Estimation of a Jacobi process

Christian Gouriéroux* — CREST
Pascale Valéry † — Université de Montréal

First version: March 2002
This version: November 2002
Compiled: 28th April 2004

ABSTRACT

In this paper 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 unfeasable. We first review different characterizations of the transition function based on nonlinear canonical decomposition, partial differential equations.. They allow for approximations of the log-likelihood function which can be used to define an approximated maximum likelihood estimator. The finite sample properties of this estimator are compared with the properties of Kessler and Sorensen's estimator based on the eigenfunctions of the generator of the diffusion. But also simulation based estimators, such as generalized method of moments (GMM) estimator, simulated method of moments (SMM) estimator or indirect inference estimator are considered.

Résumé Dans cet essai, nous considérons un processus de Jacobi échantillonné en temps discret permettant de spécifier la dynamique de processus à valeurs dans $[0, 1]$, tels qu' un coéfficient d'actualisation, une probabilité de régime, ou un prix d'état. La fonction de transition en temps discret du processus de Jacobi n'admettant pas

* University of Toronto

† Centre interuniversitaire de recherche en économie quantitative (CIREQ), Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Département de sciences économiques, Université de Montréal. Mailing address: Département de sciences économiques, Université de Montréal, C.P. 6128 succursale Centre-ville, Montréal, Québec, Canada H3C 3J7. e-mail: valeryp@magellan.umontreal.ca.

d'expression analytique, le maximum de vraisemblance ne peut alors être déterminé . Nous rappelons dans un premier temps diverses caractérisations de la fonction de transition en nous appuyant sur la décomposition canonique non linéaire, sur les équations différentielles partielles.. Elles induisent des approximations de la fonction de vraisemblance qui peuvent être utilisées afin de définir un estimateur du maximum de vraisemblance approximé. Les propriétés de petit échantillon de cet estimateur sont alors comparées à celles de l'estimateur fondé sur les fonctions propres de Kessler et Sorensen. Nous considérons également des estimateurs obtenus par simulations, tels l'estimateur des moments généralisés (GMM), l'estimateur des moments simulés (SMM) ou encore l'estimateur par inférence indirecte.
Keywords: Jacobi process, nonlinear canonical analysis, diffusion process, subordinated process, method-of-moments, simulation-based estimation methods
JEL classification: C13, C15, C22

1. Introduction

Stochastic differential equations (s.d.e.) often provide a convenient way to describe the dynamics of economic and financial data in continuous time, and a great deal of effort has been expanded searching for efficient ways to estimate these continuous time models from discrete time observations.In this framework, the Jacobi process, solution of the s.d.e. (1.1) below, is characterized by a quadratic specification of its volatility function allowing thereby for larger conditional heteroscedasticity compared with the usual interest rate term-structure models, say, the Vasicek model, the Cox-Ingersoll-Ross model.Furthermore, as this process is restricted to a finite interval, say $[0, 1]$, it naturally arises to model dynamic bounded variables such as a regime probability or a default probability. Indeed, it seems appropriate to capture the evolution of a state variable underlying a regime-shift model or a credit risk model.

The aim of the paper is to estimate the parameters of the Jacobi process and to compare the statistical properties of various estimators through Monte Carlo experiments. Although the process specified by a s.d.e. is defined in continuous time, the different estimators examined in this paper handle with discrete time sampled data. Unfortunately, the likelihood function for discrete observations generated by continuous-time diffusion processes is in most contexts unknown. Various estimation methods have been developpped during the last decades to circumvent this issue. Some of these methods are based on simulations [see Gouriéroux, Monfort and Renault (1993), Gallant and Tauchen (1996)], others on the generalized method of moments [see Hansen and Scheinkman (1995), Duffie and Glynn (1997)] others on martingale estimating functions [see Kessler and Sorensen (1999), Larsen and Sorensen (2003)], others on nonparametric density-matching techniques [see Aït-Sahalia (1996a, 1996b)], on nonparametric regression for approximate moments [see Stanton (1997)] or finally on bayesian techniques [see Eraker (1997), Jones (1997)].

But it is of common knowledge that the maximum likelihood approach is the best estimation method under a correct specification of the model. Except for a few models in finance [see Black and Scholes (1973), Vasicek (1978), Cox, Ingersoll and Ross (1985) and Cox (1975)] the likelihood function is not available in closed form and therefore the maximum likelihood is unfeasable. The solutions proposed in the literature to compute the unknown likelihood function involve either solving numerically the Fokker-Planck-Kolmogorov partial differential equation [see e.g. Lo (1988)] or by simulations [see Pedersen (1995), Santa-Clara (1995)]. But neither method produces a closed-form expression to be maximized over the parameter of interest. By contrast, Aït-Sahalia(2002) constructs a closed-form sequence of approximations to the original diffusion process in order to get close enough to a $N(0, 1)$ variable. The conditional density of the transformed process can then be expanded through a Hermite series expansion. Finally, by applying the Jacobian formula twice, Aït-Sahalia

(2002) backtrackes and infers a sequence of approximations of the transition function to the original diffusion process. Since the expression of the approximated likelihood function is explicit, the maximum likelihood approach becomes feasable.

Likewise, we develop in this paper a technique based on nonlinear canonical analysis to approximate the unknown discrete-time transition function of continuous-time parametric diffusion processes and apply it to the Jacobi process. The approximation technique consists on truncating the spectral decomposition of the transition density derived from the spectral decomposition of the infinitesimal generator associated with the diffusion process. Our technique to approximate the transition function is much simpler than that of Aït-Sahalia and is much closer tailored to the genuine diffusion process since we do not proceed with any transformation. The maximization of this approximated likelihood function over the parameter of interest yields the approximated maximum likelihood estimator denoted by AML. The finite sample properties of this estimator are then compared with the properties of the Kessler and Sorensen's estimator based on eigenfunctions of the infinitesimal generator, together with simulation-based estimators such as generalized method of moments (GMM) estimator, simulated method of moments (SMM) estimator, or indirect inference (II) estimator and also with an exact indirect (EI) estimator based on an identifying constraint.

More specifically, let y_t denote the Jacobi process solution of the following stochastic differential equation.

$$dy_t = -b(y_t - \beta)dt + \sqrt{cy_t(1-y_t)}dW_t \ , \tag{1.1}$$

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 type of continuous model is appropriate for describing the evolution of a variable between 0 and 1, such as a regime probability, a discount coefficient or a state price.

The paper is organized as follows. The next section exploits the subordination properties to induce a Jacobi process, and its distributional properties are analyzed by means of nonlinear canonical analysis. Section 3 reviews the estimation methods which can be applied to estimate the parameters of the Jacobi process. Theoretically the maximum likelihood approach provides asymptotically efficient estimators. However the transition density in discrete time has no closed form and thus the implementation of maximum likelihood approach is unfeasible. Hence, an approximated maximum likelihood method is implemented based on the spectral decomposition of the transition density. Two other consistent estimation approaches can be followed that are generalized-method-of-moments or simulation-based methods. These various principles of estimation are applied to the discretely Jacobi process to compare the statistical performance of the estimators through Monte Carlo experiments. How to simulate the Jacobi process is detailed in the fourth section before presenting the Monte Carlo results in section five. Finally concluding remarks are given in section 6. The proofs are gathered in appendices.

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)].

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. We show that 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:

$$\begin{cases} dx_{1t} = -b(x_{1t} - \beta_1)dt + \sqrt{cx_{1t}}dW_{1t} \ , \\ dx_{2t} = -b(x_{2t} - \beta_2)dt + \sqrt{cx_{2t}}dW_{2t} \ , \end{cases} \tag{2.2}$$

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:

$$\begin{cases} dy_{1t} = -(b/y_{2t})[y_{1t}(\beta_1 + \beta_2) - \beta_1]dt + [c(y_{1t}/y_{2t})(1 - y_{1t})]^{1/2}d\tilde{W}_{1t} \ , \\ dy_{2t} = -b[y_{2t} - (\beta_1 + \beta_2)]dt + \sqrt{cy_{2t}}d\tilde{W}_{2t} \ , \end{cases} \tag{2.3}$$

where $(\tilde{W}_{1t})$ and $(\tilde{W}_{2t})$ are standard brownian motions with variance-covariance matrix $\tilde{\Sigma}_{12}$ such that:

$$\tilde{\Sigma}_{12} = \begin{pmatrix} c\frac{y_{1t}}{y_{2t}}(1-y_{1t}) & 0 \\ 0 & cy_{2t} \end{pmatrix} . \tag{2.4}$$

In particular the two brownian motions $\tilde{W}_1$, $\tilde{W}_2$ are independent. Therefore the process (y_{2t}) is a CIR process with parameters b, $\beta_1+\beta_2$ and c, and (y_{1t}) is a Jacobi process after time deformation. Indeed, let us define the time deformed process:

$$y^*_{1t} = y_{1\tau_t} , \tag{2.5}$$

where the time deformation :

$$\tau_t = \int_0^t y_{2u} du , \tag{2.6}$$

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

$$dy^*_{1t} = -b(\beta_1+\beta_2)[y^*_{1t} - \frac{\beta_1}{\beta_1+\beta_2}]dt + [cy^*_{1t}(1-y^*_{1t})]^{1/2} d\tilde{W}^*_{1t} , \tag{2.7}$$

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.

2.2. Canonical decomposition

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 is defined by:

$$\mathcal{A}P(y) = \lim_{h\to 0} \frac{1}{h} E[P(y_{t+h}) - P(y_t) | y_t = y] . \tag{2.8}$$

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:

$$\mathcal{A}P(y) = \mu(y)\frac{\partial P}{\partial y}(y) + \frac{1}{2}\sigma^2(y)\frac{\partial^2 P}{\partial y^2}(y) , \tag{2.9}$$

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

$$\mathcal{A}P(y) = -b(y-\beta)\frac{\partial P}{\partial y}(y) + \frac{1}{2}cy(1-y)\frac{\partial^2 P}{\partial y^2}(y) \; . \tag{2.10}$$

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:

$$\mathcal{A}P_n = \lambda_n P_n \quad ,\forall\, n, \tag{2.11}$$

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), 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:

$$P_n(y_t) = \left[\frac{\Gamma(\tilde{\alpha}+n)(2n+\tilde{\alpha}+\tilde{\beta}-1)\Gamma(\tilde{\alpha})\Gamma(\tilde{\beta})}{n!\Gamma(\tilde{\alpha}+\tilde{\beta}+n-1)\Gamma(\tilde{\alpha}+\tilde{\beta})\Gamma(\tilde{\beta}+n)}\right]^{1/2} \sum_{m=0}^{n}(-1)^m \binom{n}{m} \frac{\Gamma(\tilde{\alpha}+\tilde{\beta}+n+m-1)}{\Gamma(\tilde{\alpha}+m)} y_t^m \tag{2.12}$$

with $\tilde{\alpha} = \frac{2b}{c}\beta$ and $\tilde{\beta} = \frac{2b}{c}(1-\beta)$. These polynomials define an orthonormal set 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 have been standardized with respect to the beta distribution $\beta(\frac{2b\beta}{c}, \frac{2b(1-\beta)}{c})$, $[E(P(y)) = 0$ and $V(P(y)) = 1]$.

2.2.2. The conditional expectation operator

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

$$\mathcal{T}P(y) = E[P(y_{t+1})|y_t = y] \; . \tag{2.13}$$

It can be seen that 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 $P_n, \;\; n \in \mathbb{N}\backslash\{0\}$.

2.2.3. Moment conditions

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

$$E[P_n(y_t)|y_{t-1}] = \exp(\lambda_n)P_n(y_{t-1}) \;\; \forall \, n \in \mathbb{N}\backslash\{0\} \; , \tag{2.14}$$

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

$$E[P_n(y_t)|y_{t-h}] = \exp(\lambda_n h)P_n(y_{t-h}) \;\; \forall h \; , \; n \in \mathbb{N}\backslash\{0\} \; . \tag{2.15}$$

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

$$E\left[\begin{pmatrix} 1 \\ y_t \\ y_t^2 \\ \vdots \\ y_t^n \end{pmatrix} |y_{t-h}\right] = A^{-1} diag \begin{bmatrix} \exp(\lambda_0 h) \\ \exp(\lambda_1 h) \\ \exp(\lambda_2 h) \\ \vdots \\ \exp(\lambda_n h) \end{bmatrix} A \begin{bmatrix} 1 \\ y_{t-h} \\ y_{t-h}^2 \\ \vdots \\ y_{t-h}^n \end{bmatrix} \; , \tag{2.16}$$

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

$$\begin{bmatrix} P_0(y_t) \\ P_1(y_t) \\ P_2(y_t) \\ \vdots \\ P_n(y_t) \end{bmatrix} = A \begin{bmatrix} 1 \\ y_t \\ y_t^2 \\ \vdots \\ y_t^n \end{bmatrix} \; . \tag{2.17}$$

Matrix A is lower triangular:

$$A = \begin{pmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ a_{10} & a_{11} & 0 & \dots & \dots & 0 \\ a_{20} & a_{21} & a_{22} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & 0 & 0 \\ a_{n0} & a_{n1} & a_{n2} & \dots & \dots & a_{nn} \end{pmatrix} \; , \tag{2.18}$$

with coefficients given by equation (2.12).

Explicit expressions of the conditional power moments can be found by solving recursively the system of moment conditions (2.15):

$$\sum_{m=0}^{n}(-1)^m\binom{n}{m}\frac{\Gamma(\tilde{\alpha}+\tilde{\beta}+n+m-1)}{\Gamma(\tilde{\alpha}+m)}E[y_t^m|y_{t-h}] = \exp(\lambda_n h)\sum_{m=0}^{n}(-1)^m\binom{n}{m}\frac{\Gamma(\tilde{\alpha}+\tilde{\beta}+n+m-1)}{\Gamma(\tilde{\alpha}+m)}y_{t-h}^m . \tag{2.19}$$

For instance we have:

$$E[y_t|y_{t-h}] = [1-\exp(-bh)]\frac{\tilde{\alpha}}{\tilde{\alpha}+\tilde{\beta}} + \exp(-bh)y_{t-h} , \tag{2.20}$$

$$\begin{aligned} E[y_t^2|y_{t-h}] &= -\frac{\tilde{\alpha}(\tilde{\alpha}+1)}{(\tilde{\alpha}+\tilde{\beta}+1)(\tilde{\alpha}+\tilde{\beta}+2)}[1-\exp[(-2b-c)h]] + \frac{2(\tilde{\alpha}+1)}{(\tilde{\alpha}+\tilde{\beta}+2)}[1-\exp(-bh)]\frac{\tilde{\alpha}}{\tilde{\alpha}+\tilde{\beta}} \\ &+\frac{2(\tilde{\alpha}+1)}{(\tilde{\alpha}+\tilde{\beta}+2)}\{\exp(-bh)-\exp[(-2b-c)h]\}y_{t-h} + \exp[(-2b-c)h]y_{t-h}^2 , \end{aligned} \tag{2.21}$$

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 (2.19). For instance, we get:

$$\begin{aligned} E[y_t^2 y_{t-h}^2] &= \frac{-\left[(\frac{2b}{c}\beta)^2+\frac{2b}{c}\beta\right]}{(\frac{2b}{c})^2+3(\frac{2b}{c})+2}(1-\exp[(-2b-c)h])k_2(\theta) \\ &+2\frac{\left[(\frac{2b}{c})\beta^2+\beta\right]}{\frac{2b}{c}+2}[1-\exp(-bh)]k_2(\theta) \\ &+2\frac{\left[(\frac{2b}{c})\beta+1\right]}{\frac{2b}{c}+2}\{\exp(-bh)-\exp[(-2b-c)h]\}k_3(\theta)+\exp[(-2b-c)h]k_4(\theta) , \end{aligned} \tag{2.22}$$

where $k_i(\theta) \equiv E_\theta(y_t^i)$ denotes the marginal power moment of degree i [see Appendix for analytical expressions of $k_i(\theta)$ and of the cross moments].

2.3. Conditional distribution 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) **An expression of the transition based on time deformation.**

The joint conditional distribution of y_{1t}, y_{2t} given $y_{1,t-h}, y_{2,t-h}$is a mixture of a noncentral chi square distribution $\chi^2(K_1 + K_2, \xi_h y_{2t-h})$[1] for y_{2t} and a beta distribution $\beta(\frac{K_1}{2} + i_1, \frac{K_2}{2} + i_2)$[2] for y_{1t} with binomial weights of the form $\mathcal{B}(i_1 + i_2, y_{1t-h})$[3], with $K_i = \frac{4b\beta_i}{c}$ and $\xi_h = \frac{4b}{c}(\exp(bh) - 1)^{-1}$ for $i = 1, 2$. Thus, the joint conditional probability density function (p.d.f) is given by

$$g_{12}^{(h)}(y_{1t}, y_{2t}|y_{1t-h}, y_{2t-h}) = \left[\frac{c}{4b}(1-\exp(-bh))\right]^2 I_{(0,1)\times\mathbb{R}^+}(y_{1t}, y_{2t}) \exp(-\xi_h \frac{y_{2t-h}}{2}) \exp(-\frac{y_{2t}}{2})$$
$$\frac{(y_{2t})^{\frac{K_1+K_2}{2}-1}}{2^{\frac{K_1+K_2}{2}}} \sum_{i_1=0}^{\infty} \sum_{i_2=0}^{\infty} \frac{(y_{2t})^{i_1+i_2}}{(i_1+i_2)!} \frac{(\xi_h y_{2t-h})^{i_1+i_2}}{2^{2(i_1+i_2)}} \frac{1}{\Gamma(\frac{K_1+K_2}{2} + i_1 + i_2)}$$
$$\frac{1}{B(\frac{K_1}{2} + i_1, \frac{K_2}{2} + i_2)} (y_{1t})^{\frac{K_1}{2}+i_1-1} (1 - y_{1t})^{\frac{K_2}{2}+i_2-1} \binom{i_1 + i_2}{i_1} (y_{1t-h})^{i_1} (1 - y_{1t-h})^{i_2} . \tag{2.23}$$

Integrating the bivariate conditional distribution $g_{12}^{(h)}(y_{1t}, y_{2t}|y_{1t-h}, y_{2t-h})$ given at equation (2.23) w.r.t the whole path of process (y_{2t}) yields the conditional density of the process (y_{1t}).

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)]:

$$g^{(h)}(y_t|y_{t-h}) = g(y_t)\left\{1 + \sum_{n=1}^{\infty} \exp(\lambda_n h) P_n(y_t) P_n(y_{t-h})\right\}, \tag{2.24}$$

[1] The density of a noncentral chi-square distribution $\chi^2(K, \lambda)$ is : $I_{\mathbb{R}^+}(y) \exp(-\lambda/2) \exp(-y/2) \frac{y^{K/2-1}}{2^{K/2}} \sum_{i=0}^{\infty} \frac{y^i \lambda^i}{i! 2^{2i}} \frac{1}{\Gamma(i+K/2)}$.

[2] The density of a beta distribution $\beta(p, q)$ is: $\frac{1}{B(p,q)} t^{p-1}(1-t)^{q-1} I_{[0,1]}(t)$.

[3] The binomial distribution $\mathcal{B}(n, p)$ is defined by: $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$.

where $\lambda_n = -bn - \frac{1}{2}cn(n-1)$, P_n are the orthonormal polynomials defined at equation (2.12) and $g(y_t)$ denotes the marginal distribution of y_t [see iv)].

iii) **The transition density as the solution of a partial differential equation.**

Let $g^{(h)}(y_t|y_{t-h})$ denote the transition density function of the Jacobi process at horizon h. It is known [see Friedman (1975), Wong and Hajek (1985), ch.4, section 7, p.172-173, Lo(1988), p.235] that $g^{(h)}(y_t|y_{t-h})$ is solution of a functional partial differential equation, (called the backward diffusion equation)[4] [see Appendix]:

$$\begin{aligned}\frac{\partial}{\partial h} g^{(h)}(y_t|y_{t-h}) &= -b(y_{t-h}-\beta)\frac{\partial}{\partial y_{t-h}} g^{(h)}(y_t|y_{t-h}) \\ &\quad + \frac{1}{2}cy_{t-h}(1-y_{t-h})\frac{\partial^2}{\partial y_{t-h}^2} g^{(h)}(y_t|y_{t-h}) \ ,\end{aligned} \tag{2.25}$$

with $t > t-h$, subject to the terminal condition

$$\lim_{h\to 0} g^{(h)}(y_t|y_{t-h}) = \delta(y_t - y_{t-h}) \ . \tag{2.26}$$

Such a partial differential equation can be solved numerically.

iv) **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 2.1]. Since $y^*_{1t} = y_{1,\tau_t}$ where y_1 and τ are independent processes, we see that 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. The result follows.

