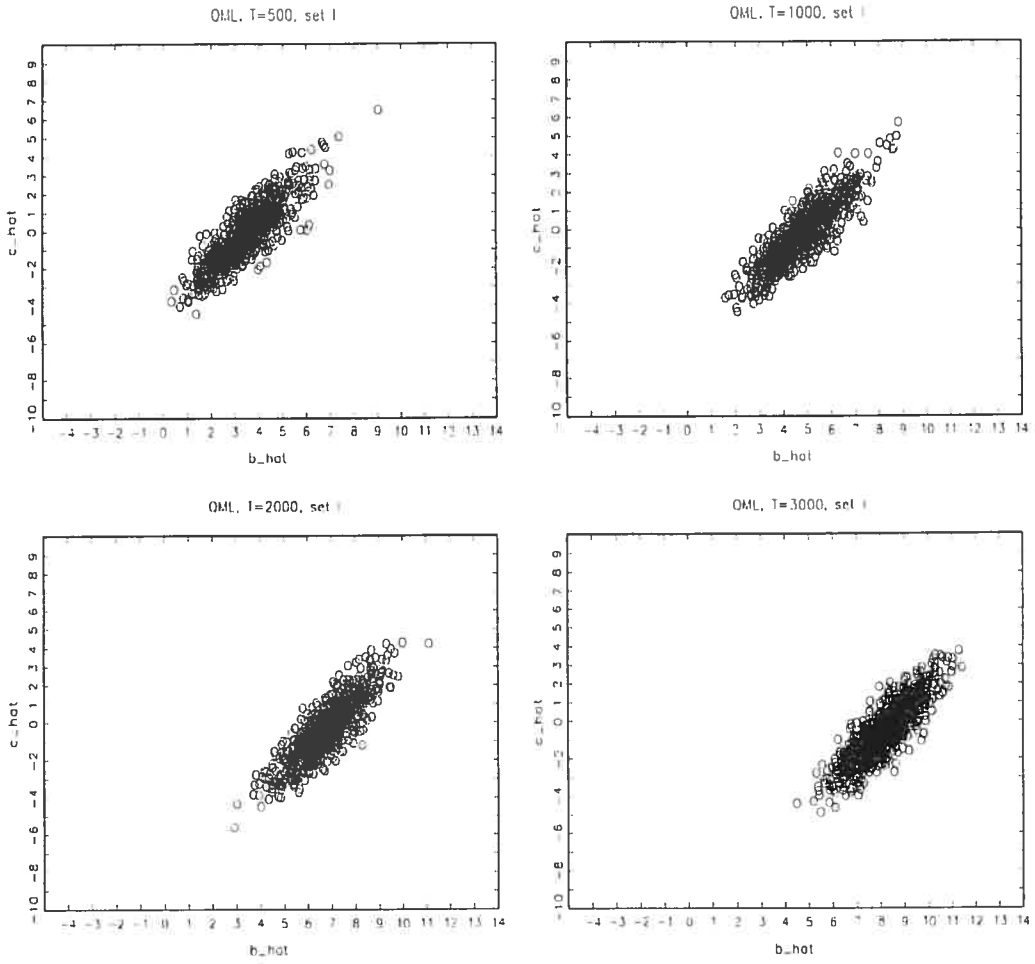


Figure 5.21: Standardized joint sample distribution of b and c, QML, set I: (parameter per column, size per row:500,1000,2000,3000).

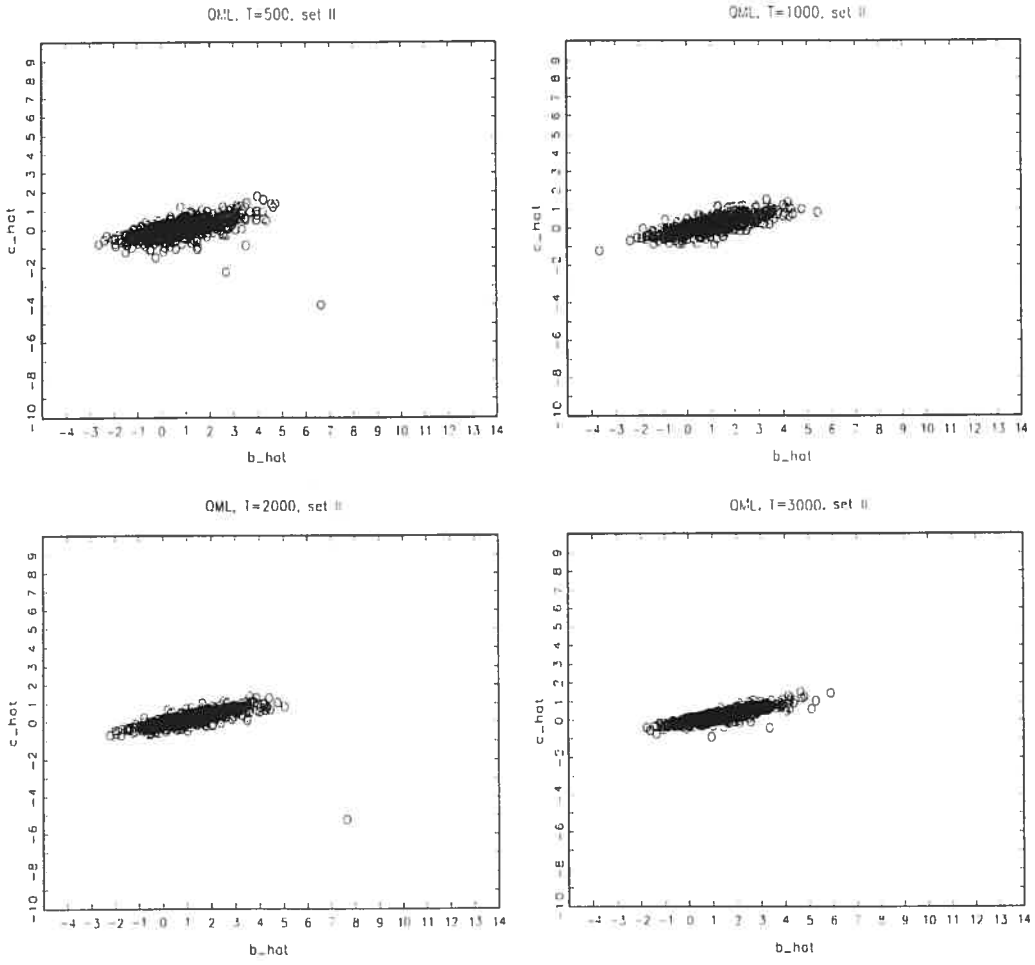


Figure 5.22: Standardized joint sample distribution of b and c, QML, set II: (parameter per column, size per row:500,1000,2000,3000).

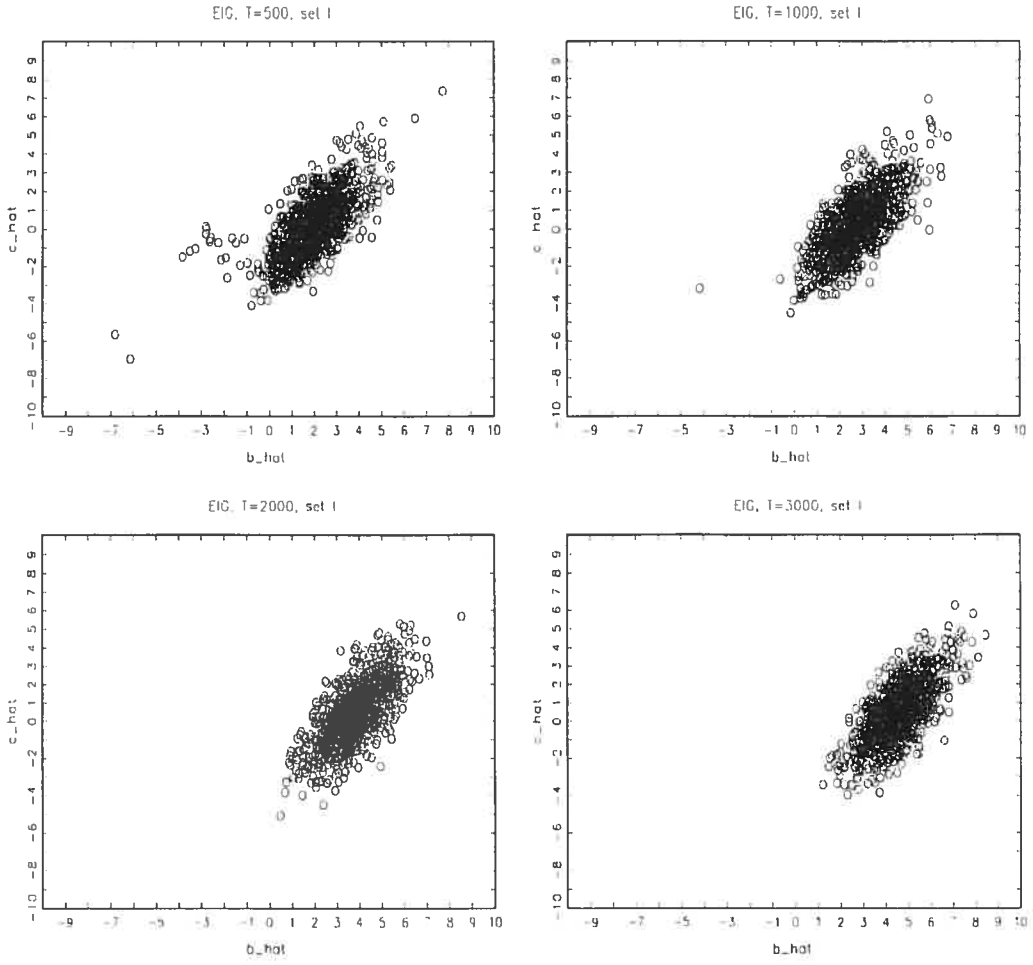


Figure 5.23: Standardized joint sample distribution of b and c, EIG, set I: (parameter per column, size per row:500,1000,2000,3000).

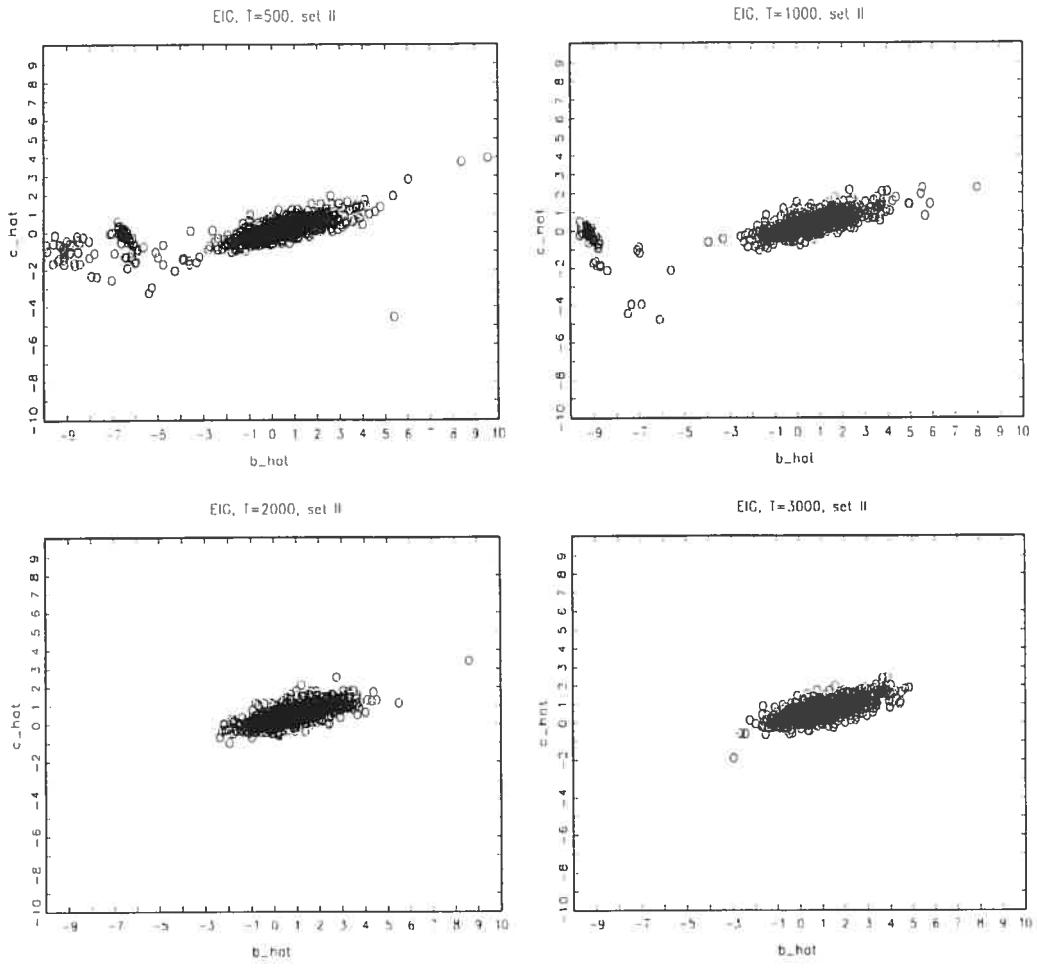


Figure 5.24: Standardized joint sample distribution of b and c, EIG, set II: (parameter per column, size per row:500,1000,2000,3000).

