

# Efficient estimation of general dynamic models with a continuum of moment conditions

Marine Carrasco<sup>a,\*</sup>, Mikhail Chernov<sup>b</sup>,  
Jean-Pierre Florens<sup>c</sup>, Eric Ghysels<sup>d</sup>

<sup>a</sup>*Université de Montréal, Montréal, QC H3C3J7, Canada*

<sup>b</sup>*Columbia Graduate School of Business, New York, NY 10027, USA*

<sup>c</sup>*University of Toulouse, 31000 Toulouse, France*

<sup>d</sup>*University of North Carolina, Chapel Hill, NC 27599, USA*

---

## Abstract

There are two difficulties with the implementation of the characteristic function-based estimators. First, the optimal instrument yielding the ML efficiency depends on the unknown probability density function. Second, the need to use a large set of moment conditions leads to the singularity of the covariance matrix. We resolve the two problems in the framework of GMM with a continuum of moment conditions. A new optimal instrument relies on the double indexing and, as a result, has a simple exponential form. The singularity problem is addressed via a penalization term. We introduce HAC-type estimators for non-Markov models. A simulated method of moments is proposed for non-analytical cases.

*JEL Classification:* C13; C22; G12

*Keywords:* Characteristic Function; Efficient Estimation; Affine Models

---

## 1 Introduction

This paper proposes a general estimation approach which combines the attractive features of method of moments estimation with the efficiency of the maximum likelihood estimator (MLE) in one framework. The method exploits the moment conditions computed via the characteristic function (CF) of a stochastic process instead of the likelihood function, as in the recent work by Chacko and Viceira (2003), Jiang and Knight (2002), and Singleton (2001). The most obvious advantage of such an approach is that in many cases of practical interest the CF is available in analytic form, while the likelihood is not. Moreover, the CF contains the same information as the likelihood function. Therefore, a clever choice of moment conditions should provide the same efficiency as ML.

The main contribution of this paper is the resolution of two major difficulties with the estimation via the CF. The first one is related to the intuition that the more moments one generates by varying the CF argument, the more information one uses, and, therefore, the estimator becomes more efficient. However, as one refines the finite grid of CF argument values, the associated covariance matrix approaches singularity. The second difficulty is that in addition to a large set of CF-based moment conditions, one requires an optimal instrument to achieve the ML efficiency. Prior work (e.g. Feuerverger and McDunnough, 1981 or Singleton, 2001) derived the optimal instrument, which is a function of the unknown probability density. Such an estimator is clearly hard to implement.

---

\* We would like to thank Yacine Aït-Sahalia, Lars Hansen, Mike Johannes, Nour Meddahi, Antonio Mele, Jun Pan, Benoit Perron, Eric Renault, Tano Santos, Ken Singleton, the conference and seminar participants at the Canadian Econometric Study Group in Quebec city, CIREQ-CIRANO-MITACS Conference on Univariate and Multivariate Models for Asset Pricing, Econometric Society North American winter meeting in New Orleans, Western Finance Association in Tucson, Chicago, Princeton, ITAM, Michigan State, Montréal, and USC for comments. We are grateful to Ken Singleton for the code implementing the approximately efficient estimator. Ruslan Bikbov and Jeremy Petranka have provided outstanding research assistance. Carrasco acknowledges financial support from the National Science Foundation, grant SES-0211418.

\* Corresponding author. Université de Montréal, Département de Sciences Economiques, CP 6128, succ. centre ville, Montréal QC H3C3J7, Canada. Tel.: +1-514-343-2394; fax: +1-514-343-7221

*Email addresses:* marine.carrasco@umontreal.ca (Marine Carrasco), mc1365@columbia.edu (Mikhail Chernov), florens@cict.fr (Jean-Pierre Florens), eghysels@unc.edu (Eric Ghysels).