3. Estimation methods

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

[4] The notation $\frac{\partial}{\partial h} g^{(h)}(y_t|y_{t-h})$ denotes the derivation w.r.t. the sole h of $g^{(h)}$.

3.1. Approximated maximum likelihood

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

$$\hat{\theta}_T^{ML} = \arg\max_{\theta} \sum_{t=1}^{T} \log g(y_t|y_{t-1};\theta) \ , \tag{3.27}$$

where $g(y_t|y_{t-1};\theta)$ denotes the transition density at horizon 1. It is conditional to 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 unfeasible. However, it is possible to approximate the likelihood along the lines described below.

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

$$\tilde{g}_N(y_t|y_{t-1};\theta) = g(y_t;\theta)\{1 + \sum_{n=1}^{N} \exp(\lambda_n(\theta))P_n(y_t;\theta)P_n(y_{t-1};\theta)\} \ , \tag{3.28}$$

for a large value of N, since

$$\lim_{N\to\infty} \tilde{g}_N(y_t|y_{t-1};\theta) = g(y_t|y_{t-1};\theta) \tag{3.29}$$

for each $\theta \in \Theta$, where $g(y_t|y_{t-1};\theta)$ has been defined at equation (2.24) and $g(y_t)$ denotes the marginal p.d.f of the process which corresponds to a Beta distribution for the Jacobi process.

Then we can define the approximated maximum likelihood estimator $\hat{\theta}_{TN}^{AML}$ of θ as:

$$\hat{\theta}_{TN}^{AML} = \arg\max_{\theta} \sum_{t=1}^{T} \log \tilde{g}_N(y_t|y_{t-1};\theta) \ . \tag{3.30}$$

This estimator based on a truncated version of the likelihood will be asymptotically equivalent to the ML estimator when N tends to infinity with T at an appropriate rate [The regularity conditions are discussed in Appendix]. This rate depends on the rate of decrease of nonlinear correlations. Finally note that the truncated canonical decomposition $g_N(y_t|y_{t-1};\theta)$ is not necessarily positive, when N is fixed. This can create numerical problems in the optimization 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 g_N by its absolute value in the expression of the log-likelihood function.

3.2. Method of moments

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

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 bring the same information as the score (due to the special canonical decomposition of a 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 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:

$$
\begin{array}{llll}
E_0 y_t = k_1(\theta_0), & E_0 y_t y_{t-1} = k_{11}(\theta_0), & & \\
E_0 y_t^2 = k_2(\theta_0), & E_0 y_t^2 y_{t-1} = k_{21}(\theta_0), & E_0 y_t^2 y_{t-1}^2 = k_{22}(\theta_0), & \\
E_0 y_t^3 = k_3(\theta_0), & E_0 y_t^3 y_{t-1} = k_{31}(\theta_0), & E_0 y_t^3 y_{t-1}^2 = k_{32}(\theta_0), & E_0 y_t^3 y_{t-1}^3 = k_{33}(\theta_0) \\
E_0 y_t^N = k_N(\theta_0), & E_0 y_t^N y_{t-1} = k_{N1}(\theta_0), & E_0 y_t^N y_{t-1}^2 = k_{N2}(\theta_0), \ \ldots & E_0 y_t^N y_{t-1}^N = k_{NN}(\theta_0)\ .
\end{array}
\tag{3.31}
$$

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

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:

$$E[y_t|y_{t-1}] = [1-\exp(-b)]\beta + \exp(-b) y_{t-1}\ , \tag{3.32}$$

and

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

$$(3.33)$$

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

$$E[y_t|y_{t-1}] = a_{11}y_{t-1} + a_{10} \tag{3.34}$$
$$E[y_t^2|y_{t-1}] = a_{22}y_{t-1}^2 + a_{21}y_{t-1} + a_{20}\ . \tag{3.35}$$

Thus the parameter of interest θ can be identified from these two conditional moments if the mapping $\theta \to (a_{11}, a_{10}, a_{22}, a_{21}, a_{20})$ is a one-to-one mapping. It is shown in Appendix that this identification condition is satisfied. More precisely the parameter 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[5].

3.2.3. An exact indirect estimator

An 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 $\hat{a}_T = [\hat{m}_T, \hat{\sigma}_T^2, \hat{\rho}_T(1)]'$ 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:

$$\hat{a}_T = a[\hat{\theta}_T^{EI}]\ . \tag{3.36}$$

3.2.4. Generalized-method-of-moments estimator

The summary statistics m, σ^2, $\rho(1,\theta)$ 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

[5] 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 vaues $-\ln(\hat{\rho}_T(h))/h$, h varying, can be the basis of a specification test for the jacobi hypothesis.

by considering a larger set of moments and applying GMM. We consider below a set of moments also including the 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 (3.31) can be rewritten under a vector form as:

$$E_0[K(y_t) - k(\theta_0)] = 0 \; . \tag{3.37}$$

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

$$K(y_t) - k(\theta) = \begin{pmatrix} y_t - k_1(\theta) \\ y_t y_{t-1} - k_{11}^{(1)}(\theta) \\ y_t^2 - k_2(\theta) \\ y_t y_{t-1}^2 - k_{12}^{(1)}(\theta) \\ y_t^2 y_{t-1} - k_{21}^{(1)}(\theta) \\ y_t^2 y_{t-1}^2 - k_{22}^{(1)}(\theta) \\ y_t^3 - k_3(\theta) \\ y_t^4 - k_4(\theta) \end{pmatrix} . \tag{3.38}$$

The GMM estimator is defined by:

$$\hat{\theta}_T^{GMM}(\Omega) = \arg\min_\theta \left(\sum_{t=1}^T [K(y_t) - k(\theta)] \right)' \hat{\Omega}^{-1} \left(\sum_{t=1}^T [K(y_t) - k(\theta)] \right) . \tag{3.39}$$

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

$$\hat{\Omega} = \hat{\Gamma}_0 + \sum_{k=1}^K \left(1 - \frac{k}{K+1} \right) (\hat{\Gamma}_k + \hat{\Gamma}_k') \; , \tag{3.40}$$

where

$$\hat{\Gamma}_k = \frac{1}{T} \sum_{t=k+1}^T [K(y_{t-k}) - k(\tilde{\theta}_T)][K(y_t) - k(\tilde{\theta}_T)]' \tag{3.41}$$

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

3.2.5. Estimating equations based on eigenfunctions

As an alternative to the maximum likelihood approach aforementioned, martingale estimating functions based on eigenfunctions for the generator of the diffusion model provide simple and well-behaved estimators for diffusion models whose likelihood is unknown. This is the technique proposed by Kessler and Sorensen (1999) and applied by Larsen and Sorensen (2003) for fitting a Jacobi process to exchange rates data. This estimator based on eigenfunctions follow the same route via spectral theory as our AML estimator. In this view, it is natural to compare the statistical properties of our AML estimator to the one adopted by Kessler and *alii*.

The optimal estimating function based on the first N eigenfunctions is [see Kessler and Sorensen (1999), Larsen and Sorensen (2003)]:

$$G_T^*(\theta) = \sum_{t=1}^{T} B(y_{t-1};\theta)C(y_{t-1};\theta)^{-1}h(y_t, y_{t-1};\theta) \ , \tag{3.42}$$

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

$$h_j(y, x;\theta) = P_j(y;\theta) - \exp(\lambda_j(\theta))P_j(x;\theta) \ , \tag{3.43}$$

where $P_j(y;\theta) = \sum_{i=0}^{j} a_{j,i}(\theta)y^i$, $\lambda_j(\theta) = -bj - \frac{1}{2}cj(j-1)$, $j = 1, \dots, N$ are the Jacobi polynomials and the associated eigenvalues, respectively. $B(x;\theta) = \{b_{ij}(x;\theta)\}$ is the $p \times N-$matrix, p is the dimension of θ, with entries:

$$b_{ij}(x;\theta) = \sum_{k=0}^{j} \partial_{\theta_i} a_{j,k}(\theta) \int_0^1 y^k g(y, x;\theta)dy - \partial_{\theta_i}(e^{\lambda_j(\theta)}P_j(x;\theta)) \ , \tag{3.44}$$

and $C(x;\theta) = \{c_{ij}(x;\theta)\}$ is the $N \times N-$matrix with entries:

$$c_{ij}(x;\theta) = \sum_{r=0}^{i}\sum_{s=0}^{j} a_{i,r}(\theta)a_{j,s}(\theta) \int_0^1 y^{r+s} g(y, x;\theta)dy - e^{\lambda_i(\theta)+\lambda_j(\theta)}P_i(x;\theta)P_j(x;\theta) \ . \tag{3.45}$$

Here the mapping $y \mapsto g(y, x; \theta)$ denotes the transition density, i.e. the conditional density of y_t given $y_{t-1} = x$. The weight matrices B and C correspond to the optimal choice in the sense of Godambe and Heyde (1987) [see also Heyde (1997)]. The motivation for this specific standardization of the estimating function (3.42) in order to get optimality is the following. First consider the set $\mathcal{G}_0$ of unbiased estimating function G of the form

$$\begin{aligned} G &= \sum_{t=1}^{T} A(y_{t-1};\theta)h(y_t, y_{t-1};\theta) \\ &= \sum_{t=1}^{T}\sum_{j=1}^{N} \alpha_j(y_{t-1};\theta)h_j(y_t, y_{t-1};\theta) \ , \end{aligned} \tag{3.46}$$

where $A(x;\theta)$ is a $p \times N$-matrix with entries $\alpha_j(x;\theta)$. One possible standardization (which is necessary if variances are to be compared) is to define the standardized version of G as

$$G^{(s)} = -E(\dot{G})(E(G^2))^{-1}G \ , \tag{3.47}$$

where the dot denotes the derivative w.r.t. θ, and the expectation is computed w.r.t. the conditional density of y_t given y_{t-1}, i.e. $g(y_t|y_{t-1})$. In order to be used as an estimating equation, the estimating function (3.46) needs to be as close to zero as possible when θ is the true value. Thus, the optimality property requires $var(G)$ ($= E(G^2)$, since $EG = 0$) to be as small as possible. On the other hand, $G(\theta + \delta\theta)$, $\delta > 0$, has to differ as much as possible from $G(\theta)$ when θ is the true value. That is, $(E[\dot{G}(\theta)])^2$, has to be as large as possible. These requirements are combined by maximizing $var(G^{(s)}) = \frac{(E[\dot{G}(\theta)])^2}{EG^2}$. Furthermore, the standardized version $G^{(s)}$ of G possesses the standard likelihood score property

$$\begin{aligned} var(G^{(s)}) &= \frac{(E[\dot{G}(\theta)])^2}{EG^2} \\ &= -E(\dot{G}^{(s)}) \ . \end{aligned} \tag{3.48}$$

Furthermore, the unknown score function of the Jacobi process will be denoted by

$$U = \frac{\partial \log L}{\partial \theta} = \sum_{t=1}^{T} \frac{\partial \log g(y_t|y_{t-1};\theta)}{\partial \theta} \ . \tag{3.49}$$

Hence, we compute

$$E(UG) = \sum_{t=1}^{T} A(y_{t-1};\theta)E\left[\frac{\partial \log g(y_t|y_{t-1};\theta)}{\partial \theta}h(y_t, y_{t-1};\theta)\right] \tag{3.50}$$

that is

$$E(UG) = \sum_{t=1}^{T}\sum_{j=1}^{N}\alpha_j(y_{t-1};\theta)\Bigg\{E\left[\frac{\partial \log g(y_t|y_{t-1};\theta)}{\partial\theta}P_j(y_t;\theta)\right] - E\left[\frac{\partial \log g(y_t|y_{t-1};\theta)}{\partial\theta}\exp(\lambda_j(\theta)) P_j(y_{t-1};\theta)\right]\Bigg\}\,. \tag{3.51}$$

If the transition function $g(y_t|y_{t-1};\theta)$ can be upper bounded by an integrable function, then, by the dominated convergence theorem, the integration and differentiation operators can be interchanged, such that

$$\begin{aligned}
E(UG) &= \sum_{t=1}^{T}\sum_{j=1}^{N}\alpha_j(y_{t-1};\theta)\Bigg\{E\left(\frac{\frac{\partial g(y_t|y_{t-1};\theta)}{\partial\theta}}{g(y_t|y_{t-1};\theta)}P_j(y_t;\theta)\right) - E\left(\frac{\frac{\partial g(y_t|y_{t-1};\theta)}{\partial\theta}}{g(y_t|y_{t-1};\theta)}\exp(\lambda_j(\theta))P_j(y_{t-1};\theta)\right) \\
&= \sum_{t=1}^{T}\sum_{j=1}^{N}\alpha_j(y_{t-1};\theta)\Bigg\{\int_0^1 \frac{\partial g(y_t|y_{t-1};\theta)}{\partial\theta}\sum_{i=0}^{j}a_{ji}(\theta)y_t^i dy_t \\
&\quad - \int_0^1 \frac{\partial g(y_t|y_{t-1};\theta)}{\partial\theta}\exp(\lambda_j(\theta))\sum_{i=0}^{j}a_{ji}(\theta)y_{t-1}^i dy_t\Bigg\} \\
&= \sum_{t=1}^{T}\sum_{j=1}^{N}\alpha_j(y_{t-1};\theta)\Bigg\{\sum_{i=0}^{j}\int_0^1 \frac{\partial g(y_t|y_{t-1};\theta)}{\partial\theta}a_{ji}(\theta)y_t^i dy_t \\
&\quad - \sum_{i=0}^{j}\int_0^1 \frac{\partial g(y_t|y_{t-1};\theta)}{\partial\theta}\exp(\lambda_j(\theta))a_{ji}(\theta)y_{t-1}^i dy_t\Bigg\} \\
&= \sum_{t=1}^{T}\sum_{j=1}^{N}\alpha_j(y_{t-1};\theta)\Bigg\{\sum_{i=0}^{j}\frac{\partial}{\partial\theta}a_{ji}(\theta)\int_0^1 y_t^i g(y_t|y_{t-1};\theta)dy_t \\
&\quad - \sum_{i=0}^{j}\frac{\partial}{\partial\theta}\Big(\exp(\lambda_j(\theta))a_{ji}(\theta)y_{t-1}^i\Big)\underbrace{\int_0^1 g(y_t|y_{t-1};\theta)dy_t}_{1}\Bigg\} \\
&= \sum_{t=1}^{T}\sum_{j=1}^{N}\alpha_j(y_{t-1};\theta)\Bigg\{\sum_{i=0}^{j}\frac{\partial}{\partial\theta}a_{ji}(\theta)\int_0^1 y_t^i g(y_t|y_{t-1};\theta)dy_t \\
&\quad - \frac{\partial}{\partial\theta}\Big(\exp(\lambda_j(\theta))P_j(y_{t-1};\theta)\Big)\Bigg\} \\
&= \sum_{t=1}^{T}\sum_{j=1}^{N}\alpha_j(y_{t-1};\theta)E[\dot{h}_j(y_t,y_{t-1};\theta)] = E\dot{G}\,.
\end{aligned}$$

(3.52)

Therefore,

$$\begin{aligned} corr(U,G) &= (E(UG))^2/(EU^2)(EG^2) \\ &= (var(G^{(s)}))/EU^2 \ , \end{aligned} \tag{3.53}$$

which is maximized if $var(G^{(s)})$ is maximized. That is, the choice of an optimal estimating function $G^* \in \mathcal{G}_0$ is giving an element of $\mathcal{G}_0$ that has maximum correlation with the generally unknown score function.

The coefficients $\{b_{ij}(x;\theta)\}$ and $\{c_{ij}(x;\theta)\}$ of the weight matrices B and C in equation (3.42), respectively, require the computation of the integrals of the type $\int_0^1 y^k g(y,x;\theta)dy$, for $k = 0,1,2,\ldots,2N$. To compute these integrals, we exploit the fact that the Jacobi polynomials $P_j(y;\theta)$ are also eigenfunctions for the conditional expectation operator associated with the eigenvalues $\exp(\lambda_j(\theta))$, i.e.:

$$\int_0^1 P_j(y;\theta)g(y,x;\theta)dy = \int_0^1 \sum_{i=0}^{j} a_{j,i}(\theta)y^i g(y,x;\theta)dy \tag{3.54}$$

$$\Leftrightarrow E[P_j(y;\theta)|x;\theta] = \sum_{i=0}^{j} a_{j,i}(\theta) \int_0^1 y^i g(y,x;\theta)dy \tag{3.55}$$

$$\Leftrightarrow e^{\lambda_j(\theta)} P_j(x;\theta) = \sum_{i=0}^{j} a_{j,i}(\theta) \int_0^1 y^i g(y,x;\theta)dy \tag{3.56}$$

$$\Leftrightarrow e^{\lambda_j(\theta)} \sum_{i=0}^{j} a_{j,i}(\theta)y^i = \sum_{i=0}^{j} a_{j,i}(\theta) \int_0^1 y^i g(y,x;\theta)dy \ . \tag{3.57}$$

We compute these integrals recursively, that is for $j = 1,2,\ldots,2N$ using the fact that $\int_0^1 g(y,x;\theta)dy = 1$ 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 (3.42) of $p = 3$ equations. Kessler and Sorensen (1999) showed that for N going to infinity, the optimal estimating function of the type (3.42) will converge to the score function.

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 4.1 for the description of the simulation procedure].

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:

$$\hat{\theta}_T^{SMM} = \arg\min_{\theta} \Psi_{ST}(\theta) \ , \tag{3.58}$$

where

$$\Psi_{ST}(\theta) = \{\sum_{t=1}^{T}[K(y_t) - \frac{1}{S}\sum_{s=1}^{S} K(y_t^s(\theta))]\}'\hat{\Omega}^{-1}\{\sum_{t=1}^{T}[K(y_t) - \frac{1}{S}\sum_{s=1}^{S} K(y_t^s(\theta))]\} \tag{3.59}$$

and

$$\hat{\Omega} = \hat{\Gamma}_0 + \sum_{k=1}^{K}\left(1 - \frac{k}{K+1}\right)(\hat{\Gamma}_k + \hat{\Gamma}_k') \ , \tag{3.60}$$

where

$$\hat{\Gamma}_k = \frac{1}{T}\sum_{t=k+1}^{T}[K(y_{t-k}) - \frac{1}{S}\sum_{s=1}^{S} K(y_t^s(\tilde{\theta}_T))][K(y_t) - \frac{1}{S}\sum_{s=1}^{S} K(y_t^s(\tilde{\theta}_T))]' \tag{3.61}$$

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. Of course in our framework where the analytical 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.

3.3.2. The indirect inference method

The indirect inference method (hereafter II) 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 (1.1). The instrumental model corresponds to the autoregression with conditional heteroscedasticity:

$$y_{t+1} - y_t = -b^*(y_t - \beta^*) + \sqrt{c^* y_t(1 - y_t)}z_{t+1} \tag{3.62}$$

where the errors z_{t+1} are independent with standard normal distribution. After a change of parameters, the instrumental model can be rewritten as an autoregressive model linear in the new parameters α, γ.

$$\frac{y_{t+1}}{\sqrt{y_t(1-y_t)}} = \alpha \frac{y_t}{\sqrt{y_t(1-y_t)}} + \frac{\gamma}{\sqrt{y_t(1-y_t)}} + \delta z_{t+1} \ . \tag{3.63}$$

The parameter $g = (\alpha, \gamma, \delta)$ can be estimated by OLS applied to the set of available data. $\hat{g}_T$ denotes the associated estimator. We can also introduce $\hat{g}_T^s(\theta)$ 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:

$$\hat{\theta}_T^{II}(\Omega) = \arg\min_\theta \left[\hat{g}_T - \frac{1}{S}\sum_{s=1}^{S} \hat{g}_T^s(\theta)\right]' \Omega \left[\hat{g}_T - \frac{1}{S}\sum_{s=1}^{S} \hat{g}_T^s(\theta)\right] \tag{3.64}$$

where Ω is a symetric nonnegative matrix defining the metric. Since the model is just identified the estimation does not depend on the choice of Ω and we select $\Omega = I_3$ in the application.

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. Then we check the accuracy of the simulated path for different sets of parameter value.