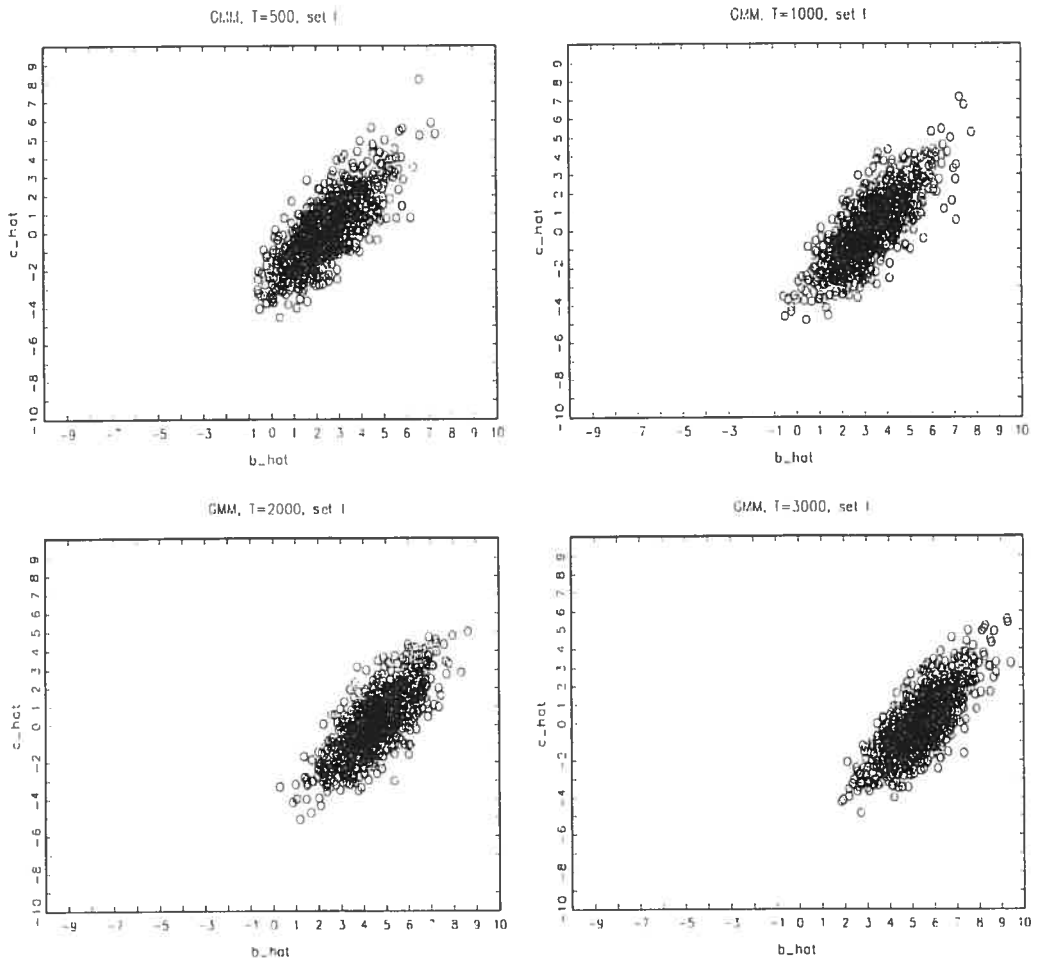


Figure 5.25: Standardized joint sample distribution of b and c, GMM, set I: (parameter per column, size per row:500,1000,2000,3000).

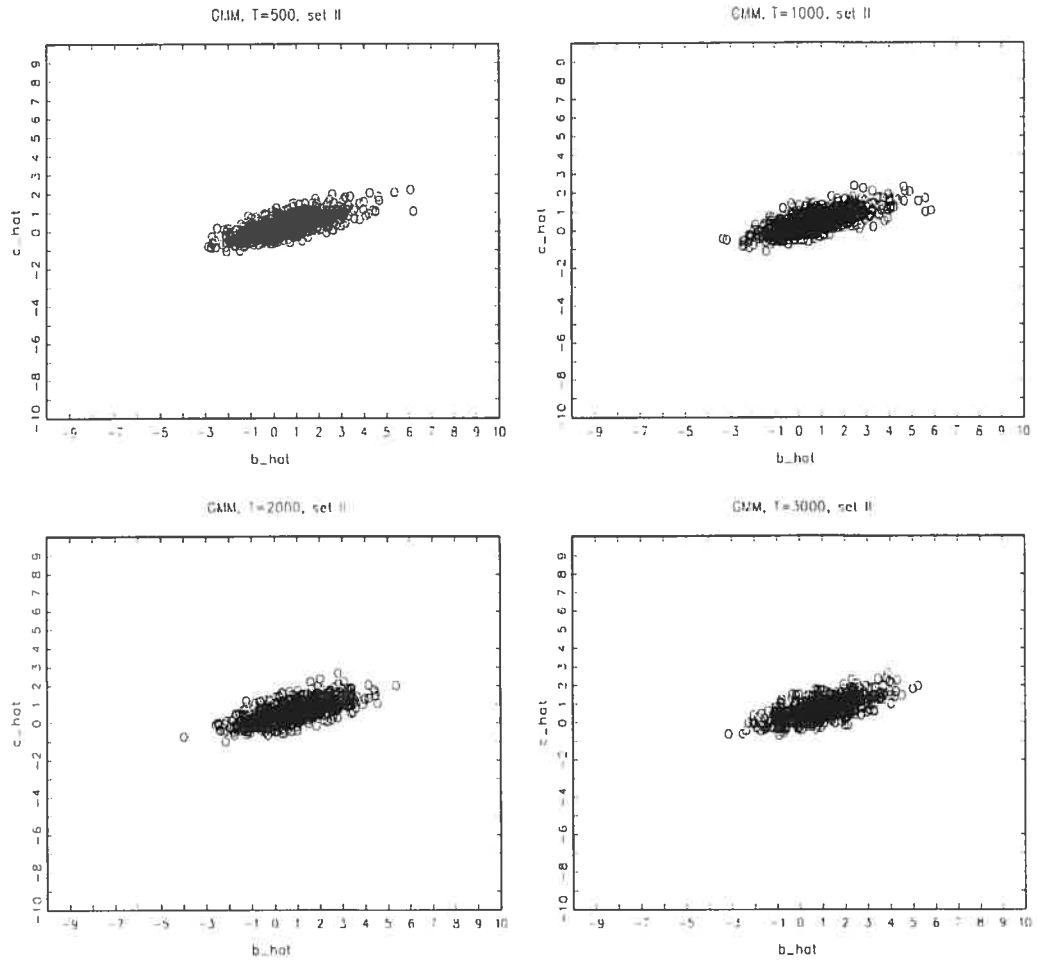


Figure 5.26: Standardized joint sample distribution of b and c, GMM, set II: (parameter per column, size per row:500,1000,2000,3000).

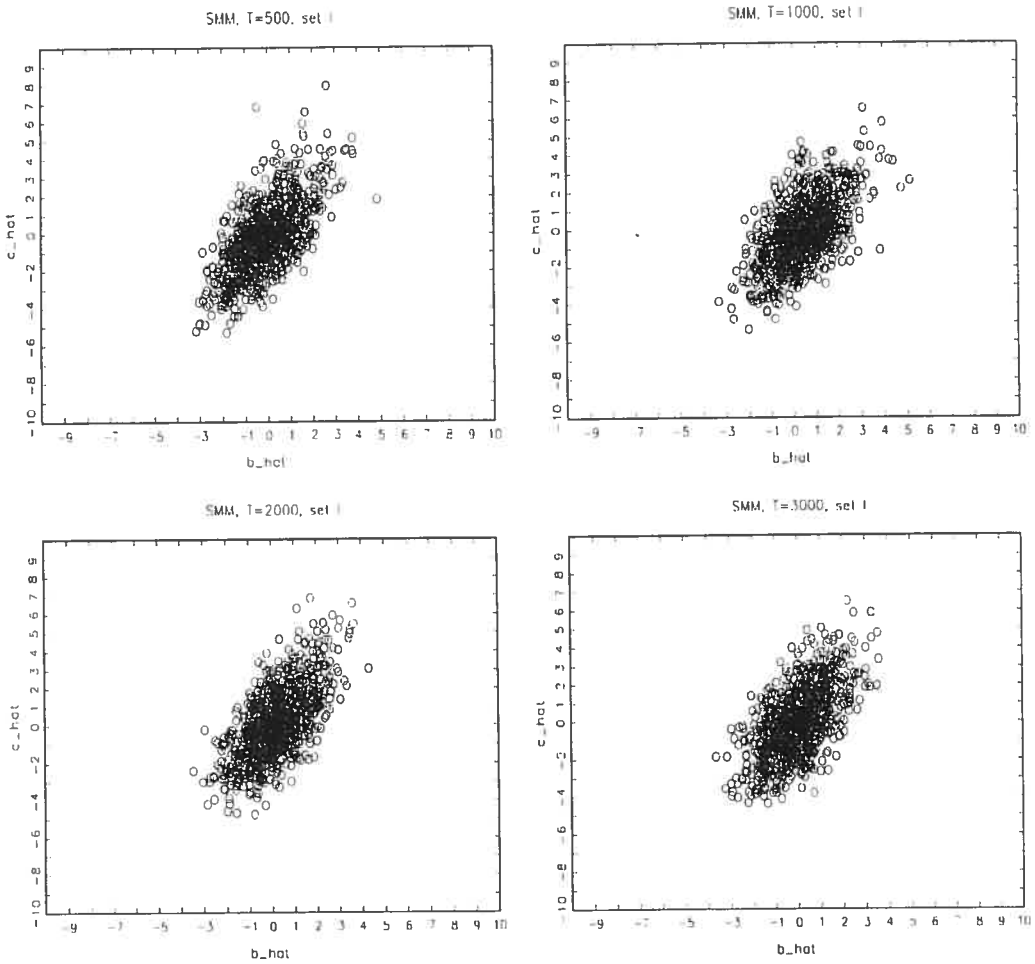


Figure 5.27: Standardized joint sample distribution of b and c, SMM, set I: (parameter per column, size per row:500,1000,2000,3000).

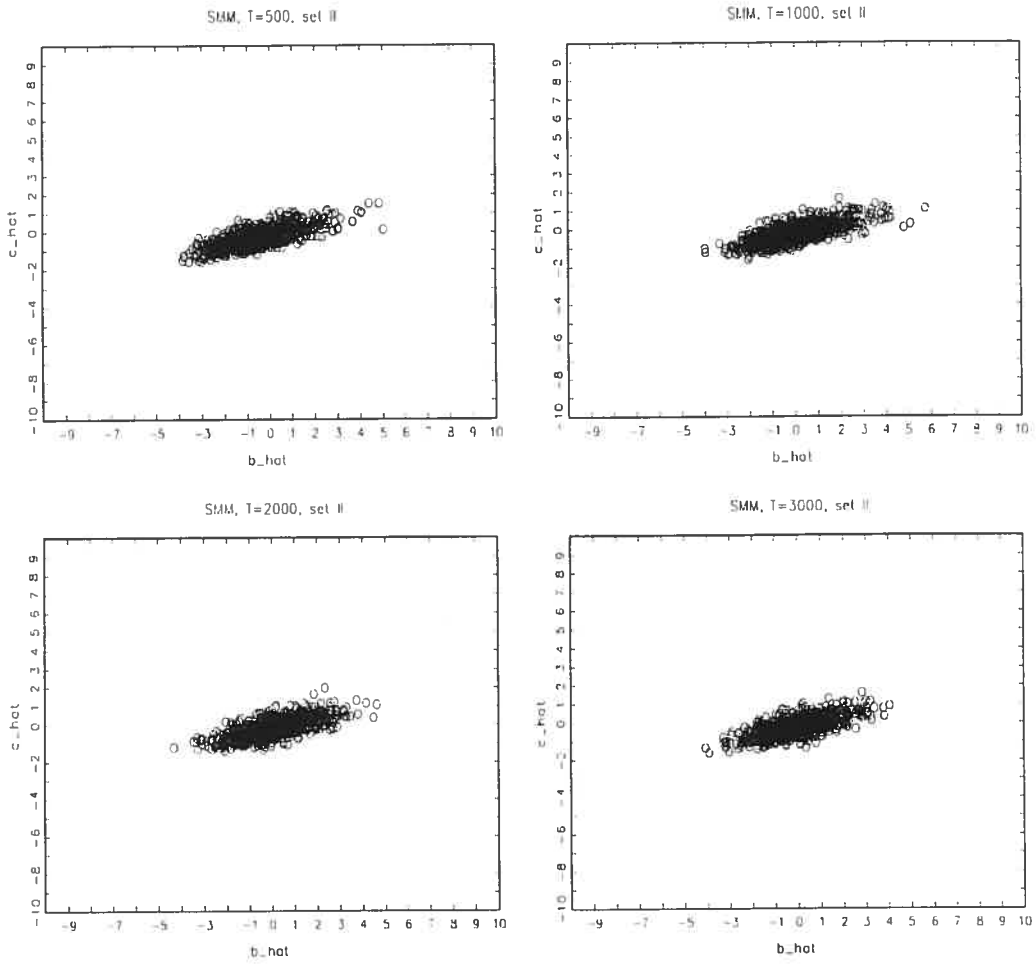


Figure 5.28: Standardized joint sample distribution of b and c, SMM, set II: (parameter per column, size per row:500,1000,2000,3000).