We use the extension of GMM to a continuum of moment conditions (referred to as C-GMM) of Carrasco and Florens (2000). Instead of selecting a finite number of grid points, the whole continuum of moment conditions is used, guaranteeing the full efficiency of the estimator. To implement the optimal C-GMM estimator, it is necessary to invert the covariance operator, which is the analog of the covariance matrix in finite dimension. Because of the infinity of moments, the covariance operator is nearly singular and its inverse is highly unstable. To stabilize the inverse, we introduce a penalization parameter,  $\alpha_T$ . This term may be thought of as the counterpart of the grid width in the discretization. We document the rate of convergence of  $\alpha_T$  and give a heuristic method for selecting it via bootstrap.

In order to find an implementable optimal instrument, our paper provides various extensions of the initial work by Carrasco and Florens (2000). While the original work deals with iid data, we derive the asymptotic properties of the C-GMM estimator applied to weakly dependent data and correlated moment functions. The moment functions may be complex valued and be functions of an index parameter taking its values in  $\mathbb{R}^d$  for an arbitrary  $d \geq 1$  in order to accommodate the specific features of CF. To solve for the optimal instrument, we distinguish two cases depending on whether the observable variables are Markov or not.

In the Markov case, the moment conditions are based on conditional CF. We propose to span the unknown optimal instrument by an infinite basis consisting of simple exponential functions. Since the estimation framework already relies on a continuum of moment conditions, adding a continuum of spanning functions does not pose any problems. As a result, we achieve ML efficiency when we use the values of conditional CF indexed by its argument as moment functions. We propose a simulated method of moments type estimator for the cases when the CF is unknown. If one is able to draw from the true conditional distribution, then the conditional CF can be estimated via simulations and ML efficiency obtains. This approach can be thought as a simple alternative to the Indirect Inference proposed by Gouriéroux, Monfort and Renault (1993), the Efficient Method of Moments (EMM) suggested by Gallant and Tauchen (1996), and the nonparametric simulated maximum likelihood method (Fermanian and Salanie, 2004).<sup>1</sup>

---

<sup>1</sup> Similarly, Altissimo and Mele (2005) have recently proposed a method to estimate diffusions

If the observations are not Markov, it is in general not possible to construct the conditional CF.<sup>2</sup> Therefore, we estimate our parameter using the joint CF of a particular number of data lags. The resulting moment function is not a martingale difference sequence. A remarkable feature of the joint CF is that the usual GMM first-stage estimator is not required. While we were not able to obtain optimal moment functions yielding ML efficiency in this case, we derived an upper bound on the variance of the resulting estimator. In the worst case scenario, if one uses the joint CF for estimation, the variance of the C-GMM estimator corresponds to that of the ML estimator based on the joint density of the same data lags. As the joint CF is usually unknown, a simulated method of moments becomes handy to do inference in the nonmarkovian case. The simulation scheme differs from that used in the Markov case. Instead of simulating conditionally on the observable data, we simulate the full time-series as it is done in Duffie and Singleton (1993).

The paper is organized as follows. The first section reviews issues related to the estimation via CF and discusses the estimator in the most simple case of moments forming martingale difference sequences. Section 3 extends the C-GMM proposed by Carrasco and Florens (2000) to the case where the moment functions are correlated. It shows how to estimate the long-run covariance and how to implement the C-GMM estimator in practice. Section 4 specializes to the cases where the moment conditions are based either on the conditional characteristic function or joint characteristic function. In Section 4.1, we derive the ML-efficient estimator based on the conditional CF. Then, in Section 4.2, we discuss the properties of the estimator based on joint CF, which is relevant for processes with latent variables. Section 5 establishes the properties of the simulation-based estimators when CF is not available in analytic form. Finally, a Monte Carlo comparison of the C-GMM estimator with other popular estimators is reported in Section 6. The last section concludes. All regularity conditions are collected in Appendix A. All the proofs are provided in Appendix B.

---

efficiently. It consists in minimizing the distance between two kernel estimates of the conditional density, one based on the actual data and the other based on simulated data.

<sup>2</sup> Bates (2003) was able to construct a conditional MLE estimator for certain types of affine models based on the CF.

## 2 Estimation methods based on the Characteristic Function

In this section we discuss the major unresolved issues pertaining to estimation via CF and explain how we propose to tackle them via GMM based on the continuum of moment conditions (C-GMM).