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 a meaning to 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:

$$y^s_{(k+1)\delta} = \begin{cases} 0.99 \ , & if \qquad y^{*s}_{(k+1)\delta} \geq 1 \ , \\ y^{*s}_{(k+1)\delta} \ , & if \quad 0 \leq y^{*s}_{(k+1)\delta} \leq 1 \ , \\ 0.01 \ , & if \qquad y^{*s}_{(k+1)\delta} < 0 \ , \end{cases} \tag{4.65}$$

where:

$$y^{*s}_{(k+1)\delta} = y^s_{k\delta} - b(y^s_{k\delta} - \beta)\delta + \sqrt{c y^s_{k\delta}(1 - y^s_{k\delta})}\sqrt{\delta}\epsilon^s_k \ , \tag{4.66}$$

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 (4.65-4.66) 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.

4.2. Simulated series

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

- set I : $(0.43, 0.5, 0.8)$
- set II : $(0.5, 0.5, 0.25)$
- set III : $(2, 0.8, 1)$.

The different sets have been selected to reproduce the three 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 marginal distribution corresponding to set III is a right-skewed beta distribution. The dynamics also differ. The processes associated with sets I and II admit rather high first order correlation, larger than 0.6 while set III exhibits a low first order correlation ($\simeq 0.13$). 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 1: Simulated paths, set I]
[Insert Figure 2: Simulated paths, set II]

Table 1: Summary statistics for y_t and beta distr.,set I

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

Due to the choice of parameter values, the process distribution can give more or less weight in a neighbourhood 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 (4.66) 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 3 and 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 3: Empirical marginal distributions, set I]
[Insert Figure 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 3 and 4) gives some information on the accuracy of the simulations as well as tables 1 and 2 reporting summary statistics .

The skewness [resp. fat tail] effects are also clearly seen on the sample distribution of $P_2(y_t)$ [resp. $P_3(y_t)$]. The sample means are close to zero and 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 and 6 provide the joint autocorrelogramms of the three series for the first 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

Table 2: Summary statistics for y_t and beta distr., set II

set II	sample moments of y_t	theoretical beta
mean	0.4960	0.5
variance	0.0493	0.05
skewness	0.0337	0
kurtosis	2.2051	2.1428

Table 3: Summary statistics for P_1,P_2,P_3, set I

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

Table 4: Summary statistics for P_1,P_2,P_3, set II

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: Sample and theoretical correlations for P_1,P_2,P_3, set I

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

Table 6: Sample and theoretical correlations for P_1,P_2,P_3, set II

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

clearly seen on the correlogramms, but the autoregressive dynamics is more difficult to detect on the autocorrelogramms shown in Figures 5 and 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 tables 5 and 6 where are reported the sample and theoretical first and second order correlations.

[Insert Figure 5: Empirical correlations, set I]
[Insert Figure 6: Empirical correlations , set II]
[Insert Figure 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 both for marginal and dynamic features.

5. Comparison of the estimators

5.1. The estimation methods

The aim of this section is to compare by Monte Carlo the finite sample properties of the estimators introduced in section 3, that are the approximated maximum likelihood (AML), the exact indirect estimator (EI), a GMM estimator, a SMM estimator and an estimator derived by indirect inference (II). Different sample sizes are considered

$T = 500, 1000, 1500, 2000, 5000$ and the number of replications in the Monte Carlo study is $M = 1000$.

The AML approach is applied with a number $N = 4$ of 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 $N = 2$ eigenfunctions. 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 eight moments described in section 3.2.4. These moments include those used in the EI approach together with higher moments associated with skewness and 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 $K(y_t^s(\theta))$ as explicited in section 3.3.1. The SMM approach is meaningful to study the behaviour 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 give 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 .

5.2. 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 experiment, we consider the bias standardized by $\sqrt{T}$ and divided by the true value of the parameter. 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.

Let us consider the sign of the bias. The bias for the volatility coefficient c is always positive across experiments for the exact methods (EI, AML, GMM) in contrast to the simulated methods (SMM, II) for which it is negative. A positive bias for c leads to overestimate the volatility and therefore to take less risk. In this sense, the negative bias is not a suitable property of the simulated methods. The mean reverting parameter b is always biased upwards for most of the methods and across experiments. However the sign of the bias for the mean parameter β is not constant and varies across experiments with a strong negative bias in set III with the asymmetric beta distribution. Note that the sign of the bias for β is the same for all methods.

Concerning the importance of the bias, we clearly see that the simulated methods (SMM, II) drastically reduce the finite sample bias compared with GMM in particular, for all parameters and across all experiments. The simulated moments seem to perform better than the indirect procedure in reducing the bias. Moreover, we can see that GMM exhibits more bias than EI, which suggests that including more moments in the estimation increases the magnitude of the bias. The parameter with the strongest bias is the mean reverting parameter b whereas the mean parameter β is much less biased.

The standardized bias seem to stabilize towards a limit when the sample size increases which reinforces the idea that the bias converges at a speed of $\sqrt{T}$. The bias is the strongest in set III with the asymmetric beta distribution for all parameters even for the mean parameter β. 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.

[Insert Tables 7-15:$\sqrt{T}\times$ bias of b,β, c, set I,II,III .]

ii) Analysis of the variance.

We consider the variance statistics standardized by the sample size and divided by the true value of the parameter to the square in order to facilate the comparison across sample size and experiment. The importance of the variance is globally the same for AML, EI and SMM. The GMM and II estimators have a little more variance. The variance of the estimators diminishes when the sample size increases or the standardized variance tends to a limit. The variances are getting very poor in set III with the asymmetric beta distribution which puts more weight on $y = 1$. When y take the value 1 very often, we do not have much information and we are close to the unidentifiability case.

We expect that GMM is better than EI at least asymptotically since GMM takes into account more moment conditions in the estimation procedure. This is not clear in set I and II but clearly rises in set III where the GMM estimator performs much better than EI. Indeed, by including moments of the form Ey_t^3, $Ey_t y_{t-1}^2$ in the GMM estimation we may capture dynamic features such as skewness, leverage effect compared to the EI estimator. Moreover we may expect that GMM and SMM are equivalent at least asymptotically but this not clear here. Besides, we observe that the AML estimator is globally better than GMM. Indeed, the estimation of the weighting matrix Ω poorly affects the estimation results. Moreover using a Bartlett estimator leads to a truncation bias in the estimation of Ω. Moreover, the more moments we include in the estimation procedure, the more risks of colinearity we may get and therefore the more trouble we have to invert the weighting matrix Ω.

[Insert Tables 16-24 :$T\times$ Variance of b,β, c, set I,II,III .]

iii) Analysis of marginal distributions.

The empirical marginal distributions of the parameters have not been standardized but are implicitely standardized by the software. The distributions are getting closer to the gaussian distribution when the sample size increases. We observe the presence of multimodes in the empirical marginal distributions in particular for AML and GMM. Two reasons can be evocated either there are multimodes in the underlying distribution of the estimator or this is due to numerical optimization issues. There may exist multiple local minima but we do not know which one to choose and how to select the same one every time. Since the EI estimators do not have multimodes, using the EI estimates as starting values could reduce the occurence of multimodes for AML and GMM. Note that the number of modes depend on the values of the parameters. More specifically, it is in set I that we observe multimodes whereas set III exhibits severe skewness of the marginal empirical distributions of the mean reverting and volatility parameters (resp. b and c) for all estimators due to the asymmetric beta distribution which puts more weight on $y = 1$. The distribution of the mean parameter β is almost not affected by this skewness effect whereas it is affected by multimodes.

[Insert Figures 8,9,10:Empirical marginal distributions of the EI estimates, set I,II,III] [Insert Figures 11,12,13:Empirical marginal distributions of the AML estimates, set I,II,III] [Insert Figures 16,17,18:Empirical marginal distributions of the GMM estimates, set I,II,III] [Insert Figures 19,20,21:Empirical marginal distributions of the SMM estimates, set I,II,III] [Insert Figure **??**:Empirical marginal distributions of the II estimates, set II]

5.3. Joint distributional properties of the estimators of b and c

The joint distributions of b and c of the EI estimators have the typical ellipsoidal shape characterizing the bivariate gaussian distribution. The GMM estimators are quite close to this typical shape. The joint distributions are extremely stretched in set III for all estimators and do not look like the typical ellipsis of the gaussian bivariate distribution. We observe again the problem of multiple solutions on the joint distributions for AML. Either we select one minimum rather than the other one, we may make a more or less serious mistake. Note that deviations from normality increases when the sample size decreases.

6. Concluding remarks

To summarize, the approximated maximum likelihood estimator (AML) exhibits the best behavior with respect to the bias and the variance. Indeed, the necessity to estimate a weighting matrix in the GMM procedure creates some numerical instability due to some difficulties to invert the weighting matrix. These numerical difficulties can be avoided by resorting to the approximated maximum likelihood. The empirical joint distribution of the AML estimators of the volatility coefficient and of the mean-reverting parameter looks similar to the bivariate gaussian distribution for set II but unfortunately exhibits multimodes for set I. One way of checking if this is due to numerical optimization issue would be to use the exact indirect estimates as starting values in the AML procedure since the marginal and joint empirical distributions of the EI estimator are free of multimodes.

Table 7: $\frac{\sqrt{T}}{b_0}\times$ Bias of b: set I

Sample size	AML	EIG	EI	GMM	SMM	II
500	8.02	4.56	5.49	5.64	6.25	-0.46
1000	11.20	6.36	7.43	7.51	4.18	-0.61
1500			8.98	9.01	3.41	
2000	15.59	8.72	10.31	10.39	2.80	-1.31
5000	24.41	13.48	16.02	16.09	-0.65	-3.83

Table 8: $\frac{\sqrt{T}}{b_0}\times$ Bias of b: set II

Sample size	AML	EIG	EI	GMM	SMM	II
500	1.55	-0.26	1.17	1.17	-0.10	
1000	1.98	-0.35	1.34	1.40	0.70	
1500			1.51	1.50	1.83	
2000	2.32	0.25	1.64	1.63	0.93	
5000	2.98	2.64	2.30	2.33	0.15	

Table 9: $\frac{\sqrt{T}}{b_0}\times$ Bias of b: set III

Sample size	AML	EIG	EI	GMM	SMM	II
500	7.25		7.48	5.17	0.86	
1000	7.01		6.94	6.68	0.80	
1500			7.69	7.55	0.71	
2000	8.56		8.66	8.40	0.26	
5000	13.03		12.87	12.72	0.07	

Table 10: $\frac{\sqrt{T}}{beta_0}\times$ Bias of β: set I

Sample size	AML	EIG	EI	GMM	SMM	II
500	-0.0025	0.17	-0.00	0.00	1.13	-0.14
1000	-0.0037	0.075	-0.03	-0.029	0.22	0.50
1500			-0.05	-0.049	0.91	
2000	-0.016	0.037	-0.05	-0.061	0.80	0.99
5000	-0.028	0.051	-0.07	-0.055	-0.25	-1.13

Table 11: $\frac{\sqrt{T}}{beta_0}\times$ Bias of β: set II

Sample size	AML	EIG	EI	GMM	SMM	II
500	-0.033	-0.19	-0.01	-0.009	1.32	
1000	-0.036	-0.28	-0.03	-0.019	0.96	
1500			-0.04	-0.033	1.21	
2000	-0.0019	-0.43	-0.04	-0.043	0.84	
5000	-0.038	-0.093	-0.04	-0.038	0.18	

Table 12: $\frac{\sqrt{T}}{beta_0}\times$ Bias of β: set III

Sample size	AML	EIG	EI	GMM	SMM	II
500	-0.43		-0.44	-0.43	0.079	
1000	-0.61		-0.61	-0.61	-0.005	
1500	-0.78		-0.76	-0.75	0.024	
2000	-0.88		-0.87	-0.87	0.099	
5000	-1.38		-1.38	-1.37	0.066	

Table 13: $\frac{\sqrt{T}}{c_0}\times$ Bias of c: set I

Sample size	AML	EIG	EI	GMM	SMM	II
500	-0.21	0.24	0.12	0.15	-1.45	1.85
1000	-0.3	0.37	0.08	0.10	-2.49	2.75
1500			0.10	0.07	-0.43	
2000	-0.49	0.52	0.08	0.08	-0.08	4.36
5000	-0.84	0.81	0.05	0.009	-0.68	7.17

Table 14: $\frac{\sqrt{T}}{c_0}\times$ Bias of c: set II

Sample size	AML	EIG	EI	GMM	SMM	II
500	0.28	0.75	1.28	1.29	-1.58	
1000	0.67	1.07	1.71	1.77	-1.62	
1500			2.10	2.15	-1.50	
2000	0.86	1.88	2.43	2.43	-0.54	
5000	1.28	3.69	3.69	3.77	-0.34	

Table 15: $\frac{\sqrt{T}}{c_0}\times$ Bias of c: set III

Sample size	AML	EIG	EI	GMM	SMM	II
500	6.95		5.33	4.43	-3.72	
1000	6.68		6.16	5.60	-3.46	
1500			6.94	6.33	-2.81	
2000	8.28		7.87	7.04	-0.98	
5000	12.74		11.72	10.73	-0.30	

Table 16: $\frac{T}{b_0^2} \times$ Variance of b: set I

Sample size	AML	EIG	EI	GMM	SMM	II
500	6.82	9.48	8.86	9.63	7.84	6.40
1000	7.01	7.25	8.85	9.23	7.25	7.97
1500			7.84	8.60	6.79	
2000	6.55	6.51	7.92	8.28	6.15	22.07
5000	6.64	6.55	7.80	8.24	5.98	37.93

Table 17: $\frac{T}{b_0^2} \times$ Variance of b: set II

Sample size	AML	EIG	EI	GMM	SMM	II
500	5.67	89.75	6.63	7.13	5.39	
1000	5.28	45.44	6.73	7.16	5.12	
1500			6.02	6.55	4.47	
2000	5.42	55.76	6.13	6.46	3.49	
5000	6.73	5.83	6.03	6.28	0.72	

Table 18: $\frac{T}{b_0^2} \times$ Variance of b: set III

Sample size	AML	EIG	EI	GMM	SMM	II
500	229.57		225	47.81	0.205	
1000	59.25		57.68	44.21	0.132	
1500			34.97	34.74	0.160	
2000	32.62		34.56	32.40	0.104	
5000	31.58		32.18	31.08	0.054	

Table 19: $\frac{T}{beta_0^2} \times$ Variance of β: set I

Sample size	AML	EIG	EI	GMM	SMM	II
500	0.79	6.87	1.80	1.76	2.29	0.88
1000	0.74	4.87	1.72	1.70	2.56	4.65
1500			1.68	1.70	2.81	
2000	0.76	4.84	1.74	1.78	2.71	9.29
5000	0.66	4.65	1.57	1.60	1.80	21.67

Table 20: $\frac{T}{beta_0^2}\times$ Variance of β: set II

Sample size	AML	EIG	EI	GMM	SMM	II
500	0.49	12.24	0.84	0.85	0.87	
1000	0.30	10.26	0.82	0.80	0.87	
1500			0.80	0.81	0.91	
2000	0.32	13.19	0.84	0.84	0.90	
5000	0.85	2.37	0.75	0.75	0.43	

Table 21: $\frac{T}{beta_0^2}\times$ Variance of β: set III

Sample size	AML	EIG	EI	GMM	SMM	II
500	0.065		0.064	0.062	0.076	
1000	0.064		0.065	0.061	0.083	
1500			0.065	0.065	0.093	
2000	0.072		0.070	0.067	0.079	
5000	0.063		0.061	0.060	0.064	

Table 22: $\frac{T}{c_0^2}\times$ Variance of c: set I

Sample size	AML	EIG	EI	GMM	SMM	II
500	3.62	5.08	4.33	4.83	2.58	30.84
1000	3.66	4.23	4.62	4.87	4.09	132.74
1500			4.20	4.62	3.79	
2000	3.56	4.15	4.25	4.47	3.79	30.45
5000	3.51	4.03	4.20	4.49	2.79	54.71

Table 23: $\frac{T}{c_0^2}\times$ Variance of c: set II

Sample size	AML	EIG	EI	GMM	SMM	II
500	3.14	25.94	3.98	4.53	3.34	
1000	2.27	6.98	4.29	4.60	3.09	
1500			3.91	4.39	2.96	
2000	2.28	6.95	3.97	4.33	2.88	
5000	1.96	3.84	3.97	4.32	2.03	

Table 24: $\frac{T}{c_0^2}\times$ Variance of c: set III

Sample size	AML	EIG	EI	GMM	SMM	II
500	214.87		62.56	46.66	1.73	
1000	58.21		42.70	42.82	1.98	
1500			35.06	34.58	1.97	
2000	33.67		34.86	31.51	1.40	
5000	31.68		31.89	29.40	0.94	