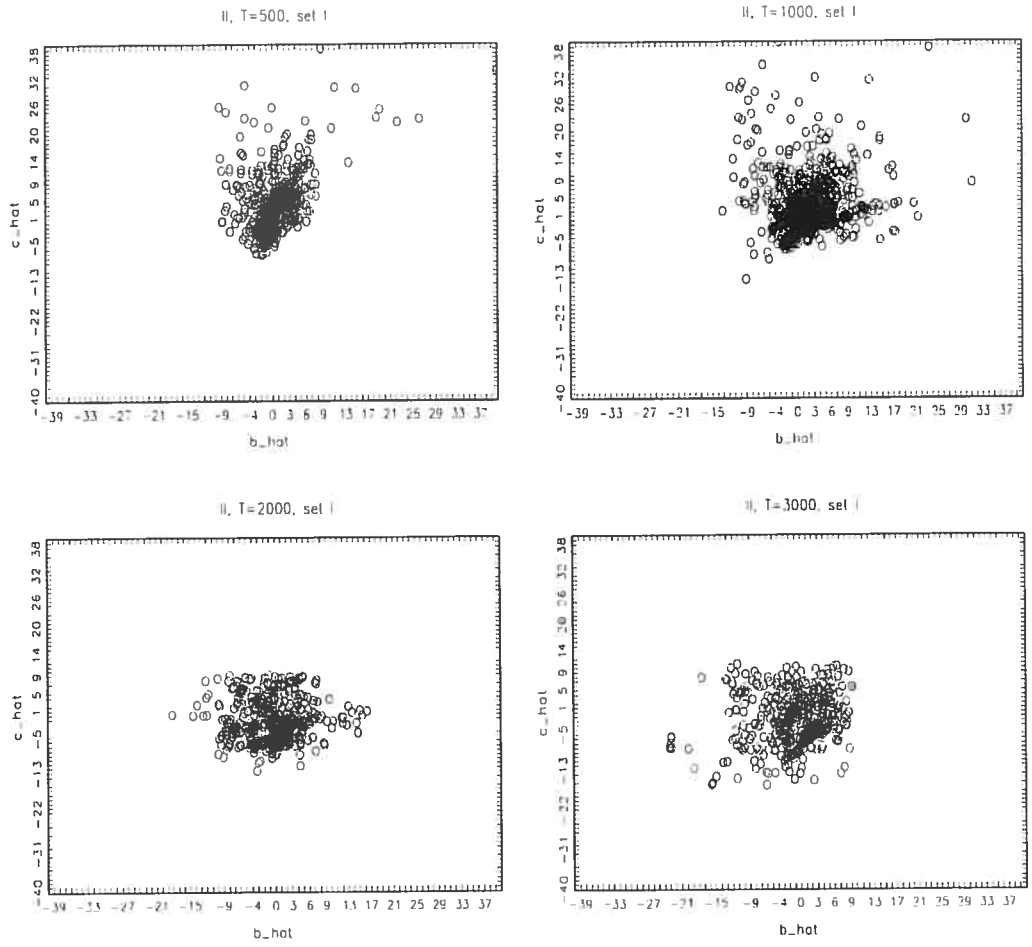


Figure 5.29: Standardized joint sample distribution of b and c, II, set I: (parameter per column, size per row:500,1000,2000,3000).

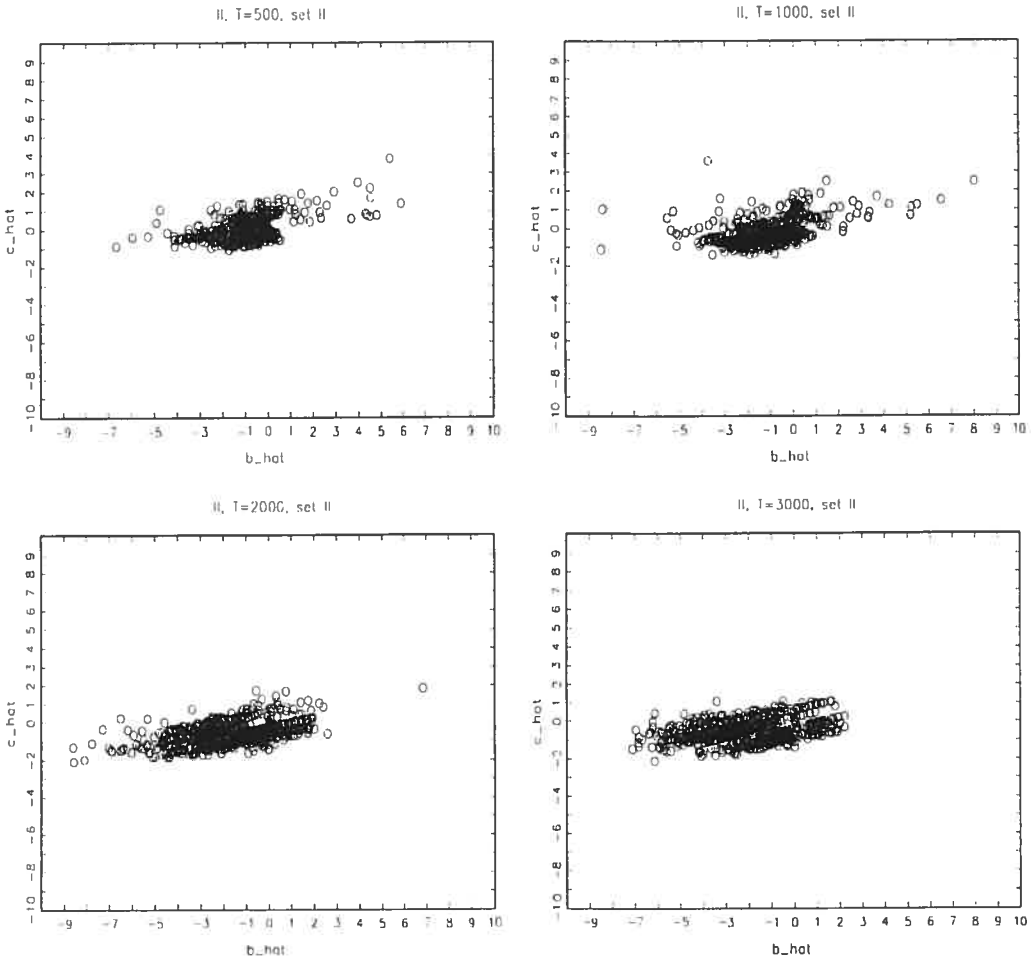


Figure 5.30: Standardized joint sample distribution of b and c, II, set II: (parameter per column, size per row:500,1000,2000,3000).

Conclusions générales

Dans des problèmes d'inférence non standard, la classe de procédures statistiques modernes qui exploite des méthodes d'inférence simulées telles que les moments simulés, l'inférence indirecte, la méthode des moments efficace, mais aussi les techniques bootstrap, Monte Carlo Markov chain ou encore les tests de Monte Carlo (maximisés) permettent de dériver des estimateurs et des procédures de tests puissants et robustes pour des modèles complexes dans des situations possiblement non standard - condition de rang des matrices violée, problème d'identification de paramètres de nuisance sur un sous-espace de l'espace des paramètres, restrictions redondantes dans des modèles nonlinéaires, présence de racines unitaires...- modulo une modification des résultats asymptotiques usuels.

Néanmoins, il reste que les techniques d'inférence simulées ne doivent pas être déconnectées des procédures statistiques conventionelles mais doivent être utilisées en combinaison avec celles-ci. En particulier elles doivent exploiter les caractéristiques dynamiques des processus étudiés - telles que celles identifiées par des techniques de décomposition canonique par exemple, qui permettent d'identifier les directions de corrélation maximale et d'obtenir tous les moments conditionels des processus - pour rapporter des outils statistiques puissants et fiables.

En particulier, l'estimation par inférence indirecte ou EMM requiert l'utilisation d'un modèle auxiliaire qui fournit une bonne approximation du modèle structurel pour produire un estimateur aussi efficace que possible [cf. Gallant and Long (1997), Tauchen (1997)]. Si le modèle auxiliaire emboîte le vrai processus générateur de données alors les estimées quasi-maximum de vraisemblance deviennent des statistiques suffisantes et dans ce cas, l'estimateur (II ou EMM) est pleinement efficace [Gallant and Tauchen (1996)].

De plus les estimateurs simulés (SMM, II, EMM, MCMC) sont très utilisés dans des applications utilisant des données des marchés financiers ou des données d'enchères.

Cependant il semble, dans des situations où le maximum de vraisemblance est évacué parce qu'infaisable, qu'aucun résultat théorique n'existe qui permettent d'affirmer

la supériorité des méthodes simulées sur des méthodes plus conventionelles [Gallant and Tauchen (1999)].

Bibliographie

ABRAMOWITZ, M., AND I. STEGUN (1965): *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables.*, vol. 55. National Bureau of Standards Applied Mathematics Series.

AIT-SAHALIA, Y. (2002): "Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach," *Econometrica*, 70, 223–262.

ANDERSEN, T., G. (1994): "Stochastic Autoregressive Volatility: A Framework for Volatility Modelling.," *Mathematical Finance*, 4(2), 75–102.

ANDERSEN, T., AND B. SØRENSEN (1996): "GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study," *Journal of Business and Economic Statistics*, 14(3), 328–352.

ANDERSEN, T. G., H.-J. CHUNG, AND B. E. SØRENSEN (1999): "Efficient Method of Moments Estimation of a Stochastic Volatility Model: A Monte Carlo Study," *Journal of Econometrics*, 91, 61–87.

ANDERSON, E., L. HANSEN, AND T. SARGENT (2003): "A Quartet of Semigroups for Model Specification, Robustness, Prices of Risk and Model Detection," *Journal of the European Economic Association*, 1.

ANDREWS, D. W. K. (1987): "Asymptotic Results for Generalized Wald Tests," *Econometric Theory*, 3, 348–358.

——— (1993a): "An Introduction to Econometric Applications of Functional Limit Theory for Dependent Random Variables," *Econometric Reviews*, 12, 183–216.

——— (1993b): "Test For Parameter Instability and Structural Change With Unknown Change Point," *Econometrica*, 61(4), 821–856.

ANDREWS, D. W. K., AND W. PLOBERGER (1994): "Optimal Tests When A Nuisance Parameter is Present Only under the Alternative," *Econometrica*, 62, 1383–1414.

BAILLIE, R., AND H. CHUNG (2001): "Estimation of GARCH Models from the Autocorrelations of the Squares of a Process," *Journal of Time Series mAnalysis*, 22(6), 631–650.

BAILLIE, R. T., T. BOLLERSLEV, AND H. O. MIKKELSEN (1996): "Fractionally integrated generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 74, 3–30.

BANSAL, R., A. R. GALLANT, R. HUSSEY, AND G. E. TAUCHEN (1995): "Nonparametric Estimation of Structural Models for High-Frequency Currency Market Data," *Journal of Econometrics*, 66, 251–287.

BARNARD, G. A. (1963): "Comment on "The Spectral Analysis of Point Processes" by M. S. Bartlett," *Journal of the Royal Statistical Society, Series B*, 25, 294.

BARNDORFF-NIELSEN, O. E., AND N. SHEPHARD (2002): "Econometric analysis of realised volatility and its use in stochastic volatility models," *Journal of the Royal Statistical Society, series B*, 63, 253–280.

——— (2003): "Realised power variation and stochastic volatility models," *Bernoulli*, 9, 243–265 and 1109–1111.

BERGER, A., AND S. WALLENSTEIN (1989): "On the Theory of $C(\alpha)$ tests," *Statistics and Probability Lettres*, 7, 419–424.

BIELECKI, T., AND M. RUTKOWSKI (2000): "Multiple Ratings Model of Defaultable Term Structure," *Mathematical Finance*, 10, 125–139.

BIRNBAUM, Z. W. (1974): "Computers and Unconventional Test-Statistics," in *Reliability and Biometry*, ed. by F. Proschan, and R. J. Serfling, pp. 441–458. SIAM, Philadelphia, PA.

BLACK, F., AND M. SCHOLES (1973): "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81, 637–654.

BOCHNER, S. (1955): "Sturm-Liouville and heat equations whose eigenfunctions and ultraspherical polynomials or associated Bessel functions," in *Proceedings of the*

Conference on Differential equations, pp. 23–48. University of Maryland Book store, College Park, Md.

BOLLERSLEV, T. (1986): "Generalized Autoregressive Conditional Heteroscedasticity," *Journal of Econometrics*, 51, 307–327.

BOLLERSLEV, T., AND H. O. MIKKELSEN (1996): "Modelling and Pricing long memory in stock market volatility," *Journal of Econometrics*, 73, 151–184.

BOLLERSLEV, T., AND J. H. WRIGHT (2000): "Semiparametric estimation of long memory volatility dependencies: The role of high-frequency data," *Journal of econometrics*, 98, 81–106.