### 2.1 Preliminaries

We assume that the stationary Markov process  $X_t$  is a  $p \times 1$ -vector of random variables which represents the data-generating process indexed by a finite dimensional parameter  $\theta \in \Theta \subset \mathbb{R}^q$ . Suppose the sequence  $X_t, t = 1, \dots, T$  is observed. Singleton (2001) proposes an estimator based on the conditional characteristic function (CCF). The CCF of  $X_{t+1}$  given  $X_t$  is defined as

$$\psi_\theta(s|X_t) \equiv E^\theta \left( e^{isX_{t+1}} | X_t \right) \quad (1)$$

and is assumed to be known. If it is not known, it can be easily recovered by simulations. Equation (1) implies that the following unconditional moment conditions are satisfied:

$$E^\theta \left[ (e^{isX_{t+1}} - \psi_\theta(s|X_t))A(X_t) \right] = 0 \text{ for all } s \in \mathbb{R}^p$$

where  $A(X_t)$  is an arbitrary instrument. Let  $Y_t = (X_t, X_{t+1})'$ . There are two issues of interest here: the choice of  $s$  and the choice of the instrument  $A(X_t)$ . Besides being a function of  $X_t$ ,  $A$  may be a function of an index  $r$  either equal to or different from  $s$ . The following two types of unconditional moment functions are of particular interest:

**SI** – the **Single Index** moment functions:  $h(s, Y_t; \theta) = A(s, X_t) \left( e^{isX_{t+1}} - \psi_\theta(s|X_t) \right)$  where  $s \in \mathbb{R}^p$  and  $A(s, X_t) = \overline{A(-s, X_t)}$

**DI** – the **Double Index** moment functions:  $h(\tau, Y_t; \theta) = A(r, X_t) \left( e^{isX_{t+1}} - \psi_\theta(s|X_t) \right)$  where  $\tau = (r, s)' \in \mathbb{R}^{2p}$  and  $A(r, X_t) = \overline{A(-r, X_t)}$

Note that in either case, the sequence of moment functions  $\{h(\cdot, Y_t; \theta)\}$  is a martingale difference sequence with respect to the filtration  $I_t = \{X_t, X_{t-1}, \dots, X_1\}$ , hence it is uncorrelated. We now discuss which choice of instruments  $A$  is optimal, i.e. yields an efficient GMM-CCF estimator, where “efficient” means as efficient as the MLE.

## 2.2 Single index moment functions

Feuerverger and McDunnough (1981), and Singleton (2001) show that the optimal **SI** instrument is:

$$A(s, X_t) = \frac{1}{(2\pi)^p} \int e^{-isx} \frac{\partial \ln f_\theta}{\partial \theta}(x|X_t) dx. \quad (2)$$

The obvious drawback of this instrument is that it requires the knowledge of the unknown conditional likelihood function,  $f_\theta$ , of  $X_{t+1}$  given  $X_t$ .<sup>3</sup> Singleton (2001) addresses the problem of the unknown score by discretizing over  $\tau$ . To simplify the exposition, assume momentarily that  $p = 1$ . The method consists in dividing an interval  $[-M\delta, M\delta] \subset \mathbb{R}$  into  $(2M + 1)$  equally spaced intervals of width  $\delta$ . Let  $\tau_j = -M\delta + \delta(j - 1)$ ,  $j = 1, 2, \dots, 2M + 1$  be the grid points. Let  $\rho(Y_t, \theta)$  be the  $(2M + 1)$  vector with element  $(e^{i\tau_j X_{t+1}} - \psi_\theta(\tau_j|X_t))$ . Applying the results of Hansen (1985), the optimal instrument for this finite set of moment conditions is

$$A(X_t) = E[\nabla_\theta \rho|X_t] E[\rho \rho'|X_t]^{-1} \quad (3)$$

which can be explicitly computed as a function of  $\psi_\theta$ .<sup>4</sup> Singleton shows that the estimator solution of  $\frac{1}{T} \sum_{t=1}^T A(X_t) \rho(Y_t, \theta) = 0$  has an asymptotic variance  $V_\delta^L$  that converges to the Cramer Rao bound as  $M$  approaches infinity and  $\delta$  goes to zero. However, no rates of convergence for  $M$  and  $\delta$  are provided. In practice, if the grid is too fine ( $\delta$  small), the