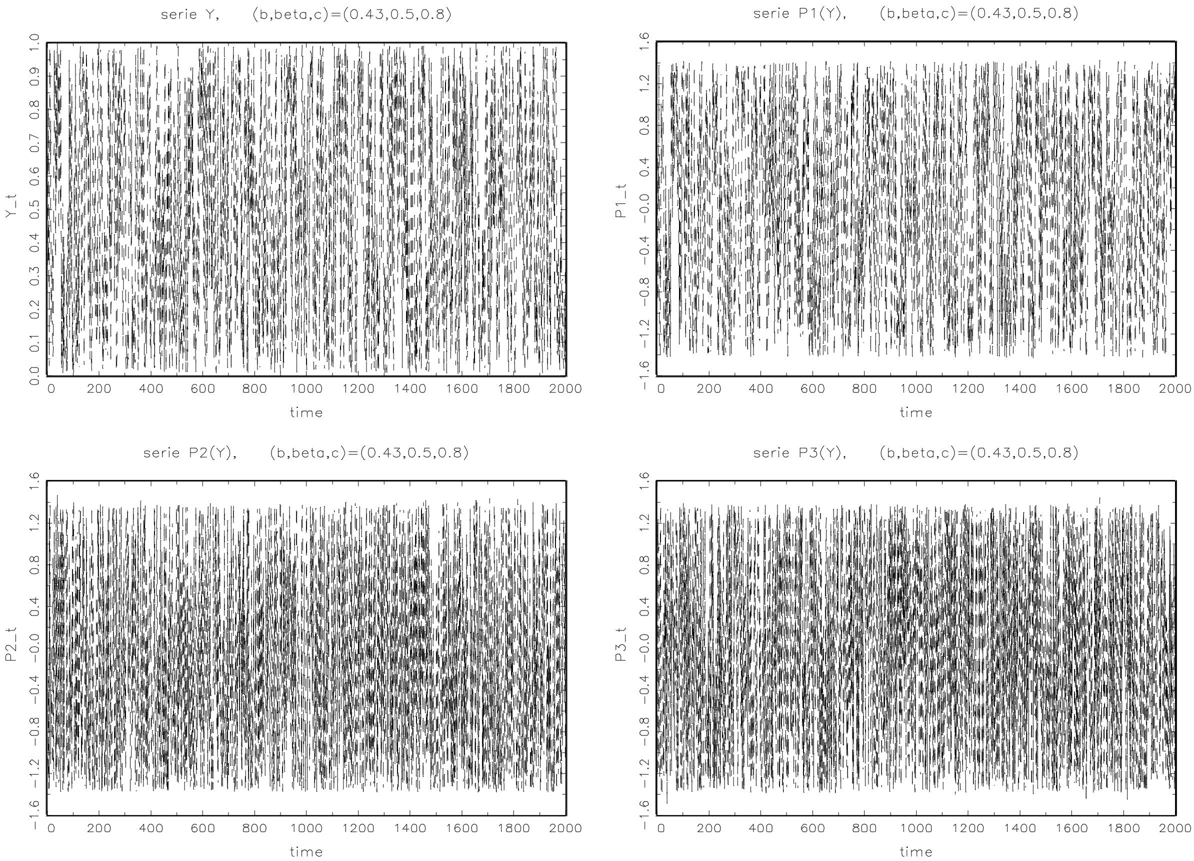


Figure 1: Simulated paths, set I

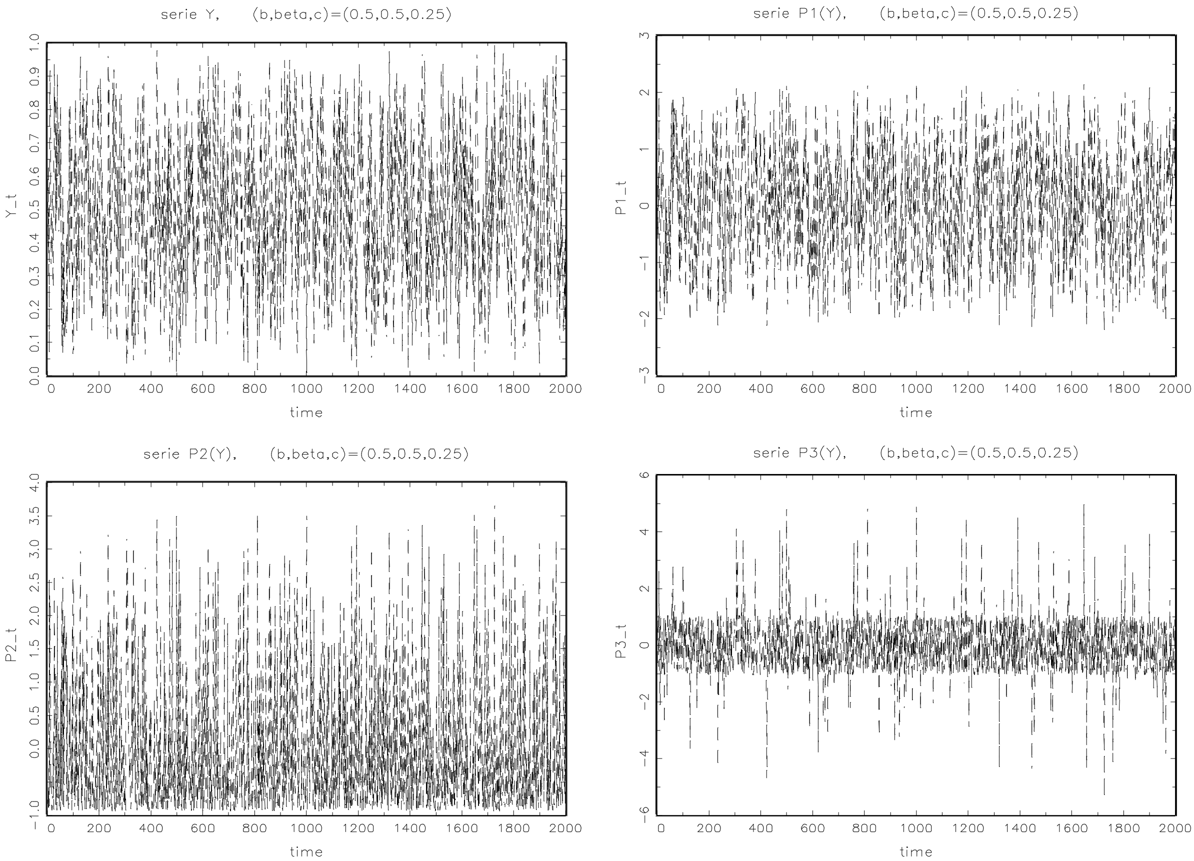


Figure 2: Simulated paths, set II

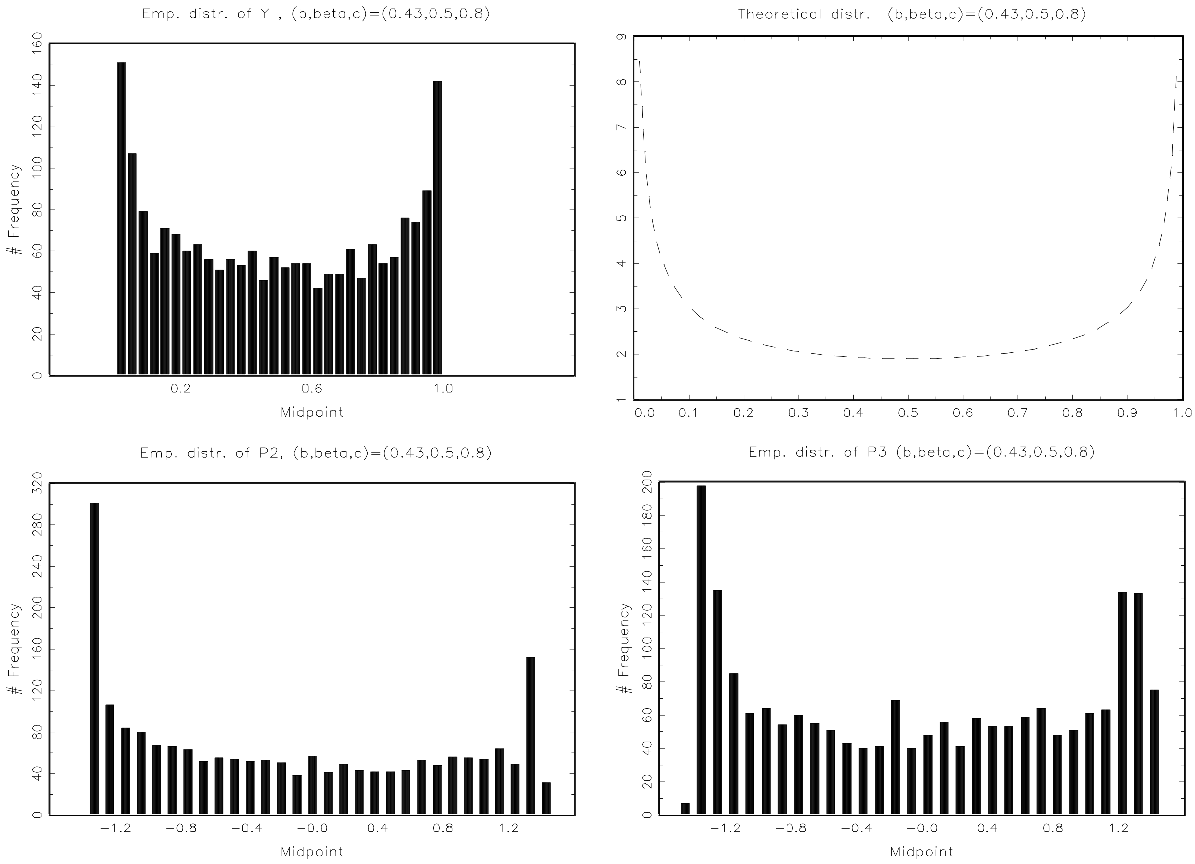


Figure 3: Empirical marginal distributions, set I

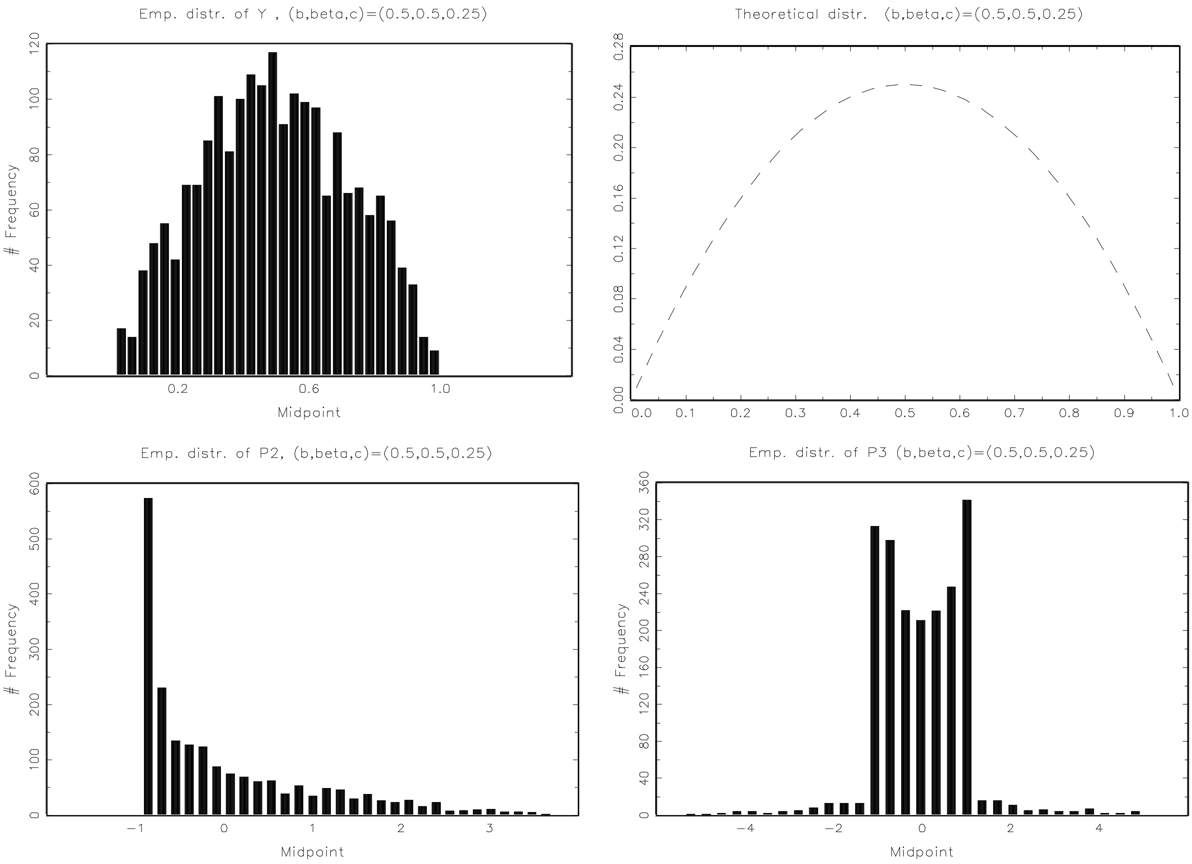


Figure 4: Empirical marginal distributions, set II

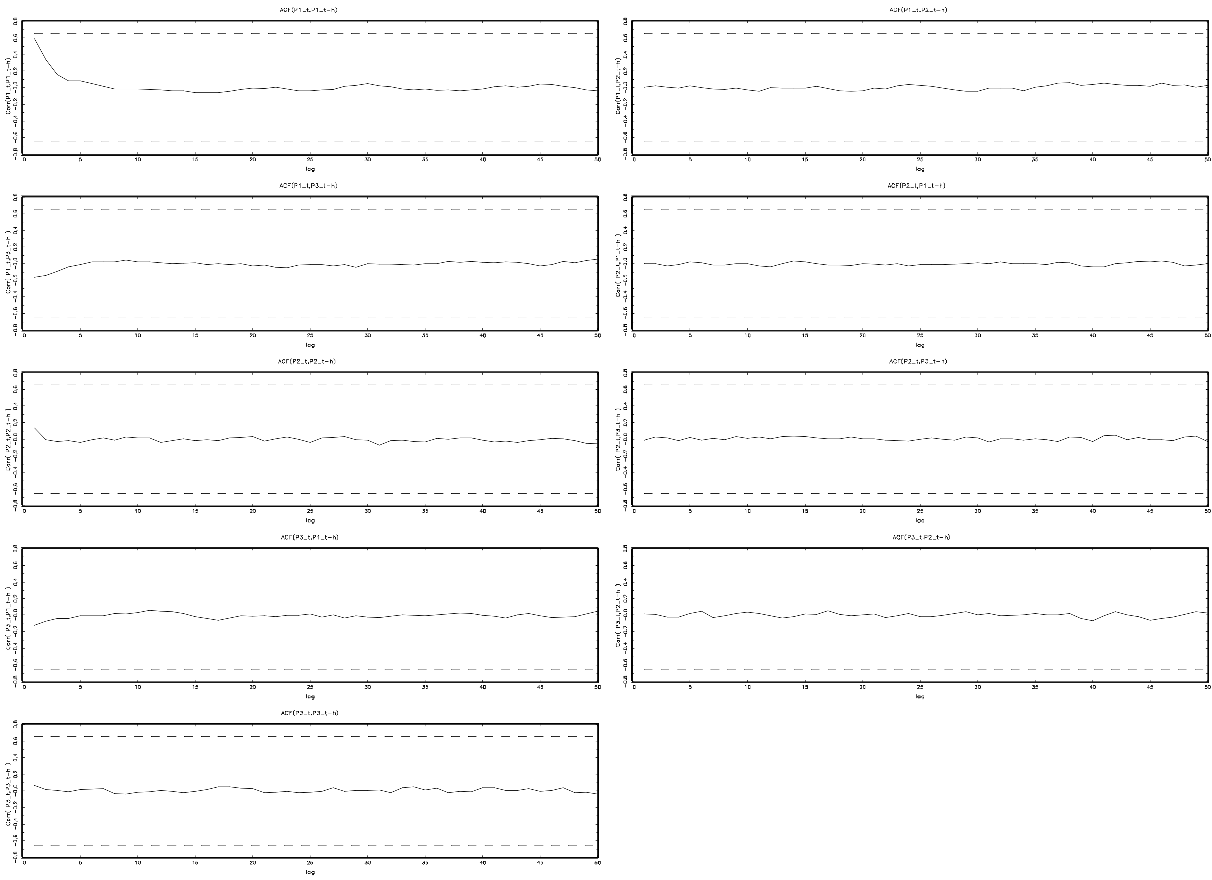


Figure 5: Cross autocorrelograms, set I

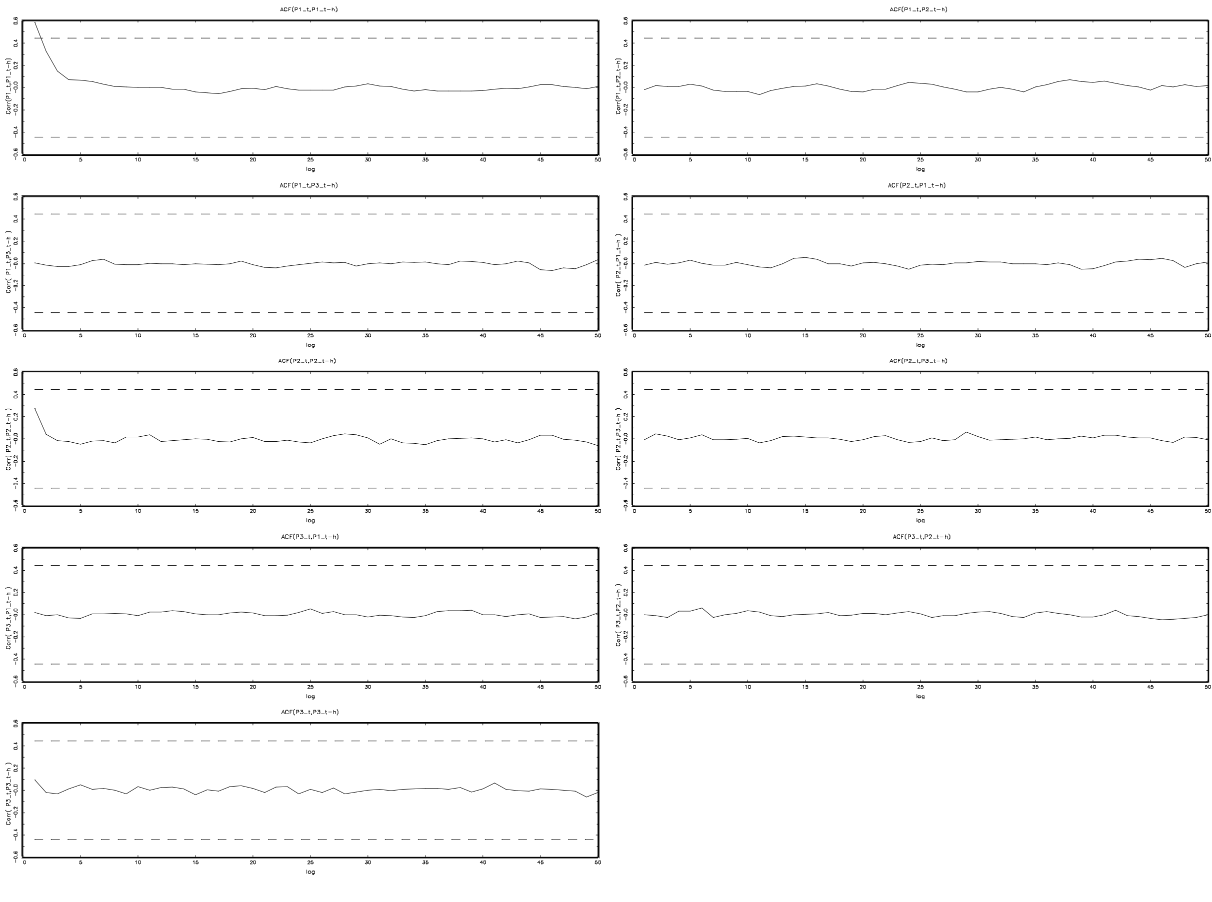


Figure 6: Cross autocorrelograms, set II

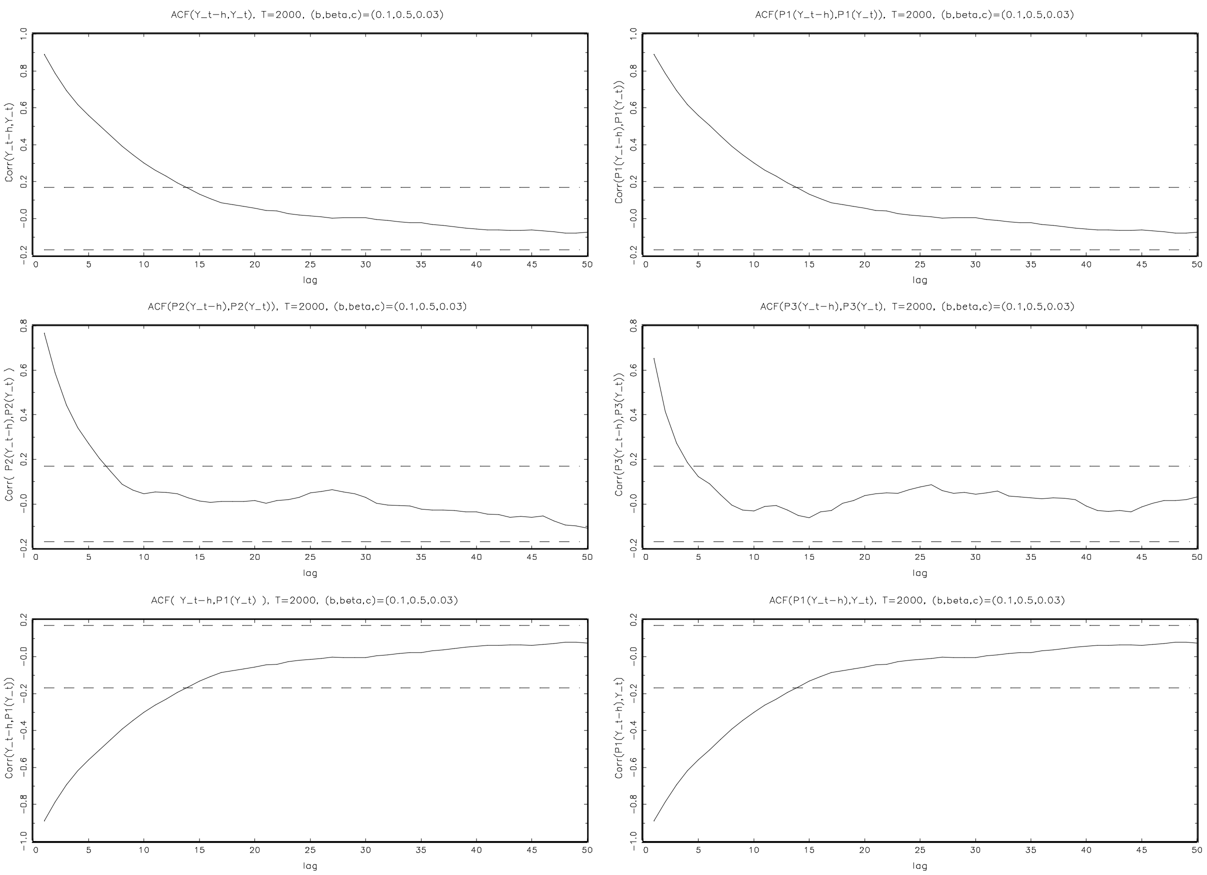


Figure 7: Cross autocorrelograms, set $(0.1, 0.5, 0.03)$

Figure 8: Marginal empirical distribution of EI, set I
(parameter per row, size per column:500,1000,2000,5000).