BOUND, J., D. A. JAEGER, AND R. M. BAKER (1995): "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogeneous Explanatory Variable is Weak," *Journal of the American Statistical Association*, 90, 443–450.

BREIDT, F. J., N. CRATO, AND P. LIMA (1998): "The Detection and Estimation of Long Memory in Stochastic Volatility," *Journal of Econometrics*, 83, 325–348.

BREUSCH, T. S., AND P. SCHMIDT (1988): "Alternative forms of the Wald test: How long is a piece of string?," *Communications in statistics, Theory and Methods*, 17, 2789–2795.

BROCKWELL, P. J., AND R. A. DAVIS (1991): *Time Series: Theory and Methods*. Springer, New York, 2 edn.

BUSE, A. (1992): "The Bias of Instrumental Variables Estimators," *Econometrica*, 60, 173–180.

CAGETTI, M., L. HANSEN, T. SARGENT, AND N. WILLIAMS (2002): "Robustness and Pricing with Uncertain Growth.," *The Review of Financial Studies*, 15.

CALZOLARI, G., G. FIORENTINI, AND E. SENTANA (2004): "Constrained Indirect Estimation," *forthcoming, Review of Economic Studies*.

CARRASCO, M., AND X. CHEN (2002): "Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models," *Econometric Theory*, 18(1), 17–39.

CARRASCO, M., AND J.-P. FLORENS (2000): "Generalization of GMM to a Continuum of Moment Conditions," *Econometric Theory*, 16, 797–834.

CARRASCO, M., L. HANSEN, AND X. CHEN (1999): "Nonlinearity and Temporal Dependence," Discussion paper.

CHAO, J., AND N. R. SWANSON (2000): "Bias and MSE of the IV Estimators under Weak Identification," Department of Economics, University of Maryland, Discussion paper.

CHEN, X., L. HANSEN, AND J. SCHEINKMAN (1998): "Shape Preserving Estimation of Diffusions," Discussion paper, University of Chicago.

CHEN, Z., AND L. EPSTEIN (1999): "Ambiguity, Risk and Asset Returns in Continuous Time," Econometrica, forthcoming.

CHERNOV, M., A. R. GALLANT, E. GHYSELS, AND G. TAUCHEN (2003): "Alternative models for stock price dynamics," *Journal of Econometrics*, 116, 225–257.

CHIB, S., S. KIM, AND N. SHEPHARD (1998): "Stochastic volatility: Likelihood inference and comparison with ARCH models," *Review of Economic studies*, 65, 361–393.

CHIB, S., F. NARDARI, AND N. SHEPHARD (2002): "Markov Chain Monte Carlo methods for stochastic volatility models," *Journal of Econometrics, series B*, 108, 281–316.

CHRISTIANSEN, P. O., D. LANDO, AND K. R. MILTERSEN (1998): "State Dependent Realignments in Target Zone Currency Regimes," *Review of Derivatives Research*, 1, 295–323.

CLEMENT, E., C. GOURIÉROUX, AND A. MONFORT (2000): "Econometric Specification of the Risk Neutral Valuation Model," *Journal of Econometrics*, 94, 117–143.

COMTE, F., L. COUTIN, AND E. RENAULT (2003): "Affine Fractional Stochastic Volatility Models with Application to Option Pricing," Discussion paper, Université de Montréal.

COMTE, F., AND E. RENAULT (1996): "Long Memory Continuous Time Models," *Journal of Econometrics*, 73, 101–149.

——— (1998): "Long memory in continuous time stochastic volatility models," *Mathematical Finance*, 8, 291–323.

CONLEY, T., L. HANSEN, E. LUTTMER, AND J. SCHEINKMAN (1997): "Short-Term Interest Rates as Subordinated Diffusions.," *The Review of Financial Studies*, 10(3), 525–577.

COX, J., J. INGERSOLL, AND S. ROSS (1985a): "An Intertemporal General Equilibrium Model of Asset Prices," *Econometrica*, 53, 363–384.

——— (1985b): "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385–408.

DAGENAIS, M., AND J.-M. DUFOUR (1991): "Invariance, Nonlinear Models and Asymptotic Tests," *Econometrica*, 59, 1601–1615.

DAI, Q., AND K. SINGLETON (2003): "Term Structure Dynamics in Theory and Reality," *forthcoming Review of Financial Studies*.

DANIELSSON, J. (1994): "Stochastic Volatility in Asset Prices: Estimation with Simulated Maximum Likelihood," *Journal of Econometrics*, 61, 375–400.

DANIELSSON, J., AND J.-F. RICHARD (1993): "Accelerated Gaussian Importance Sampler with Application to Dynamic Latent Variable Models," *Journal of Applied Econometrics*, 8, S153–S173.

DAROLLES, S., J.-P. FLORENS, AND C. GOURIÉROUX (2000): "Kernel Based Nonlinear Canonical Analysis and Time Reversibility.," Discussion paper 9855, CREST, Journal of Econometrics, forthcoming.

DAROLLES, S., J.-P. FLORENS, AND E. RENAULT (1997): "Nonlinear Principal Components and Inference on a Conditional Expectation Operator.," Discussion paper, CREST.

DAROLLES, S., AND C. GOURIÉROUX (2001): "Truncated Dynamics and Estimation of Diffusion Equations.," *Journal of Econometrics*, 102, 1–22.

DAVIDSON, J. (1994): *Stochastic Limit Theory*, Advanced Texts in Econometrics. Oxford University Press Inc., New York.

DE JONG, F., F. C. DROST, AND B. J. M. WERKER (2001): "A Jump Diffusion Model for Exchange Rates in a Target Zone," *Statistica Neerlandica*, 55, 270–300.

DEMOURA, S. (1993): "Theory and Application of Transition Operator to the Analysis of Economic Time Series.," Discussion paper, Necessary and Sufficient Conditions for Nonlinearities in Economic Dynamics. Aliasing problem. Discussion Paper.

DUFFIE, D., AND K. J. SINGLETON (1993): "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, 61, 929–952.

DUFOUR, J.-M. (1997): "Some Impossibility Theorems in Econometrics, with Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365–1389.

——— (2002): "Monte Carlo Tests with Nuisance Parameters: A General Approach to Finite-Sample Inference and Nonstandard Asymptotics in Econometrics," *Journal of Econometrics, forthcoming*.

DUFOUR, J.-M., AND L. KHALAF (1997): "Simulation Based Finite and Large Sample Inference Methods in Multiple Equation Regression Models," in *Proceedings of the Business and Economic Statistics Section of the American Statistical Association*, pp. 75–80, Washington (D.C.).

DUFOUR, J.-M., AND L. KHALAF (2002a): "Exact Tests for Contemporaneous Correlation of Disturbances in Seemingly Unrelated Regressions," *Journal of Econometrics*, 106, 143–170.

——— (2002b): "Simulation Based Finite and Large Sample Tests in Multivariate Regressions," *Journal of Econometrics*, 111(2), 303–322.

DUFOUR, J.-M., L. KHALAF, J.-T. BERNARD, AND I. GENEST (2004): "Simulation-Based Finite-Sample Tests for Heteroskedasticity and ARCH effects," *Journal of Econometrics*, 122, 317–347.

DUFOUR, J.-M., AND J. F. KIVIET (1996): "Exact Tests for Structural Change in First-Order Dynamic Models," *Journal of Econometrics*, 70, 39–68.

DUFOUR, J.-M., AND M. TAAMOUTI (2000): "Statistical Inference and Projection techniques in Simultaneous Equations Models," *Econometrica, forthcoming*.

——— (2003): "Point-Optimal Instruments and Generalized Anderson-Rubin Proce-

dures for Nonlinear Models," Département de Sciences Economiques, Université de Montréal, Discussion paper.

DUFOUR, J.-M., AND A. TROGNON (2001): "Invariant tests based on M-estimators, estimating functions, and the generalized method of moments," Discussion paper, CIREQ, University of Montreal.

DUFOUR, J.-M., AND P. VALÉRY (2004): "Finite and Large Sample Inference for a Stochastic Volatility Model," Discussion Paper, Université de Montréal.

DUFOUR, J.-M., AND P. VALÉRY (2004): "On a simple closed-form estimator for a stochastic volatility model," Discussion paper, Université de Montréal.

DWASS, M. (1957): "Modified Randomization Tests for Nonparametric Hypotheses," *Annals of Mathematical Statistics*, 28, 181–187.

ENGLE, R. F. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007.

FAN, J., AND C. ZHANG (2003): "A Reexamination of Diffusion Estimators With Applications to Financial Model Validation," *Journal of the American Statistical Association*, 98(461), 118–134.

FIORENTINI, G., E. SENTANA, AND N. SHEPHARD (2004): "Likelihood-based estimation of latent generalized ARCH structures," *Econometrica, forthcoming*.

FLORENS, J.-P., E. RENAULT, AND N. TOUZI (1998): "Testing for Embeddability by Stationary Reversible Continuous-Time Markov Processes.," *Econometric Theory*, 14, 744–769.

FOUTZ, R. V. (1980): "A Method for Constructing Exact Tests from Test Statistics that have Unknown Null Distributions," *Journal of Statistical Computation and Simulation*, 10, 187–193.

FULLER, W. A. (1976): *Introduction to statistical Time Series*. John Wiley and Sons, New York.

GAGLIARDINI, P., AND C. GOURIÉROUX (2003): "Spread Term Structure and Default Correlation," Discussion paper, CREST.

——— (2004): "Stochastic Migration Models with Application to Corporate Risk," Discussion paper, CREST.

GALLANT, A., D. HSIEH, AND G. TAUCHEN (1997): "Estimation of Stochastic Volatility Models with Diagnostics," *Journal of Econometrics*, 81(1), 159–192.

GALLANT, A. R. (1977): "Testing a Nonlinear Regression Specification: A Nonregular Case," *Journal of the American Statistical Association*, 72, 523–530.

GALLANT, A. R., AND J. LONG (1997): "Estimating Stochastic Differential Equations Efficiently by Minimum Chi-Squared," *Biometrika*, 84, 125–141.

GALLANT, A. R., AND G. TAUCHEN (1989): "Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications," *Econometrica*, 57, 1091–1120.

GALLANT, A. R., AND G. TAUCHEN (1996): "Which Moments to Match?," *Econometric Theory*, 12, 657–681.

GALLANT, A. R., AND G. TAUCHEN (1999): "The Relative Efficiency of Method of Moments Estimators," *Journal of Econometrics*, 92, 149–172.

GARCIA, R., AND P. PERRON (1996): "An Analysis of the Real Interest Rate under Regime Shifts," *Review of Economic and Statistics*, 78, 111–125.

GARCIA, R., E. RENAULT, AND D. VEREDAS (2004): "Estimation of Stable Distributions by Indirect Inference," Discussion paper, Université de Montréal.

GENTON, M. G., AND E. RONCHETTI (2001): "Robust Indirect Inference," Département d'Econométrie, Université de Genève, Discussion paper no 2001.01.

GHYSELS, E., C. GOURIÉROUX, AND J. JASIAK (1995): "Time Deformation.," in *Proceedings of the International Conference on Forcasting Financial Markets, New Advance for Exchange Rates and Asset Management*. London.

——— (1998): *Market Time and Asset Price Movements: Theory and Estimation.*Statistics in Finance. Arnold, hand, d. and jacka, s. edn., London.

GHYSELS, E., A. HARVEY, AND E. RENAULT (1996): "Stochastic Volatility," in *Statistical Methods in Finance*, ed. by C. Rao, and G. Maddala, Holland. North-Holland.

GODAMBE, V. P., AND C. C. HEYDE (1987): "Quasi-likelihood and Optimal Estimation," *International Statistical Review*, 55, 231–244.

GOFFE, W. L., G. D. FERRIER, AND J. ROGERS (1994): "Global Optimization of Statistical Functions with Simulated Annealing," *Journal of Econometrics*, 60, 65–99.

GONÇALVES, S., AND S. KILIAN (2004): "Bootstrapping Autoregressions with Conditional Heteroskedasticity of Unknown Form," *Journal of Econometrics*, 123, 89–120.

GOURIÉROUX, C. (1997): *ARCH Models and Financial Applications*, Springer Series in Statistics. Springer-Verlag, New York.

GOURIÉROUX, C., AND A. MONFORT (1995a): *Simulation Based Econometric Methods*, CORE Lecture Series. CORE Foundation, Louvain-la-Neuve.

——— (1995b): *Statistics and Econometric Models, Volumes One and Two*. Cambridge University Press, Cambridge, U.K.

GOURIÉROUX, C., A. MONFORT, AND E. RENAULT (1993): "Indirect Inference," *Journal of Applied Econometrics*, S8, 85–118.

GOURIÉROUX, C., AND J. JASIAK (2003): "Jacobi Process with Application to Switching Regimes," *forthcoming Journal of Econometrics*.

GOURIÉROUX, C., AND A. MONFORT (1989): *Statistique et Modèles économétriques*, vol. 1,2. Economica.

GOURIÉROUX, C., E. RENAULT, AND N. TOUZI (2000): *Simulation-based Inference in Econometrics. Methods and Application*.chap. Calibration by Simulation for Small Sample Bias Correction, pp. 328–358. Cambridge University Press.