<sup>3</sup> There are certain parallels between the raised issues and the estimation of univariate subordinated diffusions via an infinitesimal generator in Conley, Hansen, Luttmer, and Scheinkman (1997). They show that, assuming a continuous sampling, constructing moment conditions by applying the generator to the likelihood score of the marginal distribution is optimal and, in particular, is more efficient than building moments via the score directly. Being unable to implement in practice the corresponding optimal instrument (or test function) for the discrete sampling case, they still use the score for the empirical application.

<sup>4</sup>  $A$  depends on the unknown  $\theta$  but an estimator of  $A$  can be obtained by replacing  $\theta$  by a consistent first step estimator.

covariance matrix  $E[\rho\rho'|X_t]$  becomes singular and Singleton's estimator is not feasible. The second caveat is that optimal instruments depend on the selected grid, i.e. as one refines the grid, new instruments have to be selected. Therefore, it is not clear how it is going to impact the estimator in practice.

In this paper, we will be able to address the two raised issues (i) optimal selection of instrument and (ii) potential covariance matrix singularity, without relying on the unknown probability density function.

### 2.3 Double index moment functions

When  $r = s$  is not imposed, there is a choice of instrument that does not depend on the unknown p.d.f., while attaining the ML-efficiency. The optimal **DI** instrument is

$$A(r, X_t) = e^{irX_t}. \quad (4)$$

It gives rise to a double index moment function

$$h_t(\tau; \theta) = (e^{isX_{t+1}} - \psi_\theta(s|X_t))e^{irX_t} \quad (5)$$

where  $\tau = (r, s)' \in \mathbb{R}^{2p}$ . Such a choice of instrument is quite intuitive. Although we can not construct the optimal instrument in (2), we can span it via a set of basis functions  $\{\exp(irX_t)\}$ . The resulting GMM estimator will be as efficient as the MLE provided that the full continuum of moment conditions indexed by  $\tau$  is used. For this purpose, we use the method proposed by Carrasco and Florens (2000). This approach has two advantages: (i) the instrument  $\exp(irX_t)$  has a simple form; (ii) in contrast to Singleton's instrument (3), it does not depend on the discretization grid involved in numerical implementation of integration over  $\tau$ .<sup>5</sup>

---

<sup>5</sup> However, as we will see below, a smoothing parameter is introduced to be able to handle the full continuum of moments.

In the sequel we extend the C-GMM methodology of Carrasco and Florens (2000) so that it could be applicable to the moment function (5).

### 3 C-GMM with dependent data

This section will extend the results of Carrasco and Florens (2000) from the i.i.d. case to the case where the data are weakly dependent. We also allow the moment functions to be complex valued and be functions of an index parameter taking its values in  $\mathbb{R}^d$  for an arbitrary  $d \geq 1$  in order to accommodate the specific features of CF. The results of this section are not limited to estimation using CF but apply to a wide range of moment conditions. The first subsection proves the asymptotic normality and consistency of the C-GMM estimator, introduces the covariance operator and its regularized version, which is known to yield the C-GMM estimator with the smallest variance. The next subsection derives the convergence rate of the estimator of the covariance operator. The third subsection proposes a simple way to compute the C-GMM objective function in terms of matrices and vectors. The last subsection discusses the choice of moment conditions to achieve ML efficiency.

#### 3.1 General asymptotic theory

The data are assumed to be weakly dependent (see Assumption A.1 for a formal definition). The C-GMM estimator is based on the arbitrary set of moment conditions:

$$E^{\theta_0} h_t(\tau; \theta_0) = 0 \tag{6}$$

where  $h_t(\tau; \theta) \equiv h(\tau, Y_t; \theta)$  with  $Y_t = (X_t, X_{t+1}, \dots, X_{t+L})'$  for some finite integer  $L$ , and index  $\tau \in \mathbb{R}^d$ .<sup>6</sup> As a function of  $\tau$ ,  $h_t(\cdot; \theta_0)$  is supposed to belong to the set  $\mathbb{L}^2(\pi)$  as described in Definition A.2. Moreover all parameters are identified by the moment conditions (6), see Assumption A.3. Let  $\hat{h}_T(\tau; \theta_0) = \sum_{t=1}^T h_t(\tau; \theta_0)/T$ . In the sequel, we write the functions

<sup>6</sup> In the previous section we discussed the case corresponding to  $L = 1$ .