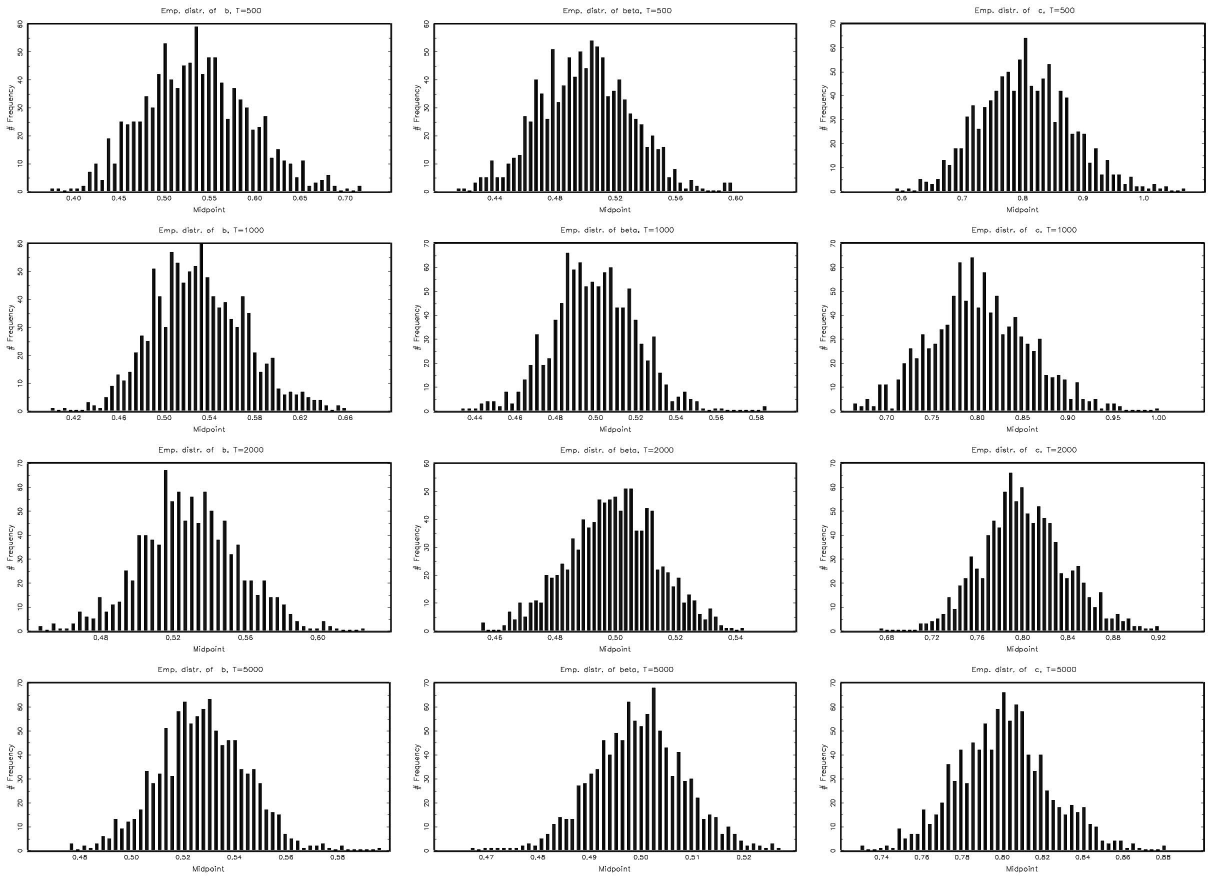


Figure 9: Marginal empirical distribution of EI, set II
(parameter per row, size per column:500,1000,2000,5000).

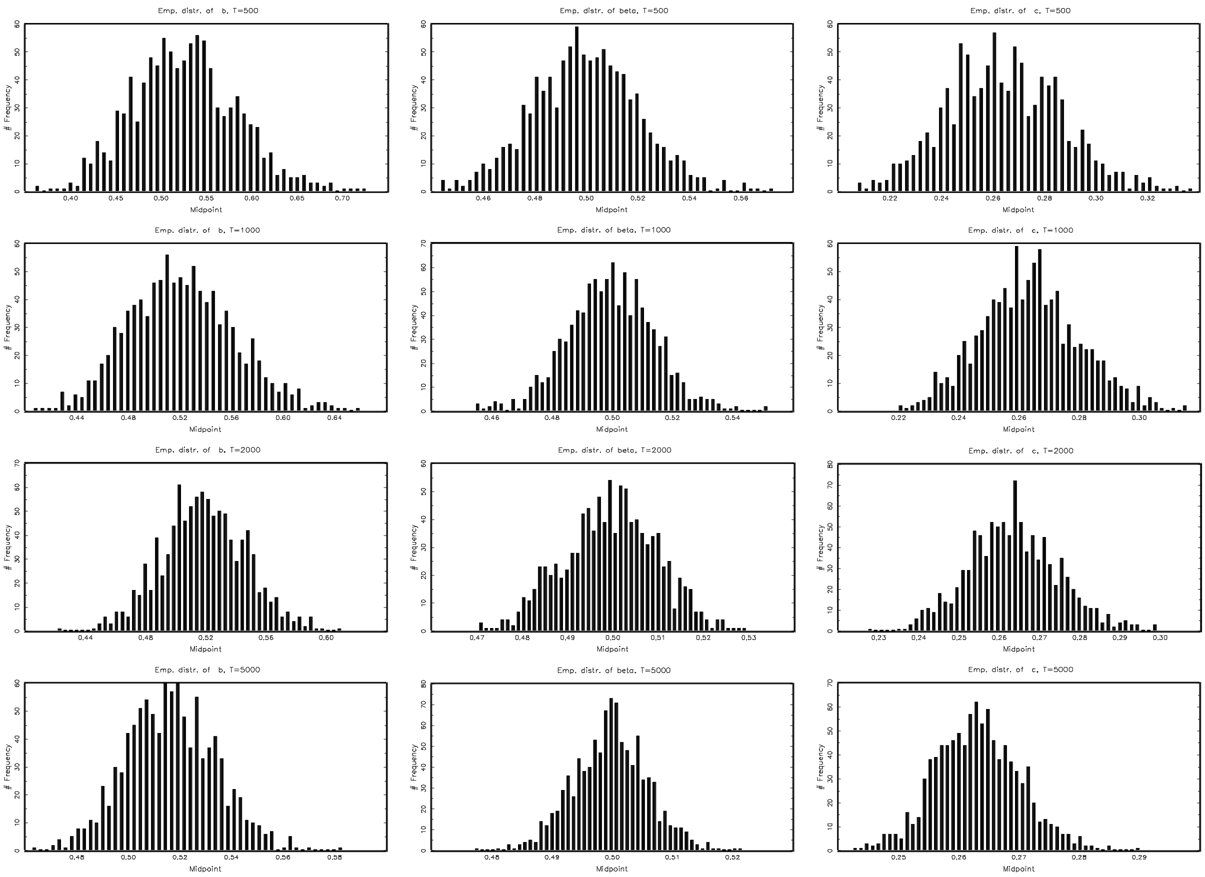


Figure 10: Marginal empirical distribution of EI, set III
(parameter per row, size per column:500,1000,2000,5000).

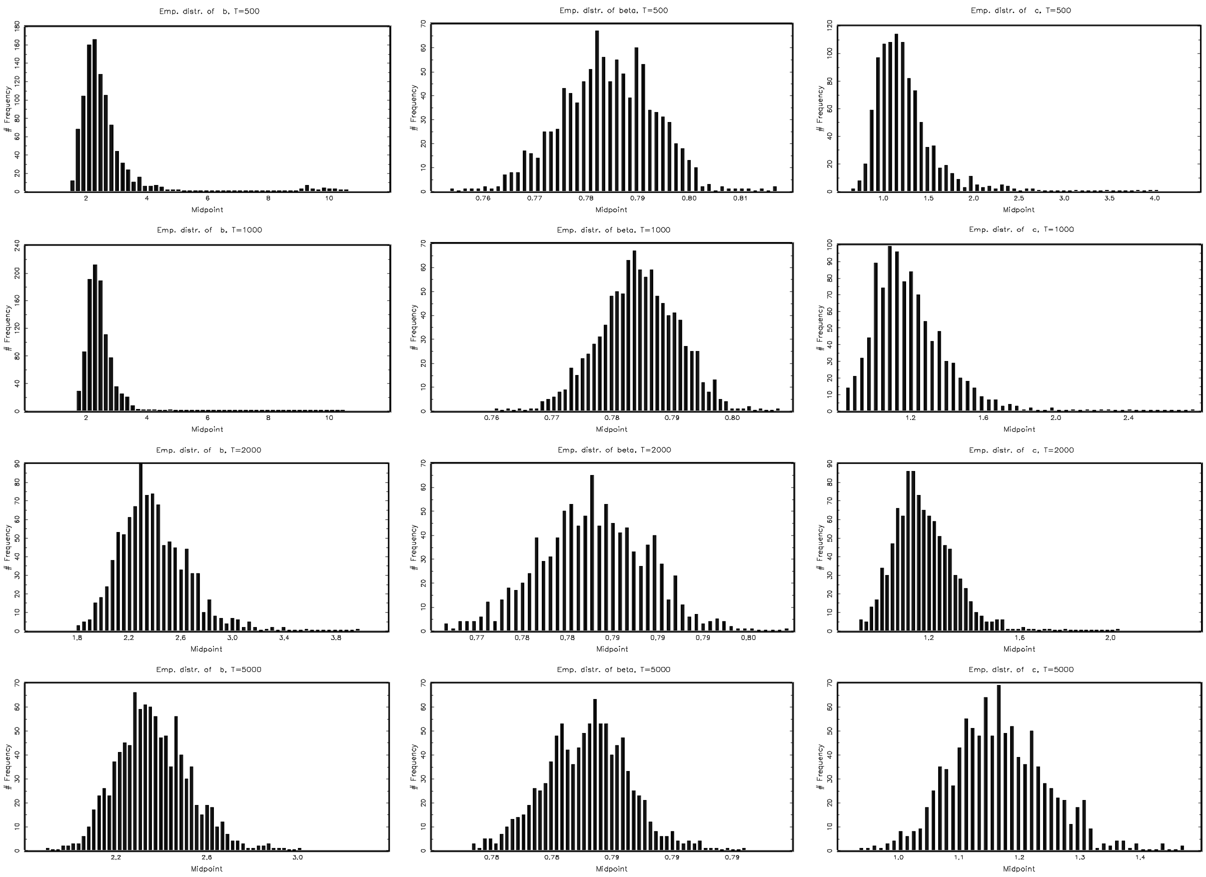


Figure 11: Marginal empirical distribution of AML, set I
(parameter per row, size per column:500,1000,2000,5000).

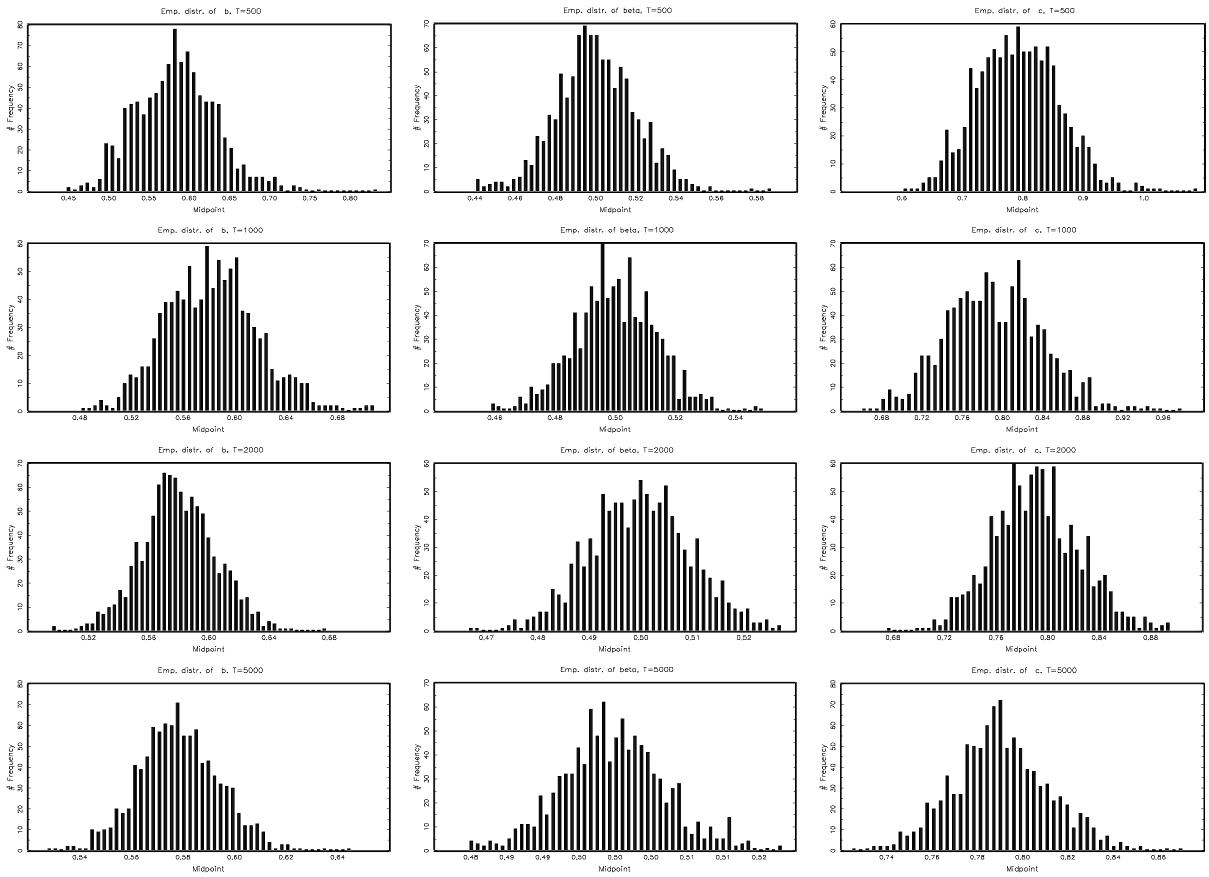


Figure 12: Marginal empirical distribution of AML, set II (parameter per row, size per column:500,1000,2000,5000).

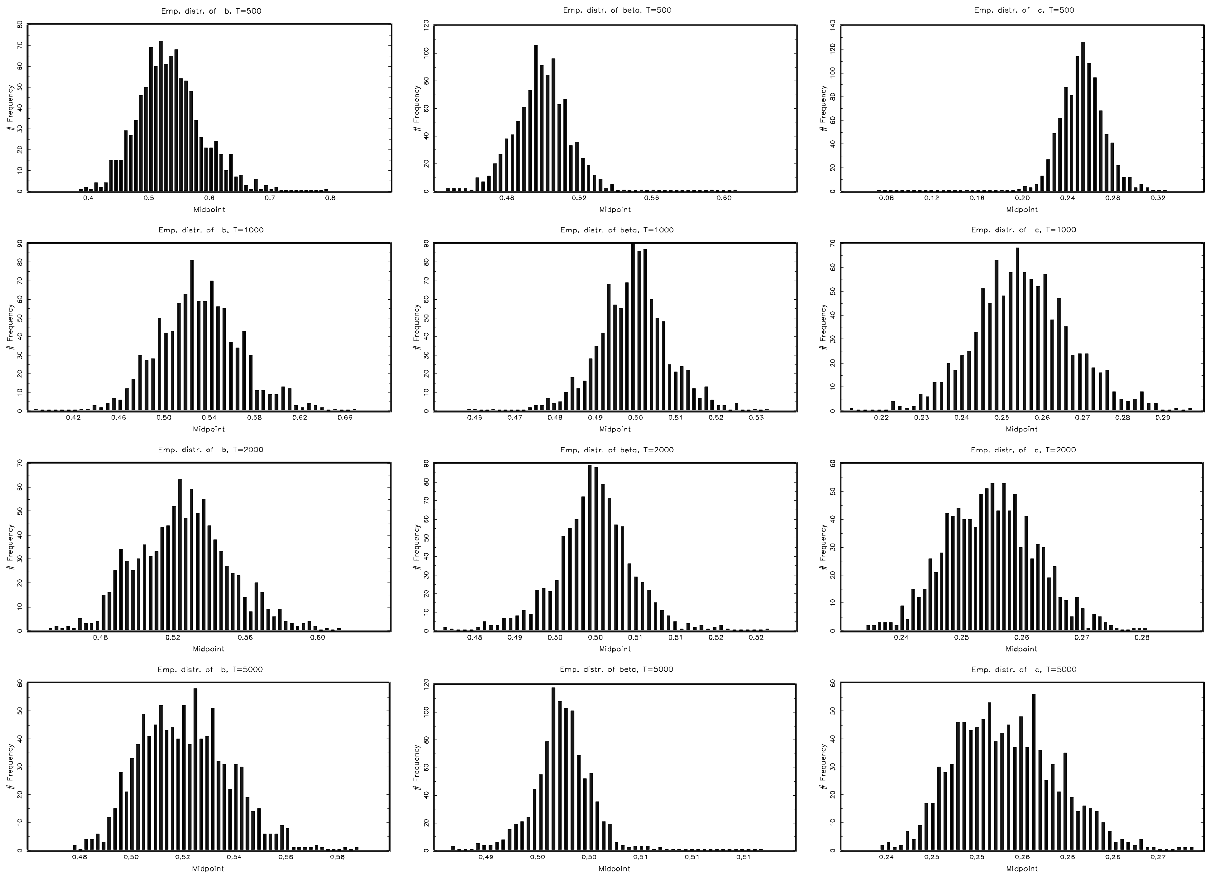


Figure 13: Marginal empirical distribution of AML, set III
(parameter per row, size per column:500,1000,2000,5000).

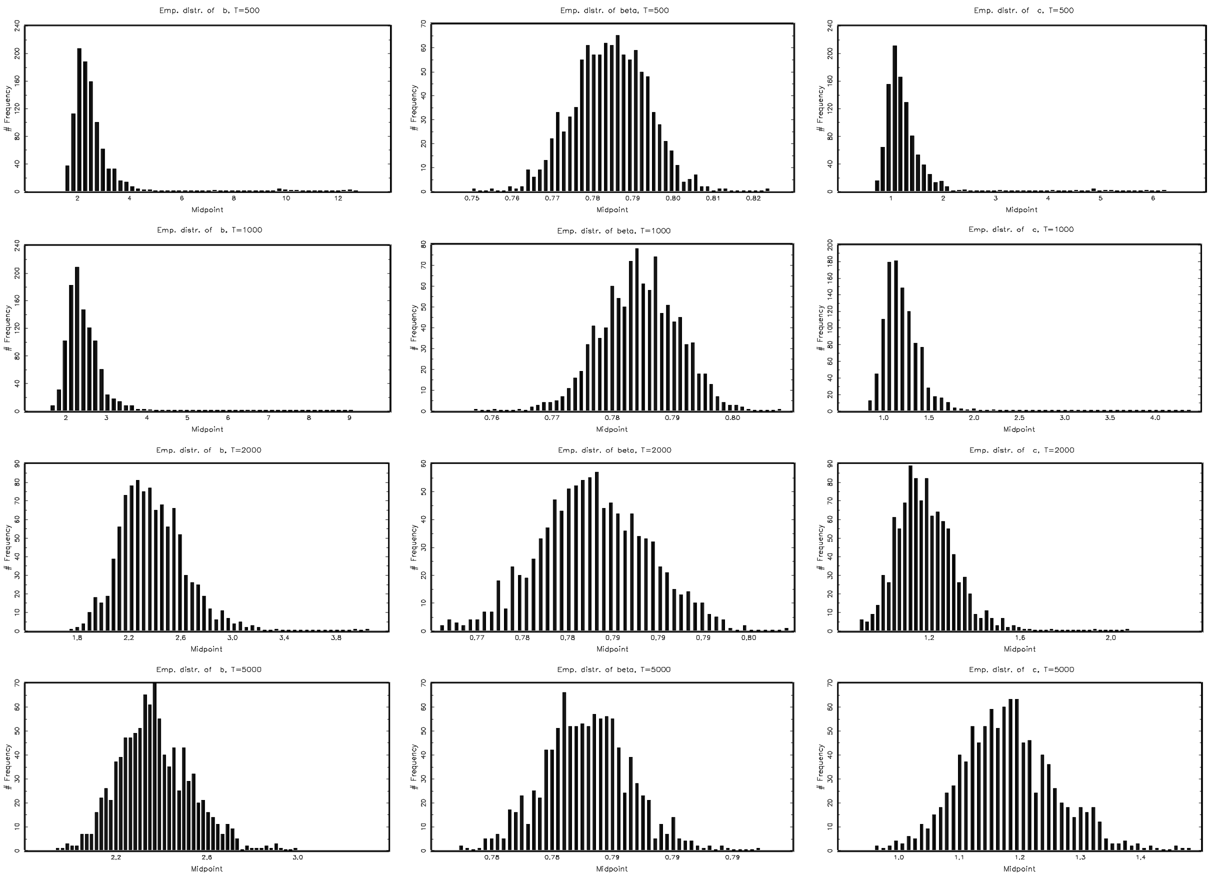


Figure 14: Marginal empirical distribution of EIG, set I
(parameter per row, size per column:500,1000,2000,5000).