GOURIÉROUX, C., E. RENAULT, AND P. VALÉRY (2002): "Diffusion Processes with Polynomial Eigenfunctions," Discussion Paper, Université de Montréal.

GRANGER, C. W. J., AND T. TERASVIRTA (1993): *Modelling Nonlinear Economic Relationship*. Oxford University Press, Oxford.

GREGORY, A. V., AND M. R. VEALL (1985): "Formulating Wald tests of nonlinear restrictions," *Econometrica*, 53, 1465–1468.

HALL, A., G. RUDEBUSCH, AND D. WILCOX (1996): "Judging Instrument Relevance in Instrumental Variables Estimation," *International Economic Review*, 37, 283–289.

HAMILTON, J. (1988): "Rational Expectations Econometric Analysis of Changes in Regime. An Investigation of the Term Structure of Interest Rates," *Journal of Economic Dynamics and Control*, 12, 385–423.

——— (1990): "Analysis of Time Series Subject to Changes in Regime," *Journal of Econometrics*, 45, 39–70.

HAMILTON, J. (1994): *Time Series Analysis*. Princeton University Press Inc., New Jersey.

HAMILTON, J. D. (1989): "A new Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357–384.

HANNAN, E. (1961): "The General Theory of Canonical correlation and and its relation to functional analysis," *Journal of Australian Mathematical analysis*, 2, 229–242.

HANSEN, B. E. (1996): "Inference When A Nuisance Parameter Is Not Identified Under the Null Hypothesis," *Econometrica*, 64(2), 413–430.

HANSEN, L., AND J. SCHEINKMAN (1995): "Back to the Future: Generating Moment Implications for Continuous Time Markov Processes.," *Econometrica*, 73, 767–807.

HANSEN, L., J. SCHEINKMAN, AND N. TOUZI (1998): "Spectral Methods for Identifying Scalar Diffusions," *Journal of Econometrics*, 86, 1–32.

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

HARVEY, A., E. RUIZ, AND N. SHEPHARD (1994): "Multivariate Stochastic Variance Models," *Review of Economic Studies*, 61, 247–264.

HARVEY, A., AND N. SHEPHARD (1996): "Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns," *Journal of Business and Economic Statistics*, 14, 429–434.

HEYDE, C. C. (1997): *Quasi-likelihood and its Application*. Springer-Verlag, New York.

HOSKING, J. (1981): "Fractional Differencing," *Biometrika*, 68, 165–186.

JACQUIER, E., N. POLSON, AND P. ROSSI (1994): "Bayesian Analysis of Stochastic Volatility Models (with discussion)," *Journal of Economics and Business Statistics*, 12, 371–417.

JARROW, R., AND F. YU (2001): "Counterparty Risk and the Pricing of Defaultable Securities," *Journal of Finance*, 56, 1756–1799.

KARLIN, S., AND J. MCGREGOR (1960): "Classical Diffusion Processes and Total Positivity," *J. Math. Anal. Appl.*, 1, 163–183.

KESSLER, M., AND M. SORENSEN (1999): "Estimating Equations Based on Eigenfunctions for a Discretely Observed Diffusion Process," *Bernoulli*, 5, 299–314.

KIM, S., AND N. SHEPHARD (1994): "Stochastic Volatility: Optimal Likelihood Inference and Comparison with ARCH Models," Discussion paper, Nuffield College, Oxford, unpublished paper.

KIM, S., N. SHEPHARD, AND S. CHIB (1998): "Stochastic volatility : Likelihood inference and comparison with ARCH models," *Review of economic studies*, 65, 361–393.

KIVIET, J., AND J.-M. DUFOUR (1997): "Exact Tests in Single Equation Autoregressive Distributed Lag Models," *Journal of Econometrics*, 80, 325–353.

KOCHERLAKOTA, S., AND K. KOCHERLAKOTA (1991): "Neyman's $C(\alpha)$ test and Rao's efficient score test for composite hypotheses," *Statistics and Probability Lettres*, 11, 491–493.

KUERSTEINER, G. M. (2001): "Optimal Instrumental Variables Estimation for ARMA Models," *Journal of Econometrics*, 104(2), 349–405.

LANCASTER, H. (1968): "The Structure of Bivariate Distributions," *Annals of Mathematical Statistics*, 3, 311–340.

LANDO, D. (1998): "On Cox processes and Credit Risky Securities," *Review of Derivatives Research*, 2, 99–120.

LARSEN, K. S., AND M. SORENSEN (2003): "Diffusion Models for Exchange Rates in a Target Zone," Discussion paper, University of Copenhagen.

LO, A. W. (1988): "Maximum Likelihood Estimation of Generalized Ito Processes with Discretely Sampled Data," *Econometric Theory*, 4, 231–247.

LUNDBERGH, S., T. TERASVIRTA, AND D. VAN DIJK (2001): "Time-Varying Smooth Transition Autoregressive Models," *Journal of Business and Economic Statistics*, Forthcoming.

LUTKEPOHL, H., AND M. M. BURDA (1997): "Modified Wald tests under nonregular conditions," *Journal of Econometrics*, 78, 315–332.

MADDALA, G., AND J. JEONG (1992): "On the Exact Small Sample Distribution of the IV Estimator," *Econometrica*, 60, 181–183.

MAHIEU, R., AND P. SCHOTMAN (1998): "An Empirical application of stochastic volatility models," *Journal of Applied Econometrics*, 13, 333–360.

MARRIOTT, F. H. C. (1979): "Barnard's Monte Carlo Tests: How Many Simulations?," *Applied Statistics*, 28, 75–77.

MEDDAHI, N. (2002): "Moments of Continuous Time Stochastic volatility models," Discussion paper, Université de Montréal.

MELINO, A., AND S. TURNBULL (1990): "Pricing Foreign Currency Options with Stochastic Volatility," *Journal of Econometrics*, 45, 239–265.

MONFARDINI, C. (1998): "Estimating Stochastic volatility Models through Indirect Inference," *The Econometrics journal*, 1(1).

NELSON, C. R., AND R. STARTZ (1990a): "The Distribution of the Instrumental Variables Estimator and Its t-ratio when the Instrument is a Poor One," *Journal of Business*, 63, S125–S140.

——— (1990b): "Some Further Results on the Exact Small Sample Properties of the Instrumental Variables Estimator," *Econometrica*, 58, 967–976.

NELSON, D. (1988): "Times series behavior of stock market volatility and returns," unpublished Ph.P dissertation, Massachusetts Institute of technology, Economics Dept.

NELSON, F. D., AND N. SAVIN (1990): "The Danger of Extrapolating Asymptotic Local Power," *Econometrica*, 58, 977–981.

NEWEY, W. K., AND D. MCFADDEN (1994): *Large Sample Estimation and Hypothesis*

Testing, vol. 4 of *Handbook of Econometrics*. Engle, R. F., McFadden, D. L., North Holland, Amsterdam.

NEWEY, W. K., AND K. D. WEST (1987a): "Hypothesis Testing with Efficient Method of Moments Estimators," *International Economic Review*, 28, 777–787.

——— (1987b): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

NEYMAN, J. (1959): "Optimal Asymptotic Tests of Composite Statistical hypotheses," in *Probability and Statistics, the Harald Cramer Volume*, ed. by U. Grenander, pp. 213–234. Uppsala: Almqvist and Wiksell.

NIELSEN, T., J. SAA REQUEJA, AND P. SANTA CLARA (1993): "Default Risk and Interest Rate Risk.," Discussion paper.

OHANISSIAN, A., J. R. RUSSELL, AND R. S. TSAY (2003): "True or Spurious Long Memory in Volatility: Does it Matter for Pricing Option," Discussion paper, University of Chicago.

PAGAN, A., A. HALL, AND V. MARTIN (1996): "Modelling the Term Structure.," *Handbook of statistics*, 14, 91–118.

PALM, F. C. (1996): *GARCH Models of Volatility*, vol. 14 of *Handbook of Statistics*. Maddala, G.S., Rao, C. R., North Holland, Amsterdam.

PHILLIPS, P. (1989): "Partially Identified Econometric Models," *Econometric Theory*, 5, 181–240.

RONCHETTI, E. (1987): "Robust $C(\alpha)$-type tests for linear models," *Sankhya: The Indian journal of Statistics*, 49(Series A), 1–16.

RUDIN, W. (1976): *Principles of Mathematical Analysis*. McGraw-Hill, Inc.

RUIZ, E. (1994): "Quasi-maximum likelihood estimation of stochastic variance models," *Journal of Econometrics*, 63, 284–306.

SARGAN, J. (1983): "Identification and Lack of Identification," *Econometrica*, 51, 1605–1633.

SCHONBUCHER, P., AND D. SCHUBERT (2001): "Copula Dependent Default Risk in Intensity Models," Discussion paper, Bonn University.

SHEPHARD, N. (1996): "Statistical Aspects of ARCH and Stochastic Volatility," in *Time Series Models in Econometrics, Finance and Other Fields*, ed. by D. Cox, O. Barndorff-Nielsen, and D. Hinkley. Chapman & Hall, London.

SMITH, A. (1993): "Estimating Nonlinear Time Series Models Using Simulated Vector Autoregressions," *Journal of Applied Econometrics*, 8, S63–S84.

STAIGER, D., AND J. STOCK (1997): "Instrumental Variable Regressions with Weak Instruments," *Econometrica*, 65(3), 557–586.

TAUCHEN, G. (1997): "New Minimum Chi-Square Methods in Empirical Finance," *Advances in Econometrics, Seventh World Congress, ed. by David Kreps, and Kenneth F. Wallis. Cambridge University Press*.

TAYLOR, S. (1986): *Modelling Financial Time Series*. John Wiley, Chichester.

——— (1994): "Modelling Stochastic Volatility," *Mathematical Finance*, 4, 183–204.

TERASVIRTA, T. (1994): "Specification, Estimation and Evaluation of Smooth Transition Autoregressive Models," *Journal of the American Statistical Association*, 89(425), 208–218.

——— (1998): "Modelling Economic Relationships with Smooth Transition Regressions," *Handbook of Applied Economic Statistics*, pp. 507–552, in A. Ullah and E.E.E. Giles (eds.),New York: Marcel Dekker.

TERASVIRTA, T., AND H. M. ANDERSON (1992): "Characterizing Nonlinearities in Business Cycles using Smooth Transition Autoregressive Models," *Journal of Applied Econometrics*, 7, S119–S136.

VASICEK, O. (1977): "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics*, 5, 177–188.

WHITE, H., AND I. DOMOWITZ (1984): "Nonlinear Regression with Dependent Observations," *Econometrica*, 52, 143–161.

WONG, C. (2002a): "Estimating stochastic volatility:new method and comparison," Discussion paper, University of Oxford, Ph.d. Thesis.

——— (2002b): "The MCMC method for an extended stochastic volatility model," Discussion paper, University of Oxford, Ph.d. Thesis.

WONG, E. (1964): "The Construction of a Class of Stationary Markoff Processes.," in *Stochastic processes in Mathematical Physics and Engineering*, ed. by R. Bellman, Sixteenth Symposium in Applied Mathematics, pp. 264–276. Providence, R I, American Mathematical Society.

WONG, E., AND B. HAJEK (1985): *Stochastic Processes in Engineering Systems*. Springer, New York.

WONG, E., AND J. THOMAS (1962): "On Polynomial Expansions of Second-Order Distributions," *J. Soc. Indust. Appl. Math.*, 10(3), 507–516.

ZIVOT, E., R. STARTZ, AND C. R. NELSON (1998): "Valid Confidence Intervals and Inference in the Presence of Weak Instruments," *International Economic Review*, 39(4), 1119–1145.