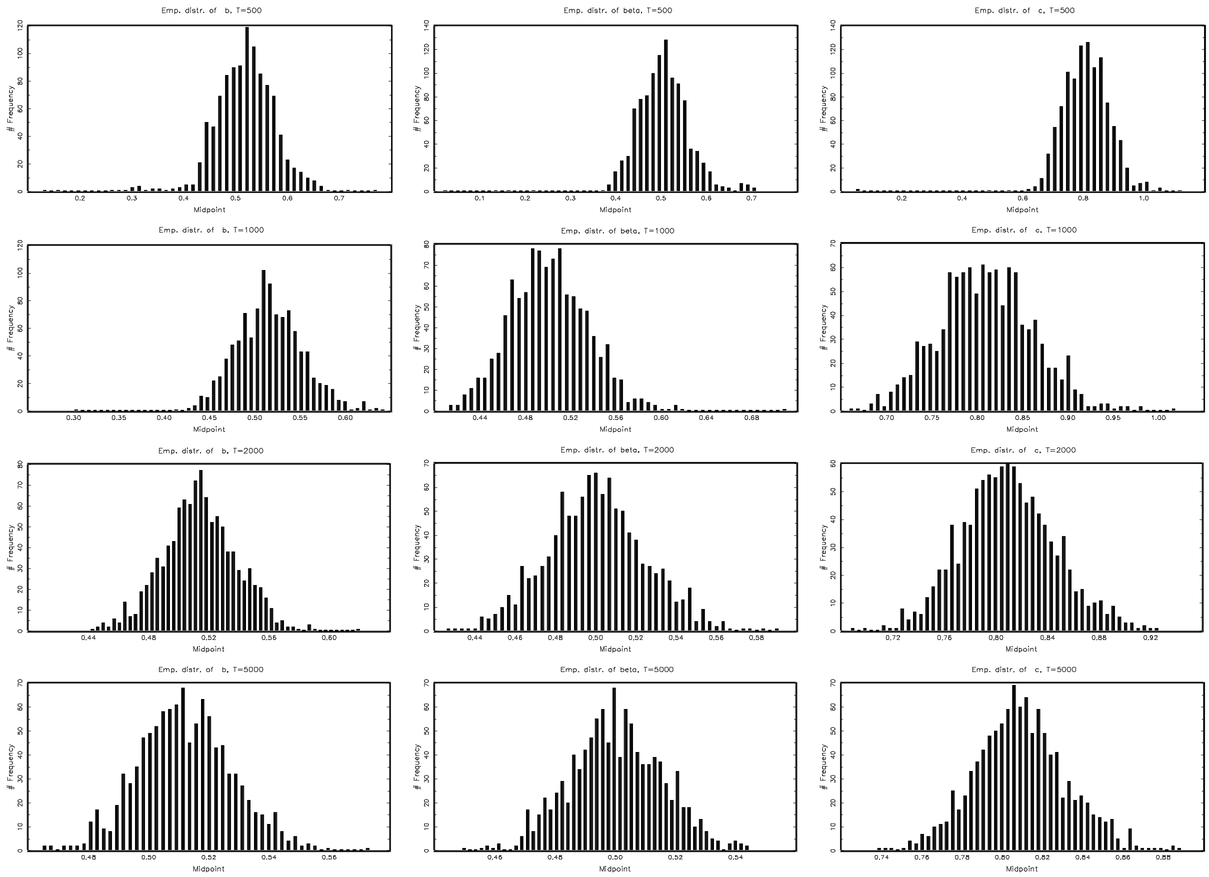


Figure 15: Marginal empirical distribution of EIG, set II (parameter per row, size per column:500,1000,2000,5000).

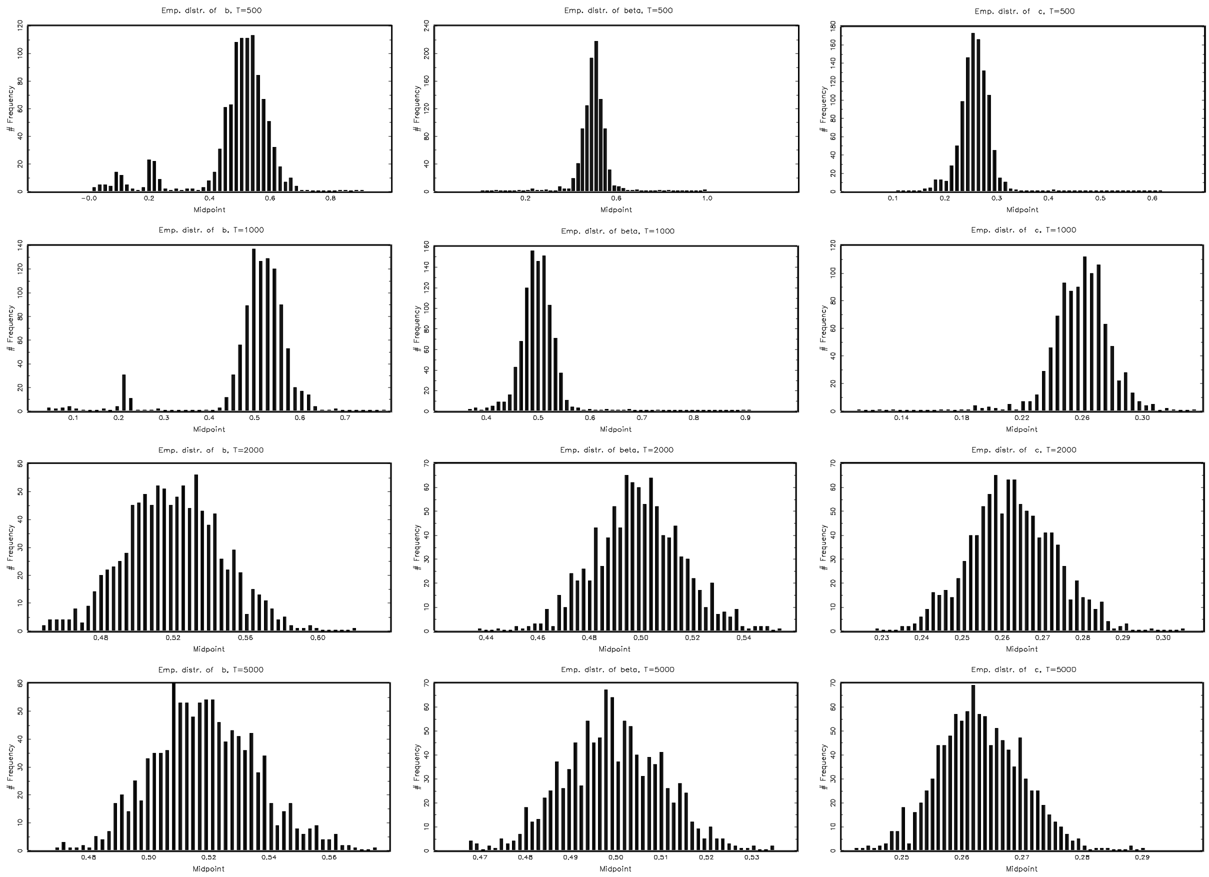


Figure 16: Marginal empirical distribution of GMM, set I
(parameter per row, size per column:500,1000,2000,5000).

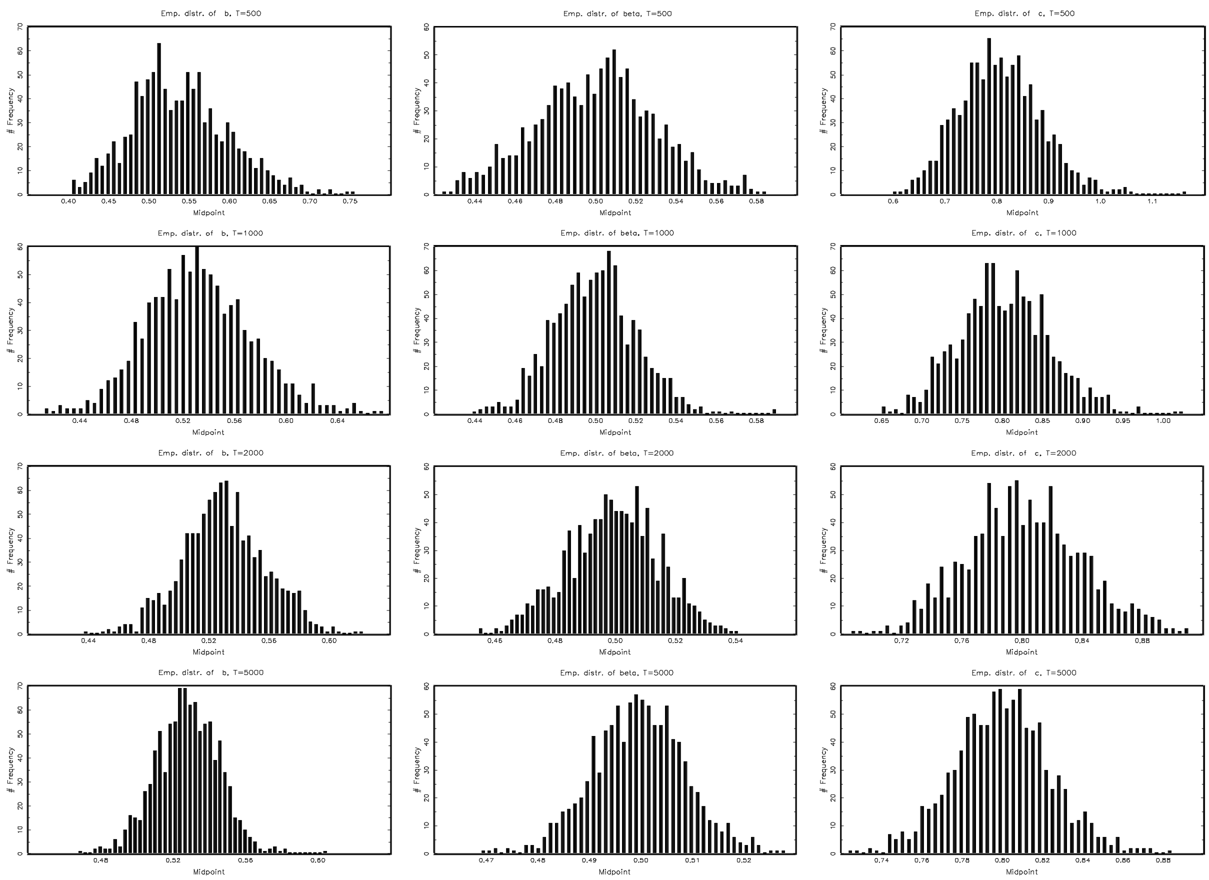


Figure 17: Marginal empirical distribution of GMM, set II
(parameter per row, size per column:500,1000,2000,5000).

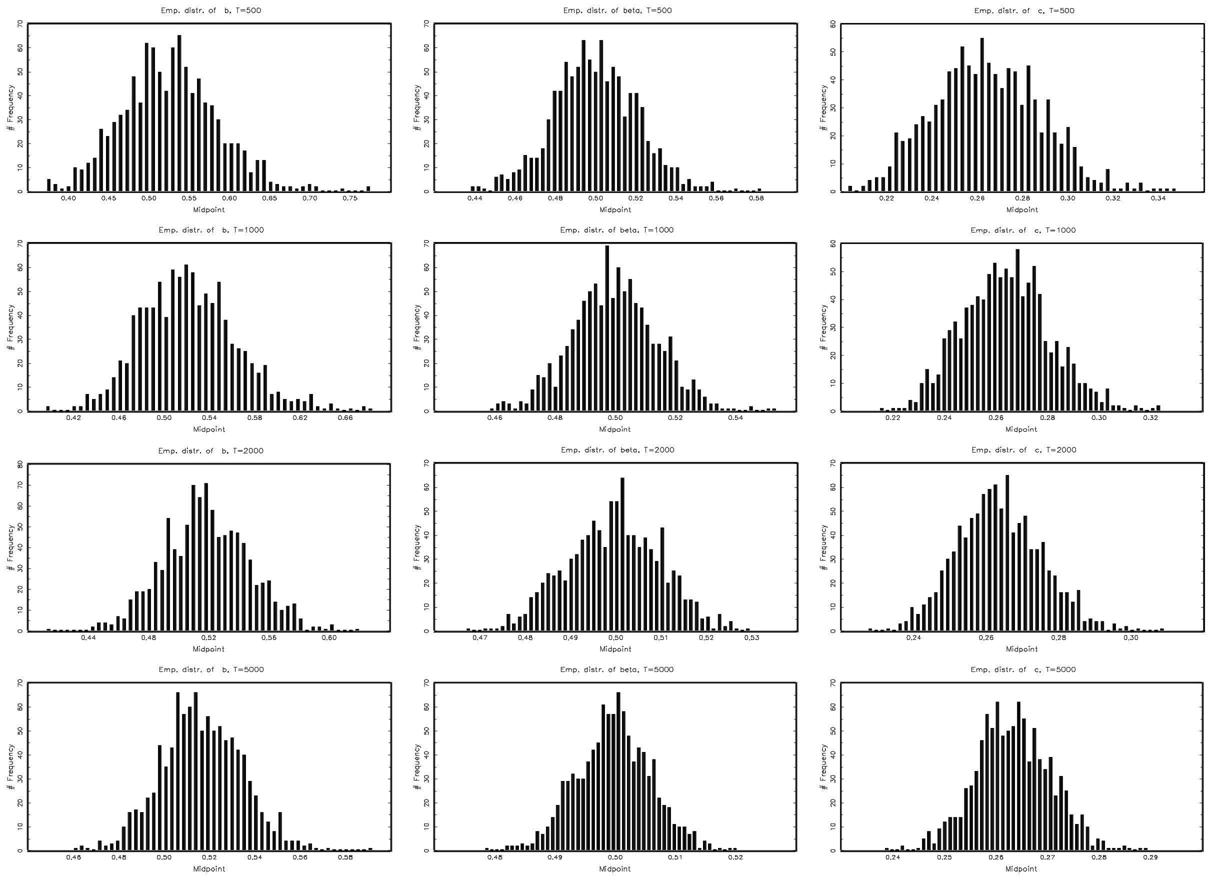


Figure 18: Marginal empirical distribution of GMM, set III (parameter per row, size per column:500,1000,2000,5000).

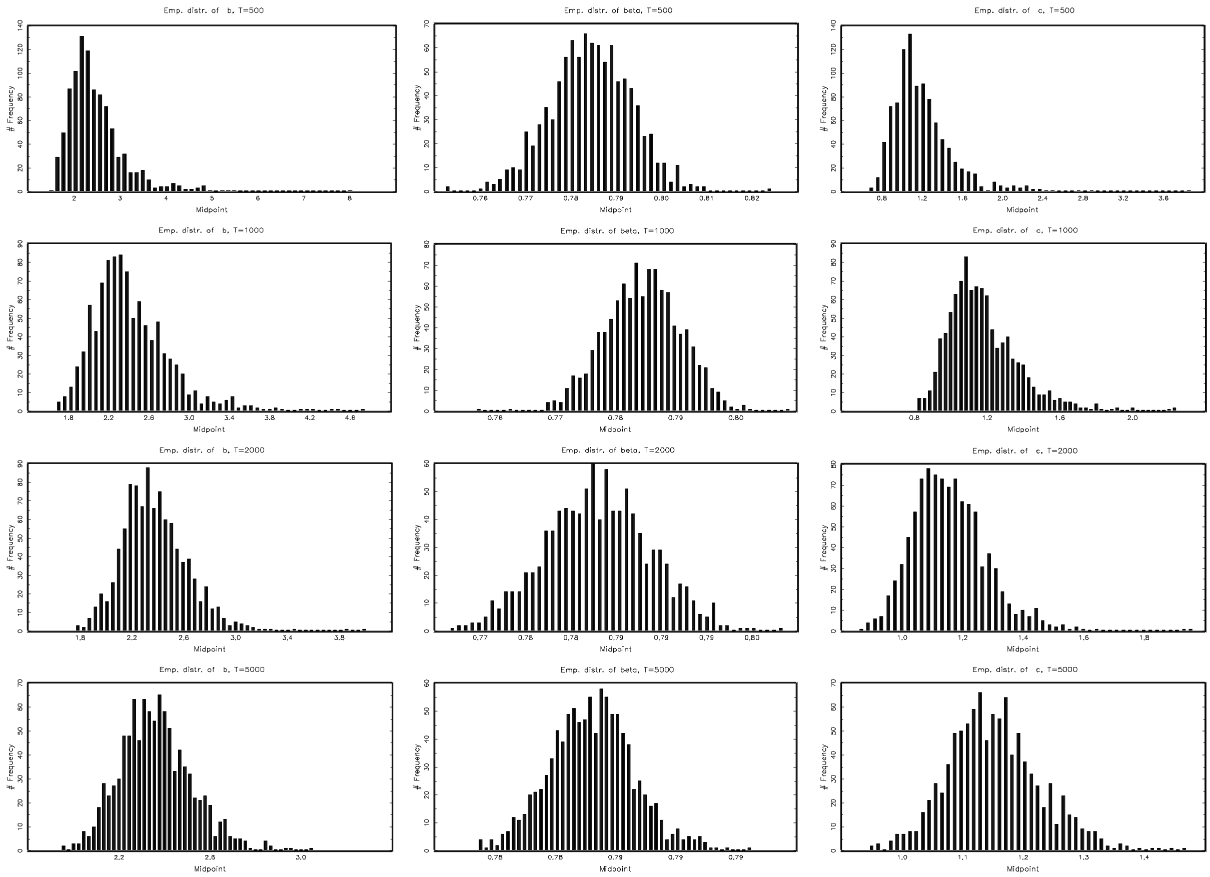


Figure 19: Marginal empirical distribution of SMM, set I
(parameter per row, size per column:500,1000,2000,5000).

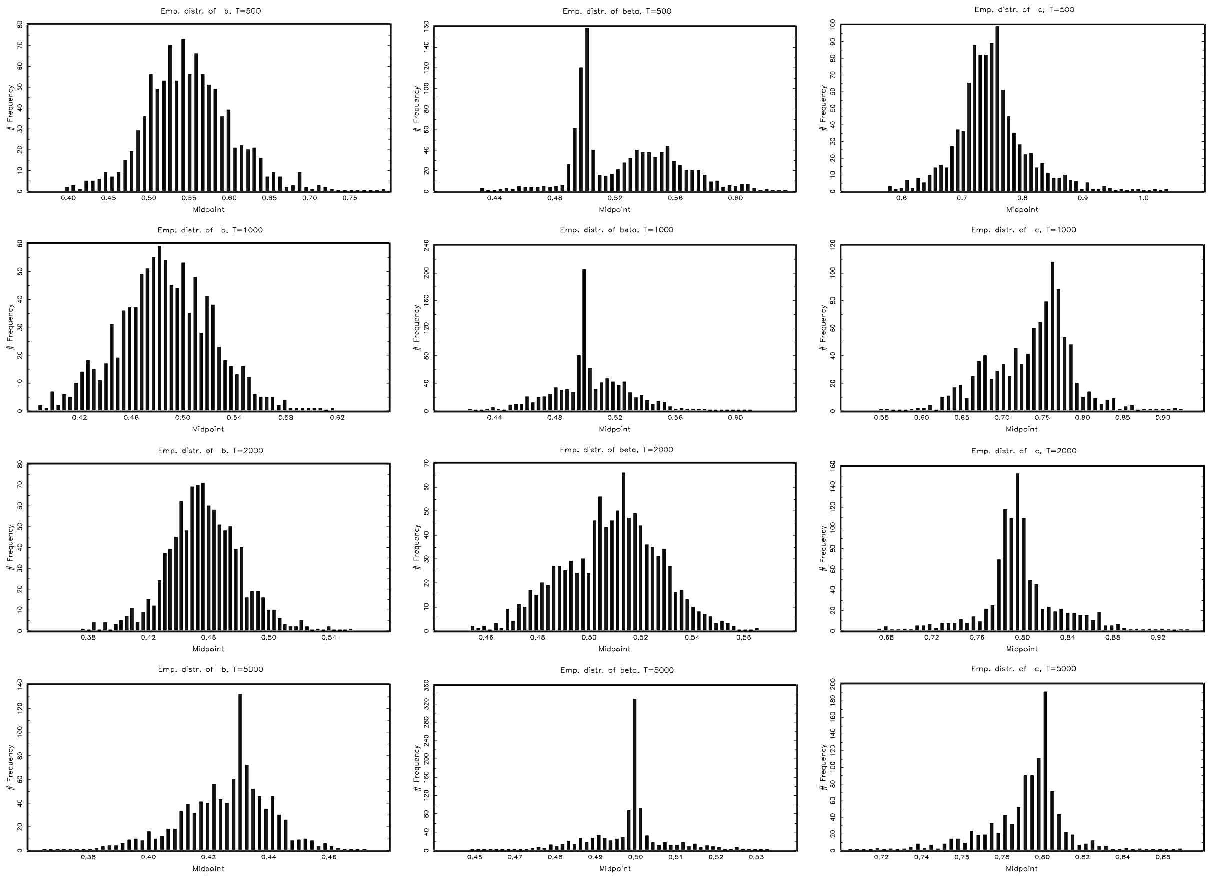


Figure 20: Marginal empirical distribution of SMM, set II
(parameter per row, size per column:500,1000,2000,5000).