Appendix: Proofs of chapter 1

PROOF OF LEMMA **1.3.1** First, if $U \sim \mathrm{N}(0,\,1)$ then $\mathsf{E}(U^{2p+1}) = 0,\, \forall\, p \in \mathbb{N}$ and $\mathsf{E}(U^{2p}) = (2p)!/[2^p p!]\ \forall\, p \in \mathbb{N}$ [see Gouriéroux and Monfort (1995a, Volume 2, page 518)]. Under Assumptions **1.2.1**,

$$\begin{aligned}
\mathsf{E}(u_t^k) &= r_y^k \mathsf{E}(z_t^k)\mathsf{E}[\exp(kw_t/2)] \\
&= r_y^k \frac{k!}{2^{(k/2)}(k/2)!} \exp\left[\frac{k^2}{4} r_w^2/2(1-a^2)\right)] \\
&= r_y^k \frac{k!}{2^{(k/2)}(k/2)!} \exp\left[\frac{k^2}{8} r_w^2/(1-a^2)\right]
\end{aligned}$$

where the second equality uses the definition of the Gaussian Laplace transform of $w_t \sim \mathrm{N}[0,\, r_w^2/(1-a^2)]$ and of the moments of the $\mathrm{N}(0,\,1)$ z_t variable. Let us now calculate the cross-product:

$$\begin{aligned}
\mathsf{E}[u_t^k u_{t+m}^l] &= \mathsf{E}[r_y^{k+l} z_t^k z_{t+m}^l \exp(k\frac{w_t}{2} + l\frac{w_{t+m}}{2})] \\
&= r_y^{k+l} \mathsf{E}(z_t^k)\mathsf{E}(z_{t+m}^l)\mathsf{E}\left[\exp(k\frac{w_t}{2} + l\frac{w_{t+m}}{2})\right] \\
&= r_y^{k+l} \frac{k!}{2^{(k/2)}(k/2)!} \frac{l!}{2^{(l/2)}(l/2)!} \exp\left[\frac{r_w^2}{8(1-a^2)}(k^2 + l^2 + 2kla^m)\right]
\end{aligned}$$

where $\mathsf{E}(w_t) = 0$, $Var(w_t) = \frac{r_w^2}{1-a^2}$ and

$$\begin{aligned}
Var(k\frac{w_t}{2} + l\frac{w_{t+m}}{2}) &= \frac{k^2}{4} Var(w_t) + \frac{l^2}{4} Var(w_{t+m}) + 2\frac{k}{2}\frac{l}{2}\mathsf{Cov}(w_t,\, w_{t+m}) \\
&= \frac{r_w^2}{4(1-a^2)}(k^2 + l^2 + 2kla^m)\,.
\end{aligned}$$

□

PROOF OF LEMMA **1.3.2** Taking the ratio of equation (1.3.7) on equation (1.3.6) to

the square produces

$$\frac{\mathsf{E}(u_t^4)}{\left(\mathsf{E}(u_t^2)\right)^2} = 3\exp[r_w^2/(1-a^2)] \,,$$

i.e.

$$r_w^2/(1-a^2) = \log\left(\frac{\mathsf{E}(u_t^4)}{3(\mathsf{E}(u_t^2))^2}\right) \equiv Q \,. \tag{.0.1}$$

Inserting $Q \equiv r_w^2/(1-a^2)$ in equation (1.3.6) yields

$$r_y = \left(\frac{\mathsf{E}(u_t^2)}{\exp(Q/2)}\right)^{1/2} = \frac{3^{1/4}\mathsf{E}(u_t^2)}{\mathsf{E}(u_t^4)^{1/4}} \,.$$

From equation (1.3.8), we have

$$\exp(\frac{r_w^2}{(1-a)}) = \frac{\mathsf{E}[u_t^2 u_{t-1}^2]}{r_y^4}$$

which, after a few manipulations, yields

$$1 + a = \frac{[\log(\mathsf{E}[u_t^2 u_{t-1}^2]) - 4\log(r_y)]}{Q}$$

or either

$$a = \frac{[\log(\mathsf{E}[u_t^2 u_{t-1}^2]) - \log(3) - 4\log(\mathsf{E}[u_t^2]) + \log(\mathsf{E}[u_t^4])]}{\log\left(\frac{\mathsf{E}[u_t^4]}{3(\mathsf{E}[u_t^2])^2}\right)} - 1 \,.$$

From the expressions of $Q \equiv r_w^2/(1-a^2)$ at equation (.0.1) and that of a above, we can deduce:

$$r_w = \left[(1-a^2)\log\left(\frac{\mathsf{E}[u_t^4]}{3(\mathsf{E}[u_t^2])^2}\right)\right]^{1/2} \,.$$

□

PROOF OF LEMMA **1.3.3** Here we derive the covariances of the components of

$X_t = (X_{1t},\, X_{2t},\, X_{3t})'$ that is

$$\begin{aligned}
\gamma_1(\tau) &= \mathrm{Cov}(X_{1t},\, X_{1,t+\tau}) = \mathrm{E}\{[u_t^2 - \mu_2(\theta)][u_{t+\tau}^2 - \mu_2(\theta)]\} \\
&= \mathrm{E}(u_t^2 u_{t+\tau}^2) - \mu_2^2(\theta) = r_y^4 \mathrm{E}\exp(w_t + w_{t+\tau}) - \mu_2^2(\theta) \\
&= r_y^4 \exp\left[\frac{r_w^2}{1-a^2}(1+a^\tau)\right] - \mu_2^2(\theta) = \mu_2^2(\theta)[\exp(\gamma a^\tau) - 1]\,,
\end{aligned}$$

where $\gamma = \frac{r_w^2}{1-a^2}$. Similarly,

$$\begin{aligned}
\gamma_2(\tau) &= \mathrm{Cov}(X_{2t},\, X_{2,t+\tau}) = \mathrm{E}\{[u_t^4 - \mu_4(\theta)][u_{t+\tau}^4 - \mu_4(\theta)]\} \\
&= \mathrm{E}(u_t^4 u_{t+\tau}^4) - \mu_4^2(\theta) = 9r_y^8 \mathrm{E}\{\exp[2(w_t + w_{t+\tau})]\} - \mu_4^2(\theta) \\
&= 9r_y^8 \exp[4\frac{r_w^2}{1-a^2}(1+a^\tau) - \mu_4^2(\theta)] = \mu_4^2(\theta)[\exp(4\gamma a^\tau) - 1]\,.
\end{aligned}$$

Finally,

$$\begin{aligned}
\gamma_3(\tau) &= \mathrm{Cov}(X_{3t},\, X_{3,t+\tau}) = \mathrm{E}\{[u_t^2 u_{t-1}^2 - \mu_{2,2}(1|\theta)][u_{t+\tau}^2 u_{t+\tau-1}^2 - \mu_{2,2}(1|\theta)]\} \\
&= \mathrm{E}[u_t^2 u_{t-1}^2 u_{t+\tau}^2 u_{t+\tau-1}^2] - \mu_{2,2}^2(1|\theta) \\
&= r_y^8 \mathrm{E}\exp(w_{t+\tau} + w_{t+\tau-1} + w_t + w_{t-1}) - \mu_{2,2}^2(1|\theta) \\
&= r_y^8 \exp[2(1+a)\gamma] \exp[\gamma(a^{\tau-1} + 2a^\tau + a^{\tau+1})] - \mu_{2,2}^2(1|\theta) \\
&= \mu_{2,2}^2(1|\theta)\{\exp[\gamma(a^{\tau-1} + 2a^\tau + a^{\tau+1})] - 1\} \\
&= \mu_{2,2}^2(1|\theta)\{\exp[\gamma(1+a)^2 a^{\tau-1}] - 1\}
\end{aligned}$$

for all $\tau \geq 2$. □

PROOF OF PROPOSITION **1.4.5**

The method-of-moments estimator $\hat{\theta}_T(\Omega)$ is solution of the following optimization problem:

$$\min_\theta M_T(\theta) = \min_\theta [\mu(\theta) - \bar{g}_T(\hat{U}_T)]' \hat{\Omega}_T [\mu(\theta) - \bar{g}_T(\hat{U}_T)]\,.$$

The first order conditions (F.O.C) associated with this problem are:

$$\frac{\partial \mu'}{\partial \theta}(\hat{\theta}_T)\hat{\Omega}_T[\mu(\hat{\theta}_T) - \bar{g}_T(\hat{U}_T)] = 0 \ .$$

An expansion of the F.O.C above around the true value θ yields

$$\frac{\partial \mu'}{\partial \theta}(\hat{\theta}_T)\hat{\Omega}_T\Big([\mu(\theta) + P(\theta)'(\hat{\theta}_T - \theta) - \bar{g}_T(\hat{U}_T)\Big] \simeq 0$$

where, after rearranging the equation,

$$\sqrt{T}\big[\hat{\theta}_T(\Omega) - \theta\big] \simeq [P(\theta)\Omega P(\theta)']^{-1}P(\theta)\Omega\sqrt{T}\big[\bar{g}_T(\hat{U}_T) - \mu(\theta)\big] \ .$$

Using Assumptions **1.4.1** to **1.4.4**, we get the asymptotic normality of $\hat{\theta}_T(\Omega)$ with asymptotic covariance matrix $V(\Omega)$ as specified in proposition **1.4.5**. □

PROOF OF PROPOSITION **1.4.7** In order to establish the asymptotic normality of $\sqrt{T}[\bar{g}_T(U_T) - \mu(\theta))]$ we shall use a Central Limit Theorem (C.L.T) for dependent processes [see Davidson (1994, Theorem 24.5, page 385)]. For that purpose, we first check the conditions under which this C.L.T holds. Setting

$$X_t \equiv \begin{pmatrix} u_t^2 - \mu_2(\theta) \\ u_t^4 - \mu_4(\theta) \\ u_t^2 u_{t-1}^2 - \mu_{2,2}(1|\theta) \end{pmatrix} = g_t(\theta) - \mu(\theta) \ ,$$

$$S_T = \sum_{t=1}^{T} X_t = \sum_{t=1}^{T} g_t(\theta) - \mu(\theta) \ ,$$

and the subfields $\mathcal{F}_t = \sigma(s_t, s_{t-1}, \ldots)$ where $s_t = (y_t, w_t)'$, we need to check three conditions:

a) $\{X_t, \mathcal{F}_t\}$ is stationary and ergodic,

b) $\{X_t, \mathcal{F}_t\}$ is a L_1-mixingale of size -1,

c)

$$\limsup_{T\to\infty} T^{-1/2}\mathsf{E}|S_T| < \infty \qquad (.0.2)$$

in order to get that $T^{-1/2}S_T = \sqrt{T}(\bar{g}_T(U_T) - \mu(\theta)) \xrightarrow{D} \mathrm{N}[0,\ \Omega_*]$.

a) By Propositions 5 and 17 from Carrasco and Chen (2002) we can say that:

(i) if $\{w_t\}$ is geometrically ergodic, then $\{(w_t,\ \ln|v_t|)\}$ is Markov geometrically ergodic with the same decay rate as the one of $\{w_t\}$;

(ii) if $\{w_t\}$ is stationary β-mixing with a certain decay rate, then $\{\ln|v_t|\}$ is β-mixing with a decay rate at least as fast as the one of $\{w_t\}$.

If the initial value v_0 follows the stationary distribution, $\{\ln|v_t|\}$ is strictly stationary β-mixing with an exponential decay rate. Since this property is preserved by any continuous transformation, $\{v_t\}$ and hence $\{v_t^k\}$ and $\{v_t^k v_{t-1}^k\}$ are strictly stationary and exponential β-mixing. We can then deduce that X_t is strictly stationary and exponential β-mixing.

b) A mixing zero-mean process is an adapted L_1-mixingale with respect to the subfields $\mathcal{F}_t$ provided it is bounded in the L_1-norm [see Davidson (1994, Theorem 14.2, page 211)]. To see that $\{X_t\}$ is bounded in the L_1-norm, we note that:

$$\mathsf{E}|v_t^2 - \mu_2(\theta)| \le \mathsf{E}(|v_t^2| + |\mu_2(\theta)|) = 2\mu_2(\theta) < \infty\ ,$$

$$\mathsf{E}|v_t^4 - \mu_4(\theta)| \le 2\mu_4(\theta) < \infty\ ,$$

$$\mathsf{E}|v_t^2 v_{t-1}^2 - \mu_{2,2}(1|\theta)| \le 2\mu_{2,2}(1|\theta) < \infty.$$

We now need to show that the L_1-mixingale $\{X_t,\ \mathcal{F}_t\}$ is of size -1. Since X_t is $\beta-$mixing, it has mixing coefficients of the type $\beta_n = c\rho^n\ ,\ c > 0,\ ,\ 0 < \rho < 1$. In order to show that $\{X_t\}$ is of size -1, we need to show that its mixing coefficients $\beta_n = O(n^{-\phi})$, with $\phi > 1$. Indeed,

$$\frac{\rho^n}{n^{-\phi}} = n^{\phi}\exp(n\log\rho) = \exp(\phi\log n)\exp(n\log\rho) = \exp(\phi\log n + n\log\rho)\ .$$

It is known that $\lim_{n\to\infty} \phi \log n + n \log \rho = -\infty$ which yields

$$\lim_{n\to\infty} \exp(\phi \log n + n \log \rho) = 0 .$$

This holds in particular for $\phi > 1$, [see Rudin (1976, Theorem 3.20(d), page 57)].

c) By Cauchy-Schwarz inequality, we have:

$$\mathsf{E}|T^{-1/2} S_T| \leq T^{-1/2} \|S_T\|_2$$

so that (.0.2) can be proven by showing that $\limsup_{T\to\infty} T^{-1}\mathsf{E}(S_T S_T') < \infty$.

We shall prove that:

$$\limsup_{T\to\infty} T^{-1}\mathsf{E}(S_T S_T') = \limsup_{T\to\infty} Var\left[\frac{1}{\sqrt{T}} S_T\right] < \infty$$

i) The first component of S_T. Set $S_{T1} = \sum_{t=1}^{T} X_{1,t}$ where $X_{1,t} \equiv u_t^2 - \mu_2(\theta)$. We compute:

$$\begin{aligned} Var\left[\frac{1}{\sqrt{T}} S_{T1}\right] &= \frac{1}{T}\left[\sum_{t=1}^{T} Var(X_{1,t}) + \sum_{\substack{t=1 \\ s\neq t}}^{T} \mathsf{Cov}(X_{1,s},\, X_{1,t})\right] \\ &= \frac{1}{T}\left[T\gamma_1(0) + 2\sum_{\tau=1}^{T}(T-\tau)\gamma_1(\tau)\right] = \gamma_1(0) + 2\sum_{\tau=1}^{T}(1-\frac{\tau}{T})\gamma_1(\tau) , \end{aligned}$$

where $\gamma \equiv r_w^2/(1-a^2)$. We must prove that $\sum_{\tau=1}^{T}(1-\frac{\tau}{T})\gamma_1(\tau)$ converge as $T \to \infty$. By Lemma 3.1.5 in Fuller (1976, page 112), it is sufficient to show that $\sum_{\tau=1}^{\infty} \gamma_1(\tau)$ converge. Using the results of Lemma **1.3.3** we have:

$$\begin{aligned} \gamma_1(\tau) &= \mu_2^2(\theta)[\exp(\gamma a^\tau) - 1] = \mu_2^2(\theta)\left[1 + \sum_{k=1}^{\infty} \frac{(\gamma a^\tau)^k}{k!} - 1\right] = \mu_2^2(\theta)\left[\gamma a^\tau \sum_{k=1}^{\infty} \frac{(\gamma a^\tau)^{k-1}}{k!}\right] \\ &= \mu_2^2(\theta)\left[\gamma a^\tau \sum_{k=0}^{\infty} \frac{(\gamma a^\tau)^k}{(k+1)!}\right] \leq \mu_2^2(\theta)\gamma a^\tau \sum_{k=0}^{\infty} \frac{(\gamma a^\tau)^k}{k!} = \mu_2^2(\theta)\gamma a^\tau \exp(\gamma a^\tau) . \end{aligned}$$

Therefore, the series

$$\begin{aligned}\sum_{\tau=1}^{\infty}\gamma_1(\tau) &\leq \mu_2^2(\theta)\gamma\sum_{\tau=1}^{\infty}a^{\tau}\exp(\gamma a^{\tau}) \leq \mu_2^2(\theta)\gamma\exp(\gamma a)\sum_{\tau=1}^{\infty}a^{\tau} \\ &= \mu_2^2(\theta)\frac{a\gamma\exp(\gamma a)}{1-a} < \infty\end{aligned}$$

converges. We deduce by Cauchy-Schwarz inequality that

$$\limsup_{T\to\infty} T^{-1/2}\mathsf{E}\Big|\sum_{t=1}^{T}\left[u_t^2-\mu_2(\theta)\right]\Big| < \infty\,.$$

The proof is very similar for the second component of S_T. We will skip to the third component of S_T.

ii) The third component of S_T. Set $S_{T3}=\sum_{t=1}^{T}X_{3,t}$ where $X_{3,t}\equiv u_t^2u_{t-1}^2-\mu_{2,2}(1|\theta)$. Likewise, we just have to show that $\sum_{\tau=1}^{\infty}\gamma_3(\tau)<\infty$ in order to prove that

$$\limsup_{T\to\infty} T^{-1/2}\mathsf{E}\Big|\sum_{t=1}^{T}\left[u_t^2u_{t-1}^2-\mu_{2,2}(1|\theta)\right]\Big| < \infty\,.$$

By lemma **1.3.3** we have for all $\tau\geq 2$:

$$\begin{aligned}\gamma_3(\tau) &= \mu_{2,2}^2(1|\theta)[\exp(\gamma(1+a)^2a^{\tau-1})-1] \\ &= \mu_{2,2}^2(1|\theta)\left\{1+\sum_{k=1}^{\infty}\frac{[\gamma(1+a)^2a^{\tau-1}]^k}{k!}-1\right\} \\ &= \mu_{2,2}^2(1|\theta)[\gamma(1+a)^2a^{\tau-1}]\sum_{k=1}^{\infty}\frac{[\gamma(1+a)^2a^{\tau-1}]^{k-1}}{k!} \\ &= \mu_{2,2}^2(1|\theta)[\gamma(1+a)^2a^{\tau-1}]\sum_{k=0}^{\infty}\frac{[\gamma(1+a)^2a^{\tau-1}]^{k}}{(k+1)!} \\ &\leq \mu_{2,2}^2(1|\theta)[\gamma(1+a)^2a^{\tau-1}]\sum_{k=0}^{\infty}\frac{[\gamma(1+a)^2a^{\tau-1}]^{k}}{k!} \\ &= \mu_{2,2}^2(1|\theta)[\gamma(1+a)^2a^{\tau-1}]\exp[\gamma(1+a)^2a^{\tau-1}]\,,\end{aligned}$$

such that

$$\begin{aligned}
\sum_{\tau=1}^{\infty}\gamma_3(\tau) &\leq \gamma_3(1)+\mu_{2,2}^2(1|\theta)\gamma(1+a)^2\sum_{\tau=2}^{\infty}a^{\tau-1}\exp[\gamma(1+a)^2a^{\tau-1}] \\
&\leq \gamma_3(1)+\mu_{2,2}^2(1|\theta)\gamma(1+a)^2\exp[\gamma(1+a)^2a]\sum_{\tau=2}^{\infty}a^{\tau-1} \\
&= \gamma_3(1)+\mu_{2,2}^2(1|\theta)\gamma(1+a)^2\exp[\gamma(1+a)^2a]\sum_{\tau=1}^{\infty}a^{\tau} \\
&= \gamma_3(1)+\mu_{2,2}^2(1|\theta)\gamma(1+a)^2\exp[\gamma(1+a)^2a]\frac{a}{1-a}<\infty\ .
\end{aligned}$$

Since $\limsup_{T\to\infty}T^{-1/2}\mathsf{E}|\sum_{t=1}^{T}X_t|<\infty$ we can therefore apply Theorem 24.5 of Davidson (1994) to each component S_{Ti} , $i=1,\ 2,\ 3$ of S_T to state that: $T^{-1/2}S_{Ti}\xrightarrow{D}\mathrm{N}(0,\ \lambda_i)$ and then by Cramér-Wold theorem establish the limiting result for the 3×1vector S_T using the stability property of the gaussian distribution, *i.e.*,

$$T^{-1/2}S_T=T^{-1/2}\sum_{t=1}^{T}X_t=\sqrt{T}\left[\bar{g}_T(U_T)-\mu(\theta)\right]\xrightarrow{D}\mathrm{N}_3(0,\ \Omega_*)\ ,$$

where

$$\Omega_*=\lim_{T\to\infty}\mathsf{E}[(T^{-1/2}S_T)^2]=\lim_{T\to\infty}\mathsf{E}\left\{T\left[\bar{g}_T(U_T)-\mu(\theta)\right]\left[\bar{g}_T(U_T)-\mu(\theta)\right]'\right\}\ .$$

□

PROOF OF PROPOSITION **1.4.8** The asymptotic equivalence of

$$\sqrt{T}\left[\bar{g}_T(\hat{U}_T)-\mu(\theta)\right]=\begin{pmatrix}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\hat{u}_t^2-\mu_2(\theta)\right]\\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\hat{u}_t^4-\mu_4(\theta)\right]\\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\hat{u}_t^2\hat{u}_{t-1}^2-\mu_{2,2}(1)(\theta)\right]\end{pmatrix}$$

with+ $\sqrt{T}(\bar{g}_T(U_T)-\mu(\theta))$ can be established by looking at each component.

1. **The component** $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\hat{u}_t^2 - \mu_2(\theta))$. We have:

$$\begin{aligned}\hat{u}_t^2 - u_t^2 &= (y_t - x_t'\hat{\beta})^2 - (y_t - x_t'\beta)^2 \\ &= (\hat{\beta}-\beta)'x_t x_t'(\hat{\beta}-\beta) - 2(\hat{\beta}-\beta)'x_t u_t \ .\end{aligned}$$

We deduce after aggregation:

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\hat{u}_t^2 - \mu_2(\theta)] &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[u_t^2 - \mu_2(\theta)] + \frac{1}{\sqrt{T}}[\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T}x_t x_t'\sqrt{T}(\hat{\beta}-\beta) \\ &\quad -2\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T}x_t u_t] \ . \qquad (.0.3)\end{aligned}$$

By (1.4.18), and Assumption **1.4.6**, and by the Law of Large Numbers (L.L.N.) $\frac{1}{T}\sum_{t=1}^{T}x_t u_t \to E(x_t u_t) = E(x_t)E(u_t) = 0$, we deduce that equation (.0.3) is equivalent to

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\hat{u}_t^2 - \mu_2(\theta)) \# \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t^2 - \mu_2(\theta))$$

asymptotically.

2. **The component** $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\hat{u}_t^4 - \mu_4(\theta))$. Noting that

$$\begin{aligned}\hat{u}_t^4 - u_t^4 &= -4(\hat{\beta}-\beta)'x_t u_t^3 + 6(\hat{\beta}-\beta)'x_t u_t^2 x_t'(\hat{\beta}-\beta) - 4(\hat{\beta}-\beta)'x_t x_t'(\hat{\beta}-\beta)(\hat{\beta}-\beta)'x_t u_t \\ &\quad +(\hat{\beta}-\beta)'x_t x_t'(\hat{\beta}-\beta)(\hat{\beta}-\beta)'x_t x_t'(\hat{\beta}-\beta)\end{aligned}$$

we get after aggregation:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\hat{u}_t^4 - \mu_4(\theta)) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t^4 - \mu_4(\theta)) + R_T$$

where

$$R_T \equiv -4\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T}x_t u_t^3 + 6\frac{1}{\sqrt{T}}\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T}x_t u_t^2 x_t'\sqrt{T}(\hat{\beta}-\beta)$$

$$-4\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T}x_t x_t'(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T}x_t u_t$$
$$+\frac{1}{\sqrt{T}}\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T}x_t x_t'\sqrt{T}(\hat{\beta}-\beta)\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T}x_t x_t'\sqrt{T}(\hat{\beta}-\beta)\,.$$

Since $\sqrt{T}(\hat{\beta}-\beta)$, $\frac{1}{T}\sum_{t=1}^{T}x_t u_t^2 x_t'$ and $\frac{1}{T}\sum_{t=1}^{T}x_t x_t'$ are asymptotically bounded, and by the L.L.N., $\frac{1}{T}\sum_{t=1}^{T}x_t u_t^3$, $\frac{1}{T}\sum_{t=1}^{T}x_t u_t$, we can conclude that R_T is an $o_p(1)$-variable which yields that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\hat{u}_t^4-\mu_4(\theta))\#\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t^4-\mu_4(\theta))$$

asymptotically.

3. **The component** $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\hat{u}_t^2\hat{u}_{t-1}^2-\mu_{2,2}(1|\theta))$. Noting that:

$$\begin{aligned}\hat{u}_t^2\hat{u}_{t-1}^2-u_t^2u_{t-1}^2 &= -2(\hat{\beta}-\beta)'[x_t u_t u_{t-1}^2+x_{t-1}u_t^2u_{t-1}]+(\hat{\beta}-\beta)'x_t u_{t-1}^2x_t'(\hat{\beta}-\beta)\\ &\quad+(\hat{\beta}-\beta)'x_{t-1}u_t^2x_{t-1}'(\hat{\beta}-\beta)+4(\hat{\beta}-\beta)'x_t u_t u_{t-1}x_{t-1}'(\hat{\beta}-\beta)\\ &\quad-2(\hat{\beta}-\beta)'x_t x_t'(\hat{\beta}-\beta)(\hat{\beta}-\beta)'x_{t-1}u_{t-1}-2(\hat{\beta}-\beta)'x_{t-1}x_{t-1}'(\hat{\beta}-\beta)(\hat{\beta}-\beta)'x_t u_t\\ &\quad+(\hat{\beta}-\beta)'x_t x_t'(\hat{\beta}-\beta)(\hat{\beta}-\beta)'x_{t-1}x_{t-1}'(\hat{\beta}-\beta)\end{aligned}$$

yields after aggregation

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\hat{u}_t^2\hat{u}_{t-1}^2)-\mu_{2,2}(1|\theta))=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t^2u_{t-1}^2-\mu_{2,2}(1|\theta))+R_T$$

where

$$\begin{aligned}R_T \;\equiv\; &-2\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T}[x_t u_t u_{t-1}^2+x_{t-1}u_t^2u_{t-1}]\\ &+\frac{1}{\sqrt{T}}\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T}x_t u_{t-1}^2x_t'\sqrt{T}(\hat{\beta}-\beta)\\ &+\frac{1}{\sqrt{T}}\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T}x_{t-1}u_t^2x_{t-1}'\sqrt{T}(\hat{\beta}-\beta)+4\frac{1}{\sqrt{T}}\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\end{aligned}$$

$$\sum_{t=1}^{T} x_t u_t u_{t-1} x'_{t-1} \sqrt{T}(\hat{\beta} - \beta)$$
$$-2\sqrt{T}(\hat{\beta} - \beta)' \frac{1}{T} \sum_{t=1}^{T} x_t x'_t \sqrt{T}(\hat{\beta} - \beta) \sqrt{T}(\hat{\beta} - \beta)' \frac{1}{T} \sum_{t=1}^{T} x_{t-1} u_{t-1}$$
$$-2\sqrt{T}(\hat{\beta} - \beta)' \frac{1}{T} \sum_{t=1}^{T} x_{t-1} x'_{t-1} \sqrt{T}(\hat{\beta} - \beta) \sqrt{T}(\hat{\beta} - \beta)' \frac{1}{T} \sum_{t=1}^{T} x_t u_t$$
$$+\frac{1}{\sqrt{T}} \sqrt{T}(\hat{\beta} - \beta)' \frac{1}{T} \sum_{t=1}^{T} x_t x'_t \sqrt{T}(\hat{\beta} - \beta) \sqrt{T}(\hat{\beta} - \beta)' \frac{1}{T} \sum_{t=1}^{T} x_{t-1} x'_{t-1} \sqrt{T}(\hat{\beta} - \beta) .$$

By (1.4.18), Assumption **1.4.6** , and by the L.L.N. applied to $\frac{1}{T} \sum_{t=1}^{T} x_{t-1} u_t^2 u_{t-1}$ which converges to $E[x_{t-1} u_t^2 u_{t-1}] = E[u_t^2 E(u_{t-1}|\mathcal{F}_{t-2}) E(x_{t-1}|\mathcal{F}_{t-2})] = 0$, we deduce that R_T is an $o_p(1)$-variable. Therefore, we have the asymptotic equivalence below.