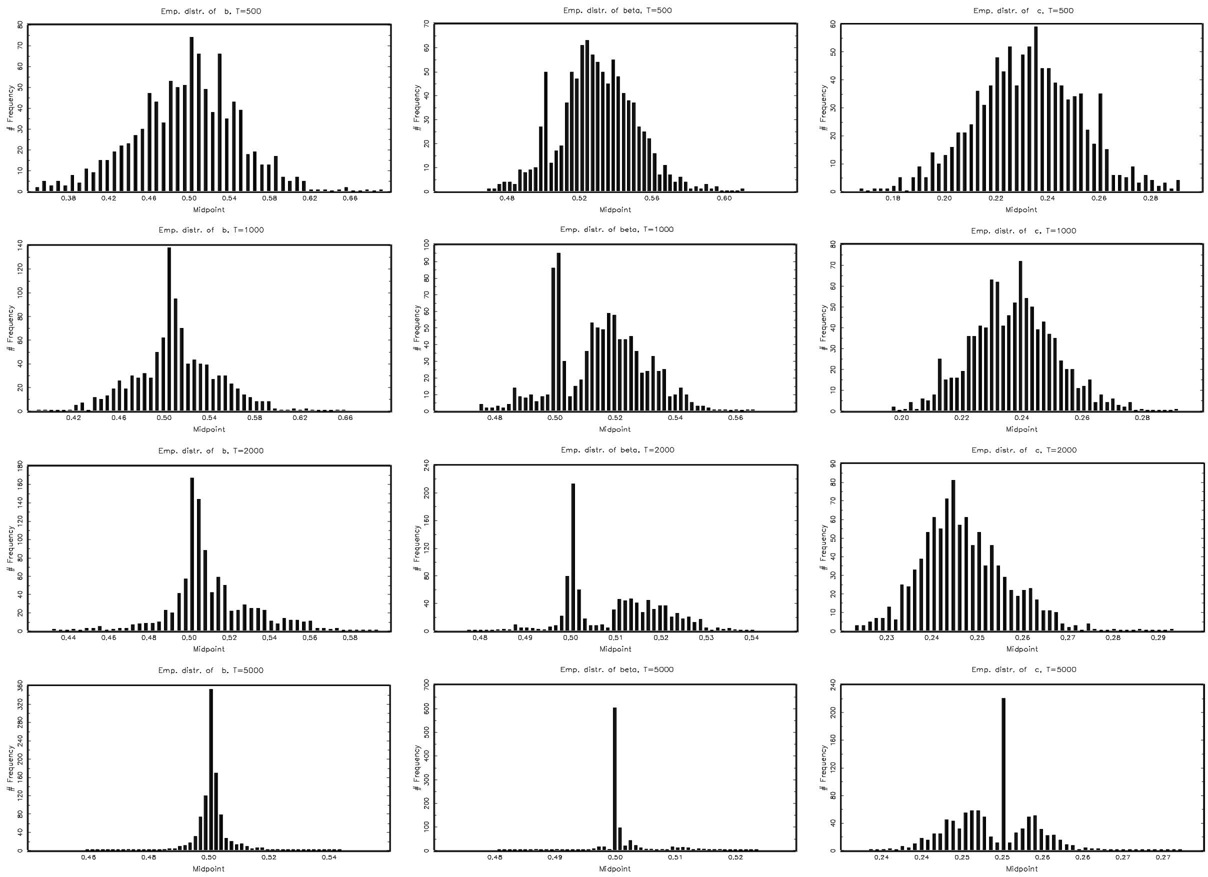


Figure 21: Marginal empirical distribution of SMM, set III
(parameter per row, size per column:500,1000,2000,5000).

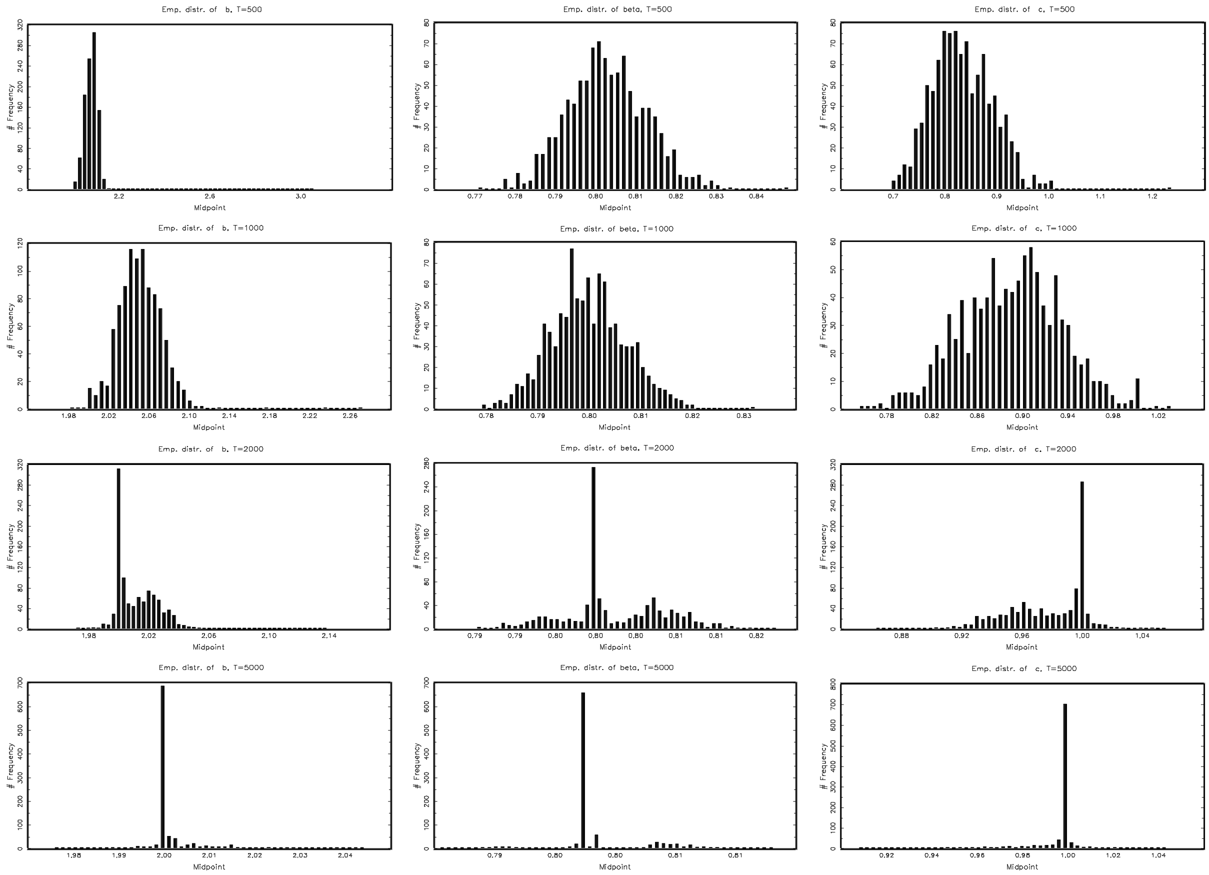
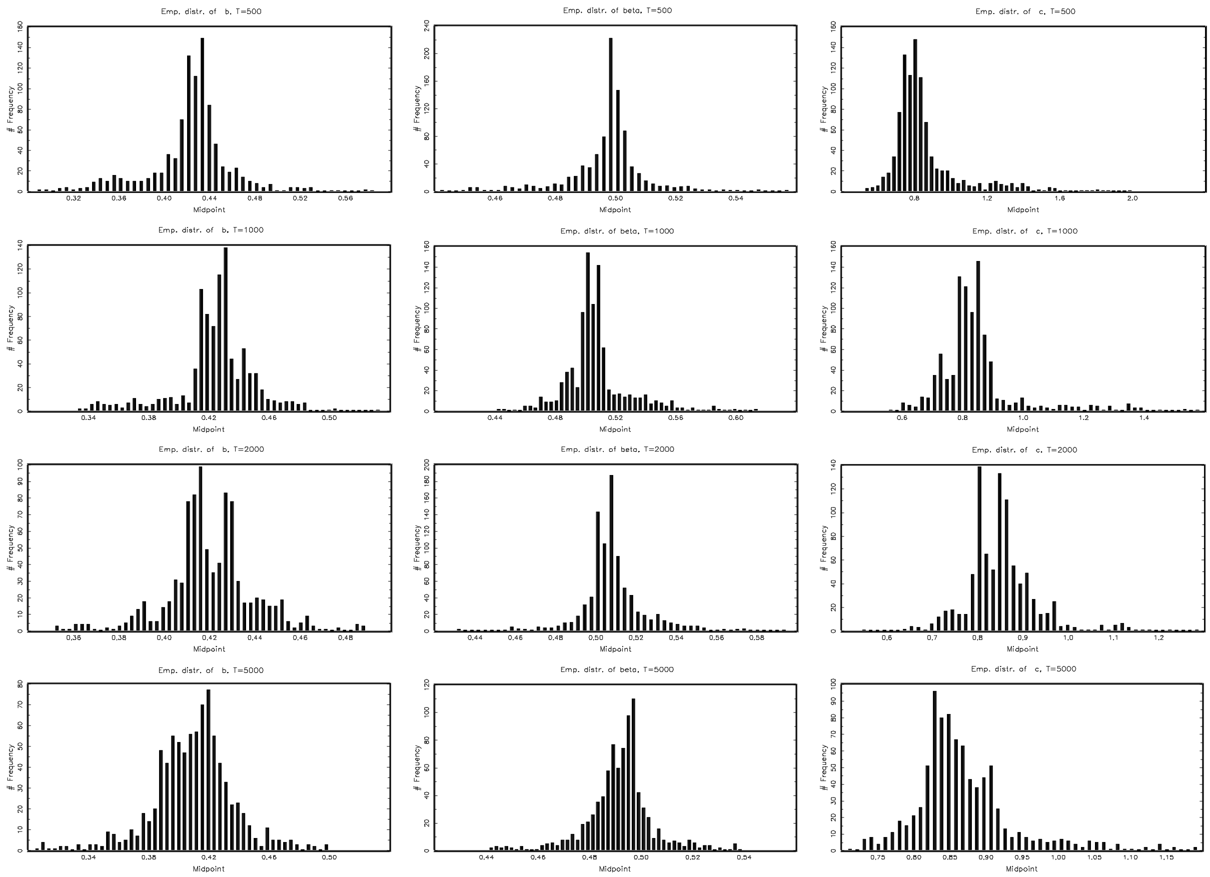


Figure 22: Marginal empirical distribution of II, set I
(parameter per row, size per column:500,1000,2000,5000).



A. Appendix

Derivation of the conditional joint p.d.f of the pair $(\mathbf{y_{1t}}, \mathbf{y_{2t}})$

Knowing that $x_{it}/\tilde{c}_h \sim \chi^2(K_i; \xi_h x_{it-h})$ with $K_i = \frac{4b\beta_i}{c}$, $\tilde{c}_h = \frac{c}{4b}(1-\exp(-bh))$ and $\xi_h = \frac{4b}{c}(\exp(bh)-1)^{-1}$ for $i = 1, 2$ whose p.d.f are given by

$$
\begin{aligned}
f(x_{it}|x_{it-h}) &= \frac{c}{4b}(1-\exp(-bh)) I_{\mathbb{R}^+}(x_{it}) \exp(-\frac{\xi_h x_{it-h}}{2}) \exp(-\frac{x_{it}}{2}) \frac{x_{it}^{K_i/2-1}}{2^{K_i/2}} \\
& \sum_{j_i=0}^{\infty} x_{it}^{j_i} \frac{(\xi_h x_{it-h})^{j_i}}{(j_i)! 2^{2j_i}} \frac{1}{\Gamma(\frac{K_i}{2}+j_i)} \,. \qquad \text{(A.67)}
\end{aligned}
$$

Since x_{1t} and x_{2t} are mutually independent and by the Jacobian formula we deduce the conditional joint p.d.f for (y_{1t}, y_{2t}), i.e.

$$
\begin{aligned}
g_{12}^{(h)}(y_{1t}, y_{2t}|y_{1t-h}, y_{2t-h}) &= \left[\frac{c}{4b}(1-\exp(-bh))\right]^2 I_{(0,1)\times\mathbb{R}^+}(y_{1t}, y_{2t}) \exp(-\xi_h \frac{y_{2t-h}}{2}) \exp(-\frac{y_{2t}}{2}) \\
& \frac{(y_{1t}y_{2t})^{\frac{K_1}{2}-1}}{2^{\frac{K_1}{2}}} \frac{(y_{2t}(1-y_{1t}))^{\frac{K_2}{2}-1}}{2^{\frac{K_2}{2}}} \\
& \sum_{i_1=0}^{\infty} \sum_{i_2=0}^{\infty} (y_{1t}y_{2t})^{i_1} \frac{(\xi_h y_{1t-h} y_{2t-h})^{i_1}}{i_1! 2^{2i_1}} \\
& (y_{2t}(1-y_{1t}))^{i_2} \frac{(\xi_h y_{2t-h}(1-y_{1t-h}))^{i_2}}{i_2! 2^{2i_2}} \\
& \frac{1}{\Gamma(\frac{K_1}{2}+i_1)\Gamma(\frac{K_2}{2}+i_2)} y_{2t} \qquad \text{(A.68)}
\end{aligned}
$$

Multiplying the numerator and the denominator by $(i_1+i_2)!\Gamma(\frac{K_1+K_2}{2}+i_1+i_2)$ and after rearranging the expression we get:

$$
\begin{aligned}
g_{12}^{(h)}(y_{1t}, y_{2t}|y_{1t-h}, y_{2t-h}) &= \left[\frac{c}{4b}(1-\exp(-bh))\right]^2 I_{(0,1)\times\mathbb{R}^+}(y_{1t}, y_{2t}) \exp(-\xi_h \frac{y_{2t-h}}{2}) \exp(-\frac{y_{2t}}{2}) \\
& \frac{(y_{2t})^{\frac{K_1+K_2}{2}-1}}{2^{\frac{K_1+K_2}{2}}} \sum_{i_1=0}^{\infty} \sum_{i_2=0}^{\infty} \frac{(y_{2t})^{i_1+i_2}}{(i_1+i_2)!} \frac{(\xi_h y_{2t-h})^{i_1+i_2}}{2^{2(i_1+i_2)}} \\
& \frac{1}{\Gamma(\frac{K_1+K_2}{2}+i_1+i_2)} \frac{1}{B(\frac{K_1}{2}+i_1, \frac{K_2}{2}+i_2)} (y_{1t})^{\frac{K_1}{2}+i_1-1} (1-y_{1t})^{\frac{K_2}{2}+i_2-1} \\
& \binom{i_1+i_2}{i_1} (y_{1t-h})^{i_1} (1-y_{1t-h})^{i_2} \,. \qquad \text{(A.69)}
\end{aligned}
$$

Letting $h \to \infty$ in equation (A.68) and after a few manipulations we deduce the marginal distributions.

$$\begin{aligned} g_{12}(y_{1t}, y_{2t}) &= (\frac{c}{4b})^2 \frac{1}{2} I_{(0,1)\times\mathbb{R}^+}(y_{1t}, y_{2t})(y_{1t})^{\frac{K_1}{2}-1}(1-y_{1t})^{\frac{K_2}{2}-1} \\ &\quad \frac{1}{\Gamma(\frac{K_1}{2})\Gamma(\frac{K_2}{2})}(\frac{y_{2t}}{2})^{\frac{K_1+K_2}{2}-1}\exp(-\frac{y_{2t}}{2}) \ . \end{aligned} \tag{A.70}$$

Integrating the above equation w.r.t. y_{2t} yields

$$g_1(y_{1t}) = (\frac{c}{4b})^2 I_{(0,1)}(y_{1t}) y_{1t}^{\frac{K_1}{2}-1}(1-y_{1t})^{\frac{K_2}{2}-1}\frac{\Gamma(\frac{K_1+K_2}{2})}{\Gamma(\frac{K_1}{2})\Gamma(\frac{K_2}{2})} \tag{A.71}$$

where we recognize a beta distribution $B(\frac{K_1}{2}, \frac{K_2}{2})$ for $(\frac{4b}{c})^2 y_{1t}$ and a gamma distribution $\gamma(\frac{K_1+K_2}{2})$ for $\frac{y_{2t}}{2}$ as induced by the transformations $y_{1t} = \frac{x_{1t}}{x_{1t}+x_{2t}}$ and $y_{2t} = x_{1t} + x_{2t}$.

$\square$

Derivation of the stationary distribution of the Jacobi process from the subordinated process $\mathbf{y}^*_{\mathbf{1t}}$.

The stationary p.d.f of y_{1t} defined at equation (2.3) is independent from y_{2t} since

$$\begin{aligned} p(y_{1t}) &= \frac{m(y_{1t})}{\int_0^1 m(y_{1t})dy_{1t}} = \frac{\frac{1}{s(y_{1t})\frac{cy_{1t}}{y_{2t}}(1-y_{1t})}}{\int_0^1 \frac{1}{s(y_{1t})\frac{cy_{1t}}{y_{2t}}(1-y_{1t})}dy_{1t}} \\ &= \frac{\frac{1}{s(y_{1t})cy_{1t}(1-y_{1t})}}{\int_0^1 \frac{1}{s(y_{1t})cy_{1t}(1-y_{1t})}dy_{1t}} \end{aligned} \tag{A.72}$$

where

$$\begin{aligned} s(y_{1t}) &= \exp(-\int^{y_{1t}} \frac{\frac{-2b}{u_2}(u_1(\beta_1+\beta_2)-\beta_1)}{c\frac{u_1}{u_2}(1-u_1)}du_1) \\ &= \exp(-\int^{y_{1t}} \frac{-2b(u_1(\beta_1+\beta_2)-\beta_1)}{cu_1(1-u_1)}du_1 \ . \end{aligned} \tag{A.73}$$

Therefore y_{1t} and y^*_{1t} have the same stationary distribution obtained at equation (A.71) above. Thus, the subordinated process $\mathbf{y}^*_{\mathbf{1t}}$ solution of the s.d.e

$$dy^*_{1t} = -b(\beta_1+\beta_2)(y^*_{1t} - \frac{\beta_1}{\beta_1+\beta_2})dt + (cy^*_{1t}(1-y^*_{1t}))^{1/2}dW^*_{1t}$$

follows the stationary distribution $B(\frac{K_1}{2}, \frac{K_2}{2}) = B(\frac{2b\beta_1}{c}, \frac{2b\beta_2}{c})$. We can then deduce the stationary distribution for the Jacobi process solution of the s.d.e (1.1) by identifying the coefficients of both equations and we easily deduce that y_t solution of the s.d.e (1.1) has the stationary distribution $B(\frac{2b\beta}{c}, \frac{2b(1-\beta)}{c})$. $\square$

The transition density as the solution of a partial differential equation.
Using the canonical decomposition of the transition density, we have:

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

whose differentiation w.r.t h yields

$$\begin{aligned}
\frac{\partial}{\partial h} g^{(h)} &= g(y_t)\left\{\sum_{n=1}^{\infty} \lambda_n \exp(\lambda_n h) P_n(y_t) P_n(y_{t-h})\right\} \\
&= g(y_t)\left\{\sum_{n=1}^{\infty} \exp(\lambda_n h) \mathcal{A}_{t-h} P_n(y_{t-h}) P_n(y_t)\right\} \\
&= \mathcal{A}_{t-h} g(y_t)\left\{1 + \sum_{n=1}^{\infty} \exp(\lambda_n h) P_n(y_{t-h}) P_n(y_t)\right\}
\end{aligned} \tag{A.74}$$

whose the last line follows from the linearity property of the differential operator $\mathcal{A}_{t-h}$ applied to y_{t-h} which can be rewritten as:

$$\begin{aligned}
\frac{\partial}{\partial h} g^{(h)}(y_t|y_{t-h}) &= \mathcal{A}_{t-h} g^{(h)}(y_t|y_{t-h}) \\
&= m(y_{t-h}) \frac{\partial}{\partial y_{t-h}} g^{(h)}(y_t|y_{t-h}) + \frac{1}{2}\sigma^2(y_{t-h}) \frac{\partial^2}{\partial y_{t-h}^2} g^{(h)}(y_t|y_{t-h})
\end{aligned} \tag{A.75}$$

$\square$

Identification issue
Our concern here is to identify the parameter of the Jacobi process. Note that taking the unconditional expectation of equation (3.32), yields under stationarity assumption that:

$$E(Y_t) = \beta = \frac{a_{10}}{1 - a_{11}} \equiv m \, . \tag{A.76}$$

b can be deduced from a_{11} as

$$b = -\frac{\log(a_{11})}{h} \tag{A.77}$$

where the coefficients a_{11} and a_{10} can be estimated by considering the regression equation

$$Y_t = a_{11} Y_{t-h} + a_{10} + u_{1t} \; . \tag{A.78}$$

Further, c can be identified from a_{22} since

$$c = -2b - \frac{\log(a_{22})}{h} \tag{A.79}$$

where a_{22} can be estimated from the regression equation

$$Y_t^2 = a_{22} Y_{t-h}^2 + a_{21} Y_{t-h} + a_{20} + u_{2t} \; . \tag{A.80}$$

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 (A.80). □

Equivalency of the exact indirect and GMM estimation procedures

A seemingly unrelated estimation procedure applied to the system of equations below

$$Y_t = a_{11}(\theta) Y_{t-1} + a_{10}(\theta) + u_{1t} \tag{A.81}$$
$$Y_t^2 = a_{22}(\theta) Y_{t-1}^2 + a_{21}(\theta) Y_{t-1} + a_{20}(\theta) + u_{2t} \; . \tag{A.82}$$

is equivalent to :

$$\min_\theta \frac{1}{T} \sum_{t=1}^{T} U_t(\theta) U_t'(\theta) = \min_\theta \begin{pmatrix} \frac{1}{T}\sum_{t=1}^T u_{1t}^2 & \frac{1}{T}\sum_{t=1}^T u_{1t}u_{2t} \\ \frac{1}{T}\sum_{t=1}^T u_{1t}u_{2t} & \frac{1}{T}\sum_{t=1}^T u_{2t}^2 \end{pmatrix} \tag{A.83}$$

where $U_t \equiv (u_{1t}, u_{2t})'$.