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\hat{u}_t^2 \hat{u}_{t-1}^2 - \mu_{2,2}(1)(\theta)) \# \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (u_t^2 u_{t-1}^2 - \mu_{2,2}(1)(\theta)) .$$

Thus,

$$\begin{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\hat{u}_t^2 - \mu_2(\theta)) \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\hat{u}_t^4 - \mu_4(\theta)) \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\hat{u}_t^2 \hat{u}_{t-1}^2 - \mu_{2,2}(1)(\theta)) \end{pmatrix} \overset{asy}{\#} \begin{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (u_t^2 - \mu_2(\theta)) \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (u_t^4 - \mu_4(\theta)) \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (u_t^2 u_{t-1}^2 - \mu_{2,2}(1)(\theta)) \end{pmatrix}$$

and from equation (1.3.3), we get the asymptotic equivalence

$$\sqrt{T}(\bar{g}_T(\hat{U}_T) - \mu(\theta)) \overset{asy}{\#} \sqrt{T}(\bar{g}_T(U_T) - \mu(\theta)) ,$$

with $\bar{g}_T(U_T)$ defined as in equation (1.4.11). □

Appendix: Proofs of chapter 3

Analytical expressions of the derivatives

The analytical expressions of the derivatives of the moment conditions defining the binding function are given by:

$$\frac{\partial \mu_2}{\partial a} = \frac{a}{(1-a^2)^2} r_w^2 r_y^2 \exp[\frac{r_w^2}{2(1-a^2)}] ,$$

$$\frac{\partial \mu_2}{\partial r_w} = \frac{r_w}{(1-a^2)} r_y^2 \exp[\frac{r_w^2}{2(1-a^2)}] ,$$

$$\frac{\partial \mu_2}{\partial r_y} = 2 r_y \exp[\frac{r_w^2}{2(1-a^2)}]$$

$$\frac{\partial \mu_4}{\partial a} = 12 \frac{a}{(1-a^2)^2} r_w^2 r_y^4 \exp[\frac{2r_w^2}{(1-a^2)}] ,$$

$$\frac{\partial \mu_4}{\partial r_w} = 12 \frac{r_w}{(1-a^2)} r_y^4 \exp[\frac{2r_w^2}{(1-a^2)}] ,$$

$$\frac{\partial \mu_4}{\partial r_y} = 12 r_y^3 \exp[\frac{2r_w^2}{(1-a^2)}] ,$$

$$\frac{\partial \mu_{2,2}}{\partial a} = \frac{r_w^2}{(1-a)^2} r_y^4 \exp[\frac{r_w^2}{(1-a)}] ,$$

$$\frac{\partial \mu_{2,2}}{\partial r_w} = \frac{2r_w}{1-a} r_y^4 \exp[\frac{r_w^2}{(1-a)}] ,$$

$$\frac{\partial \mu_{2,2}}{\partial r_y} = 4 r_y^3 \exp[\frac{r_w^2}{(1-a)}] .$$

All these derivatives evaluated at $a = 0$, $r_w = 0$ gives the results stated in equation (3.3.23).

Appendix: Proofs of chapter 5

Moments used in the GMM estimation

$$k_1(\theta) \equiv E(Y_t) = \beta \, .$$

$$k_2(\theta) \equiv E(Y_t^2) = -\frac{[(\frac{2b}{c}\beta)^2 + \frac{2b}{c}\beta]}{(\frac{2b}{c})^2 + 3(\frac{2b}{c}) + 2} + 2\frac{[\frac{2b}{c}\beta + 1]}{\frac{2b}{c} + 2} k_1(\theta) \, .$$

$$\begin{aligned} k_3(\theta) \equiv E(Y_t^3) \;\; = \;\; & \frac{[(\frac{2b}{c}\beta)^3 + 3(\frac{2b}{c}\beta)^2 + 2(\frac{2b}{c}\beta)]}{(\frac{2b}{c}+4)(\frac{2b}{c}+3)(\frac{2b}{c}+2)} \\ & -3\frac{[(\frac{2b}{c}\beta)^2 + 3(\frac{2b}{c}\beta) + 2]}{(\frac{2b}{c}+4)(\frac{2b}{c}+3)} k_1(\theta) \\ & +3\frac{\frac{2b}{c}\beta + 2}{\frac{2b}{c}+4} k_2(\theta) \end{aligned}$$

$$\begin{aligned} k_4(\theta) \equiv E(Y_t^4) \;\; = \;\; & -\frac{[(\frac{2b}{c}\beta)^4 + 6(\frac{2b}{c}\beta)^3 + 11(\frac{2b}{c}\beta)^2 + 6(\frac{2b}{c}\beta)]}{(\frac{2b}{c}+6)(\frac{2b}{c}+5)(\frac{2b}{c}+4)(\frac{2b}{c}+3)} \\ & +4\frac{[(\frac{2b}{c}\beta)^3 + 6(\frac{2b}{c}\beta)^2 + 11(\frac{2b}{c}\beta) + 6]}{(\frac{2b}{c}+4)(\frac{2b}{c}+5)(\frac{2b}{c}+6)} k_1(\theta) \\ & -6\frac{[(\frac{2b}{c}\beta)^2 + 5(\frac{2b}{c}\beta) + 6]}{(\frac{2b}{c}+5)(\frac{2b}{c}+6)} k_2(\theta) + 4\frac{\frac{2b}{c}\beta + 3}{\frac{2b}{c}+6} k_3(\theta) \end{aligned}$$

$$k_{11}^{(h)}(\theta) \equiv E[y_t y_{t-h}] = \exp(-bh)k_2(\theta) + [1 - \exp(-bh)]\beta^2$$

$$k_{12}^{(h)}(\theta) \equiv E[y_t y_{t-h}^2] = \exp(-bh)k_3(\theta) + [1 - \exp(-bh)]\beta k_2(\theta)$$

$$k_{21}^{(h)}(\theta) \equiv E[y_t^2 y_{t-h}] = -\frac{[(\frac{2b}{c}\beta)^2 + (\frac{2b}{c}\beta)]}{[(\frac{2b}{c})^2 + 3(\frac{2b}{c}) + 2]}(1 - \exp[(-2b - c)h])k_1(\theta)$$

$$+2\frac{[\frac{2b}{c}\beta^2+\beta]}{\frac{2b}{c}+2}[1-\exp(-bh)]k_1(\theta)+2\frac{\frac{2b}{c}\beta+1}{\frac{2b}{c}+2}\{\exp(-bh)-\exp[(-2b-c)h]\}k_2(\theta)$$
$$+\exp[(-2b-c)h]k_3(\theta)$$

$$k_{22}^{(h)}(\theta)\equiv E[y_t^2y_{t-h}^2]=\frac{-\left[(\frac{2b}{c}\beta)^2+\frac{2b}{c}\beta\right]}{(\frac{2b}{c})^2+3(\frac{2b}{c})+2}(1-\exp[(-2b-c)h])k_2(\theta)$$
$$+2\frac{\left[(\frac{2b}{c})\beta^2+\beta\right]}{\frac{2b}{c}+2}[1-\exp(-bh)]k_2(\theta)+2\frac{\left[(\frac{2b}{c})\beta+1\right]}{\frac{2b}{c}+2}\{\exp(-bh)-\exp[(-2b-c)h]\}k_3(\theta)$$
$$+\exp[(-2b-c)h]k_4(\theta)\ .$$

Identification issue

Our concern here is to identify the parameter of the Jacobi process. Note that taking the unconditional expectation of equation (5.3.2.2), yields under stationarity assumption that:

$$E(Y_t)=\beta=\frac{a_{10}}{1-a_{11}}\equiv m\ .$$

b can be deduced from a_{11} as

$$b=-\frac{\log(a_{11})}{h}$$

where the coefficients a_{11} and a_{10} can be estimated by considering the regression equation

$$Y_t=a_{11}Y_{t-h}+a_{10}+u_{1t}\ .$$

Further, c can be identified from a_{22} since

$$c=-2b-\frac{\log(a_{22})}{h}$$

where a_{22} can be estimated from the regression equation

$$Y_t^2 = a_{22}Y_{t-h}^2 + a_{21}Y_{t-h} + a_{20} + u_{2t} \ . \tag{.0.4}$$

In other words, denoting by $\hat{m}_T$ an estimator of $E(Y_t)$, $\hat{\rho}(h)_T$ an estimator of $Cov(Y_t, Y_{t-h})/Var(Y_{t-h})$ and $\hat{\sigma}_T^2$ an estimator of $Var(Y_t)$, we can deduce that $\hat{\beta}_T = \hat{m}_T$, $\hat{b}_T = -\frac{\log(\hat{\rho}_T(h))}{h}$ since $a_{11} = Cov(Y_t, Y_{t-h})/Var(Y_{t-h}) = \rho(h)$,, $\hat{\sigma}_T^2 = \hat{\sigma}_1^2/(1-\hat{a}_{11}^2)$ where $\sigma_1^2 \equiv Var(u_{1t})$. And finally, $\hat{c} = 2\frac{\log(\hat{\rho}(h))}{h} - \frac{\log(1-\frac{\hat{a}_{21}\hat{m}+\hat{a}_{20}}{\hat{\sigma}^2+\hat{m}^2})}{h}$ where $a_{22} = 1 - \frac{\hat{a}_{21}\hat{m}+\hat{a}_{20}}{\hat{\sigma}^2+\hat{m}^2}$ and a_{21} and a_{20} can be estimated from equation (.0.4).

Computation of the weighting matrix C

The weighting matrix $C(x;\theta) = \{c_{ij}(x;\theta)\}$ has been defined as:

$$C(x;\theta) = E_\theta[h(y,x;\theta)h(y,x;\theta)'|x] \ .$$

Therefore its entries correspond to, where for simplicity of notations $\lambda_i(\theta) \equiv \lambda_i$:

$$\begin{aligned} c_{ij}(x;\theta) &= E_\theta[h_i(y,x;\theta)h_j(y,x;\theta)|x] \\ &= \int_0^1 P_i(y;\theta)P_j(y;\theta)f(y|x;\theta)dy - \exp(\lambda_i)P_i(x;\theta)\int_0^1 P_j(y;\theta)f(y|x;\theta)dy \\ &\quad - \exp(\lambda_j)P_j(x;\theta)\int_0^1 P_i(y;\theta)f(y|x;\theta)dy + \exp(\lambda_i)\exp(\lambda_j)P_i(x;\theta)P_j(x;\theta) \ , \end{aligned}$$

where by exploiting the fact that the polynomials are also eigenfunctions for the conditional operator with eigenvalues $\exp(\lambda_j)$, $c_{ij}(x;\theta)$ is equal to:

$$c_{ij}(x;\theta) = \int_0^1 P_i(y;\theta)P_j(y;\theta)f(y|x;\theta)dy - \exp(\lambda_i)\exp(\lambda_j)P_i(x;\theta)P_j(x;\theta) \ .$$

Finally, since the eigenfunctions are polynomials we have:

$$c_{ij}(x;\theta) = \sum_{r=0}^{i}\sum_{s=0}^{j} a_{i,r}(\theta)a_{j,s}(\theta)\int_0^1 y^{r+s} f(y,|x;\theta)dy - \exp(\lambda_i)\exp(\lambda_j)P_i(x;\theta)P_j(x;\theta)\ .$$

Computation of the integrals

To compute these integrals, we exploit again the fact that the Jacobi polynomials $P_j(y;\theta)$ are also eigenfunctions for the conditional expectation operator associated with the eigenvalues $\exp(\lambda_j(\theta))$ in the following way:

$$\begin{aligned}
\int_0^1 P_j(y;\theta)f(y|x;\theta)dy &= \int_0^1 \sum_{k=0}^{j} a_{j,k}(\theta)y^k f(y|x;\theta)dy \\
\Leftrightarrow E[P_j(y;\theta)|x;\theta] &= \sum_{k=0}^{j} a_{j,k}(\theta)\int_0^1 y^k f(y|x;\theta)dy \\
\Leftrightarrow e^{\lambda_j(\theta)}P_j(x;\theta) &= \sum_{k=0}^{j} a_{j,k}(\theta)\int_0^1 y^k f(y|x;\theta)dy \\
\Leftrightarrow e^{\lambda_j(\theta)}\sum_{k=0}^{j} a_{j,k}(\theta)y^k &= \sum_{k=0}^{j} a_{j,k}(\theta)\int_0^1 y^k f(y|x;\theta)dy\ .
\end{aligned}$$