First let us have an insight on the first component that is

$$\min_\theta \frac{1}{T} \sum_{t=1}^{T} u_{1t}^2 \Longleftrightarrow \min_\theta \frac{1}{T} \sum_{t=1}^{T} (Y_t - a_{11}(\theta) Y_{t-1} - a_{10}(\theta))^2 \tag{A.84}$$

$$\Longleftrightarrow \min_{\theta}\left\{\frac{1}{T}\sum_{t=1}^{T}Y_t^2 + a_{11}(\theta)^2\frac{1}{T}\sum_{t=1}^{T}Y_{t-1}^2 + a_{10}^2(\theta) - 2a_{11}(\theta)\frac{1}{T}\sum_{t=1}^{T}Y_tY_{t-1}\right.$$
$$\left. + 2a_{11}(\theta)a_{10}(\theta)\frac{1}{T}\sum_{t=1}^{T}Y_{t-1} - 2a_{10}(\theta)\frac{1}{T}\sum_{t=1}^{T}Y_t\right\} \quad \text{(A.85)}$$

Denoting by $\hat{k}_i$ the sample moments we can write equation (A.84) as:

$$\Longleftrightarrow \min_{\theta}\left\{\hat{k}_2 + a_{11}(\theta)^2\hat{k}_2 + a_{10}^2(\theta) - 2a_{11}(\theta)\hat{k}_{11} + 2a_{11}(\theta)a_{10}(\theta)\hat{k}_1 - 2a_{10}(\theta)\hat{k}_1\right\} . \quad \text{(A.86)}$$

Introducing the theoretical moments and noting that $a_{10}(\theta) = [1-a_{11}(\theta)]k_1(\theta)$ yields:

$$\min_{\theta}\left\{\hat{X}(\theta) + [1 + a_{11}^2(\theta)]k_2(\theta) - 2a_{11}(\theta)k_{11}(\theta) - a_{10}^2(\theta)\right\} , \quad \text{(A.87)}$$

where

$$\hat{X}(\theta) = [1 + a_{11}^2(\theta)](\hat{k}_2 - k_2(\theta)) - 2a_{11}(\theta)(\hat{k}_{11} - k_{11}(\theta)) - 2a_{10}(\theta)[1 - a_{11}(\theta)](\hat{k}_1 - k_1(\theta)) . \quad \text{(A.88)}$$

But the previous equation can be rewritten as:

$$\min_{\theta}\left\{\hat{X}(\theta) + E[(Y_t - a_{11}(\theta))^2] - a_{10}^2(\theta)\right\} \quad \text{(A.89)}$$

which is equivalent to

$$\min_{\theta}\left\{\hat{X}(\theta) + E[(a_{10}(\theta) + u_{1t})^2] - a_{10}^2(\theta)\right\} \quad \text{(A.90)}$$

which yields

$$\min_{\theta}\left\{\hat{X}(\theta) + \sigma_1^2\right\} \quad \text{(A.91)}$$

or equivalently,

$$\min_{\theta}\hat{X}(\theta) \quad \text{(A.92)}$$

whose mimimum is attained for the theoretical moments equal to the sample analogs.

Second, let us have a look at the covariance component.

$$\min_{\theta} \frac{1}{T}\sum_{t=1}^{T} u_{1t}u_{2t} \Longleftrightarrow \min_{\theta} \frac{1}{T}\sum_{t=1}^{T}[Y_t - a_{11}(\theta)Y_{t-1} - a_{10}(\theta)][Y_t^2 - a_{22}(\theta)Y_{t-1}^2 - a_{21}(\theta)Y_{t-1} - a_{20}(\theta)] \ . \tag{A.93}$$

Expanding this product and using the notation $\hat{k}_i$ for the sample moments we get after some rearrangement:

$$\begin{aligned}\min_{\theta}\Big\{ & [1 + a_{11}(\theta)a_{22}(\theta)]\hat{k}_3 + [a_{11}(\theta)a_{21}(\theta) - a_{10}(\theta) + a_{10}(\theta)a_{22}(\theta)]\hat{k}_2 \\ & + [-a_{20}(\theta) + a_{11}(\theta)a_{20}(\theta) + a_{10}(\theta)a_{21}(\theta)]\hat{k}_1 \\ & - a_{22}(\theta)\hat{k}_{12} - a_{21}(\theta)\hat{k}_{11} - a_{11}(\theta)\hat{k}_{21} + a_{10}(\theta)a_{20}(\theta)\Big\} \ . \end{aligned} \tag{A.94}$$

Introducing the theoretical moments $k_i(\theta)$ and rearrangeing equation (A.94) yields:

$$\begin{aligned}\min_{\theta}\Big\{ & \hat{X}(\theta) + a_{10}(\theta)a_{20}(\theta) + [1 + a_{11}(\theta)a_{22}(\theta)]k_3(\theta) + [a_{11}(\theta)a_{21}(\theta) - a_{10}(\theta) + a_{10}(\theta)a_{22}(\theta)]k_2(\theta) \\ & + [-a_{20}(\theta) + a_{11}(\theta)a_{20}(\theta) + a_{10}(\theta)a_{21}(\theta)]k_1(\theta) \\ & - a_{22}(\theta)k_{12}(\theta) - a_{21}(\theta)k_{11}(\theta) - a_{11}(\theta)k_{21}(\theta)\Big\} \end{aligned} \tag{A.95}$$

where

$$\begin{aligned}\hat{X}(\theta) = {} & [1 + a_{11}(\theta)a_{22}(\theta)](\hat{k}_3 - k_3(\theta)) + [a_{11}(\theta)a_{21}(\theta) - a_{10}(\theta) + a_{10}(\theta)a_{22}(\theta)](\hat{k}_2 - k_2(\theta)) \\ & + [-a_{20}(\theta) + a_{11}(\theta)a_{20}(\theta) + a_{10}(\theta)a_{21}(\theta)](\hat{k}_1 - k_1(\theta)) \\ & - a_{22}(\theta)(\hat{k}_{12} - k_{12}(\theta)) - a_{21}(\theta)(\hat{k}_{11} - k_{11}(\theta)) - a_{11}(\theta)(\hat{k}_{21} - k_{21}(\theta)) \end{aligned} \tag{A.96}$$

Using $a_{10}(\theta) = [1 - a_{11}(\theta)]k_1(\theta)$ and $a_{20}(\theta) = [1 - a_{22}(\theta)]k_2(\theta) - a_{21}(\theta)k_1(\theta)$ in the previous equation yields after some simplification:

$$\begin{aligned}\min_{\theta}\Big\{ & \hat{X}(\theta) + (1 + a_{11}(\theta)a_{22}(\theta))k_3(\theta) - a_{22}(\theta)k_{12}(\theta) - a_{11}(\theta)k_{21}(\theta) \\ & + (-1 + a_{11}(\theta) + a_{22}(\theta) - a_{11}(\theta)a_{22}(\theta))k_1(\theta)k_2(\theta) \\ & + [-a_{21}(\theta)k_{11}(\theta) + a_{21}(\theta)(1 - a_{11}(\theta))k_1^2(\theta) + a_{11}(\theta)a_{21}(\theta)k_2(\theta)]\Big\} \end{aligned} \tag{A.97}$$

where the expression in brackets cancels out after replacing $a_{11}(\theta)$ by $\frac{k_{11}(\theta)-k_1^2(\theta)}{k_2(\theta)-k_1^2(\theta)}$. Therefore, we have:

$$\min_{\theta}\left\{\hat{X}(\theta)+(1+a_{11}(\theta)a_{22}(\theta))k_3(\theta)-a_{22}(\theta)k_{12}(\theta)-a_{11}(\theta)k_{21}(\theta)-[1-a_{22}(\theta)][1-a_{11}(\theta)]k_1(\theta)k_2(\theta)\right\} \tag{A.98}$$

which can be rewritten as:

$$\min_{\theta}\left\{\hat{X}(\theta)-E[Y_t-a_{11}(\theta)Y_{t-1}]E[Y_t^2-a_{22}(\theta)Y_{t-1}^2]+E[(Y_t^2-a_{22}(\theta)Y_{t-1}^2)(Y_t-a_{11}(\theta)Y_{t-1})]\right\} \tag{A.99}$$

which is still equivalent to:

$$\min_{\theta}\left\{\hat{X}(\theta)-E[a_{10}(\theta)+u_{1t}]E[a_{21}(\theta)Y_{t-1}+a_{20}(\theta)+u_{2t}]+E[(a_{21}(\theta)Y_{t-1}+a_{20}(\theta)+u_{2t})(a_{10}(\theta)+u_{1t})]\right\} \tag{A.100}$$

which yileds

$$\min_{\theta}\hat{X}(\theta)+\sigma_{12}\Longleftrightarrow\min_{\theta}\hat{X}(\theta) \tag{A.101}$$

since $\sigma_{12}\equiv E(u_{1t}u_{2t})$ does not depend on θ.

Likewise before

$$\min_{\theta}\frac{1}{T}\sum_{t=1}^{T}u_{2t}^2\Longleftrightarrow\min_{\theta}\frac{1}{T}\sum_{t=1}^{T}(Y_t^2-a_{22}(\theta)Y_{t-1}^2-a_{21}(\theta)Y_{t-1}-a_{20}(\theta))^2 \tag{A.102}$$

yileds after some computation

$$\min_{\theta}\left\{\hat{X}(\theta)-a_{20}^2(\theta)+E[a_{21}(\theta)Y_{t-1}+a_{20}(\theta)+u_{2t}]^2+a_{21}^2(\theta)k_2(\theta)\right.$$
$$\left.+2E[-a_{21}(\theta)Y_{t-1}(Y_t^2-a_{22}(\theta)Y_{t-1}^2)]\right\}. \tag{A.103}$$

Using $Y_t^2-a_{22}(\theta)Y_{t-1}^2=a_{21}(\theta)Y_{t-1}+a_{20}(\theta)+u_{2t}$ yields after some simplification:

$$\min_{\theta}\hat{X}(\theta)+\sigma_2^2\Longleftrightarrow\min_{\theta}\hat{X}(\theta) \tag{A.104}$$

since $\sigma_2^2\equiv Var(u_{2t})$ does not depend on θ. $\square$

Moments used in the GMM estimation

$$k_1(\theta) \equiv E(Y_t) = \beta \ . \tag{A.105}$$

$$k_2(\theta) \equiv E(Y_t^2) = -\frac{[(\frac{2b}{c}\beta)^2 + \frac{2b}{c}\beta]}{(\frac{2b}{c})^2 + 3(\frac{2b}{c}) + 2} + 2\frac{[\frac{2b}{c}\beta + 1]}{\frac{2b}{c} + 2} k_1(\theta) \ . \tag{A.106}$$

$$\begin{aligned} k_3(\theta) \equiv E(Y_t^3) \quad = \quad & \frac{[(\frac{2b}{c}\beta)^3 + 3(\frac{2b}{c}\beta)^2 + 2(\frac{2b}{c}\beta)]}{(\frac{2b}{c}+4)(\frac{2b}{c}+3)(\frac{2b}{c}+2)} \\ & -3\frac{[(\frac{2b}{c}\beta)^2 + 3(\frac{2b}{c}\beta) + 2]}{(\frac{2b}{c}+4)(\frac{2b}{c}+3)} k_1(\theta) \\ & +3\frac{\frac{2b}{c}\beta + 2}{\frac{2b}{c} + 4} k_2(\theta) \end{aligned} \tag{A.107}$$

$$\begin{aligned} k_4(\theta) \equiv E(Y_t^4) \quad = \quad & -\frac{[(\frac{2b}{c}\beta)^4 + 6(\frac{2b}{c}\beta)^3 + 11(\frac{2b}{c}\beta)^2 + 6(\frac{2b}{c}\beta)]}{(\frac{2b}{c}+6)(\frac{2b}{c}+5)(\frac{2b}{c}+4)(\frac{2b}{c}+3)} \\ & +4\frac{[(\frac{2b}{c}\beta)^3 + 6(\frac{2b}{c}\beta)^2 + 11(\frac{2b}{c}\beta) + 6]}{(\frac{2b}{c}+4)(\frac{2b}{c}+5)(\frac{2b}{c}+6)} k_1(\theta) \\ & -6\frac{[(\frac{2b}{c}\beta)^2 + 5(\frac{2b}{c}\beta) + 6]}{(\frac{2b}{c}+5)(\frac{2b}{c}+6)} k_2(\theta) + 4\frac{\frac{2b}{c}\beta + 3}{\frac{2b}{c} + 6} k_3(\theta) \end{aligned} \tag{A.108}$$

$$k_{11}^{(h)}(\theta) \equiv E[y_t y_{t-h}] = \exp(-bh) k_2(\theta) + [1 - \exp(-bh)]\beta^2 \tag{A.109}$$

$$k_{12}^{(h)}(\theta) \equiv E[y_t y_{t-h}^2] = \exp(-bh) k_3(\theta) + [1 - \exp(-bh)]\beta k_2(\theta) \tag{A.110}$$

$$\begin{aligned} k_{21}^{(h)}(\theta) \equiv E[y_t^2 y_{t-h}] \quad = \quad & -\frac{[(\frac{2b}{c}\beta)^2 + (\frac{2b}{c}\beta)]}{[(\frac{2b}{c})^2 + 3(\frac{2b}{c}) + 2]}(1 - \exp[(-2b-c)h]) k_1(\theta) \\ & +2\frac{[\frac{2b}{c}\beta^2 + \beta]}{\frac{2b}{c} + 2}[1 - \exp(-bh)] k_1(\theta) \\ & +2\frac{\frac{2b}{c}\beta + 1}{\frac{2b}{c} + 2}\{\exp(-bh) - \exp[(-2b-c)h]\} k_2(\theta) + \exp[(-2b-c)h] k_3(\theta) \end{aligned} \tag{A.111}$$

$$k_{22}^{(h)}(\theta) \equiv E[y_t^2 y_{t-h}^2] = \frac{-\left[(\frac{2b}{c}\beta)^2 + \frac{2b}{c}\beta\right]}{(\frac{2b}{c})^2 + 3(\frac{2b}{c}) + 2}(1 - \exp[(-2b - c)h])k_2(\theta)$$
$$+2\frac{\left[(\frac{2b}{c})\beta^2 + \beta\right]}{\frac{2b}{c} + 2}[1 - \exp(-bh)]k_2(\theta)$$
$$+2\frac{\left[(\frac{2b}{c})\beta + 1\right]}{\frac{2b}{c} + 2}\{\exp(-bh) - \exp[(-2b - c)h]\}k_3(\theta) + \exp[(-2b - c)h]k_4(\theta) \quad \text{(A.112}$$

$\square$

References

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.

BACON, D. W., AND D. G. WATTS (1971): “Estimating the transition between two intersecting straight lines,” *Biometrika*, 58, 525–534.

BAUM, L. E., AND T. PETRIE (1966): “Statistical inference for probabilistic functions of a finite state,” *Annals of mathematical statistics*, 37, 1554–1563.

BEC, F., M. BEN SALEM, AND M. CARRASCO (2001): “Tests for unit-root versus Threshold specification with an application to the PPP,” Discussion paper, University of Rochester.

BREEDEN, D. (1979): “An intertemporal asset pricing model with stochastic consumption and investment opportunities,” *Journal of Financial Economics*, 7, 265–296.

CAGETTI, M., L. HANSEN, T. SARGENT, AND N. WILLIAMS (2002): “Robustness and Pricing with Uncertain Growth.,” *The Review of Financial Studies*, 15.

CARRASCO, M., AND J.-P. FLORENS (2000): “Generalization of GMM to a continuum of moment conditions,” *Econometric Theory*, 16, 797–834.

CHAN, K. S., AND H. TONG (1986): “On estimating Thresholds in autoregressive models,” *Journal of time Series analysis*, 7, 178–190.

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.

DAVID, A. (1997): “Fluctuating confidence in stock markets: Implications for returns and volatility,” *Journal of Financial and Quantitative Analysis*, 32(4), 457–462.

Diebold, F., X., J.-H. Lee, and C. Weinbach, G. (1994): *Nonstationary time series analysis and cointegration*chap. Regime switching with time-varying transition probabilities. Oxford University Press, c. hargreaves edn., 283-302.

Filardo, A., J. (1994): "Business cycle phases and their transitional dynamics," *Journal of business and economic statistics*, 12, 299–308.

Garcia, R., and P. Perron (1996): "An analysis of the real interest rate under regime shifts," *Review of economic and statistics*, 78, 111–125.

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.

Granger, C. W. J., and T. Terasvirta (1993): *Modelling nonlinear Economic relationship.* Oxford University Press, Oxford.

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, L., and J. Scheinkman (1995): "Back to the Future: Generating Moment Implications for Continuous Time Markov Processes.," *Econmetrica*, 73, 767–807.

Lancaster, H. (1968): "The structure of bivariate distributions," *Annals of mathematical statistics*, 3, 311–340.

Lo, A. W. (1988): "Maximum likelihood estimation of generalized Ito processes with discretely sampled data," *Econometric Ttheory*, 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.

Maddala, D. S. (1977): *Econometrics.* McGraw-Hill, New York.

Michael, P., A. R. Nobay, and D. A. Peel (1997): "Transactions costs and nonlinear adjustment in real exchange rates: an empirical investigation," *Journal of Political Economy*, 105(4), 862–879.

Rose, A. K. (1988): "Is the real interest rate stable ?," *Journal of Finance*, 43, 1095–1112.

Sclove, S. L. (1983): "Time-series segmentation: a model and a method," *Information sciences*, 29, 7–25.

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.

Van Dijk, D., and P. H. Franses (1999): "Modeling multiple regimes in the business cycle," *Macroeconomic Dynamics*, 3, 311–340.

Vasicek, O. (1977): "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics*, 5, 177–188.

Veronesi, P. (1999): "Stock market overreaction to bad news in good times: a rational expectations equilibrium model," *The Review of Financial Studies*, 12(5), 975–1007.

——— (2002): "The Peso Problem Hypothesis and Stock market returns: New implications from a rational expectations equilibrium model," *Forthcoming on the Journal of Economic Dynamics and Control.*

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.