$h_t(\cdot; \theta_0)$ ,  $\hat{h}_T(\cdot; \theta_0)$  as  $h_t(\theta_0)$  and  $\hat{h}_T(\theta_0)$  or to simplify  $h_t$  and  $\hat{h}_T$ .  $\{h_t(\theta_0)\}$  is supposed to satisfy the set of Assumptions A.4, in particular  $h_t$  should be a measurable function of  $Y_t$ . Since  $L$  is finite,  $h_t$  inherits the mixing properties of  $X_t$ . Finally,  $h_t$  is assumed to be scalar because the CF itself is scalar and hence we do not need results for a vector  $h_t$ . If  $h_t$  is a vector, we can get back to a scalar function by defining  $\tilde{h}_t(i, \tau)$  as the  $i$ th component of  $h_t(\tau)$ , then  $\tilde{h}_t$  is a scalar function indexed by  $(i, \tau) \in \{1, 2, \dots, M\} \times \mathbb{R}^d$ .

These assumptions allow us to establish the asymptotic normality of the moment functions.

**Lemma 3.1** *Under regularity conditions A.1 to A.3, and A.4(i)(ii) we have*

$$\sqrt{T}\hat{h}_T(\theta_0) \Rightarrow \mathcal{N}(0, K)$$

as  $T \rightarrow \infty$ , in  $\mathbb{L}^2(\pi)$  where  $\mathcal{N}(0, K)$  is the Gaussian random element of  $\mathbb{L}^2(\pi)$  with a zero mean and the covariance operator  $K : \mathbb{L}^2(\pi) \rightarrow \mathbb{L}^2(\pi)$  satisfying

$$(Kf)(\tau) = \sum_{j=-\infty}^{\infty} \int E^{\theta_0} [h_1(\tau, \theta_0) \overline{h_j(\lambda, \theta_0)}] f(\lambda) \pi(\lambda) d\lambda \quad (7)$$

for all  $f$  in  $\mathbb{L}^2(\pi)$ .<sup>7</sup> Moreover the operator  $K$  is a Hilbert-Schmidt operator.<sup>8</sup>

We can now establish the standard properties of GMM estimators: consistency, asymptotic normality and optimality.

**Proposition 3.1** *Assume the regularity conditions A.1 to A.4 hold. Moreover, let  $B$  be a one-to-one bounded linear operator defined on  $\mathbb{L}^2(\pi)$  or a subspace of  $\mathbb{L}^2(\pi)$ . Let  $B_T$  be a sequence of random bounded linear operators converging to  $B$ . The C-GMM estimator*

$$\hat{\theta}_T = \underset{\theta \in \Theta}{\operatorname{argmin}} \|B_T \hat{h}_T(\theta)\|$$

<sup>7</sup> Definition A.1 describes a Hilbert-space valued random element.

<sup>8</sup> For a definition and the properties of Hilbert-Schmidt operators, see Dunford and Schwartz (1988). As  $K$  is a Hilbert-Schmidt operator, it can be approached by a sequence of bounded operators denoted  $K_T$ . This property will become important when we discuss how to estimate  $K$ .

has the following properties:

(1)  $\hat{\theta}_T$  is consistent and asymptotically normal such that

$$\sqrt{T} (\hat{\theta}_T - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V)$$

with

$$\begin{aligned} V &= \left\langle BE^{\theta_0} (\nabla_{\theta} h), BE^{\theta_0} (\nabla_{\theta} h)' \right\rangle^{-1} \\ &\quad \times \left\langle BE^{\theta_0} (\nabla_{\theta} h), (BKB^*) BE^{\theta_0} (\nabla_{\theta} h)' \right\rangle \\ &\quad \times \left\langle BE^{\theta_0} (\nabla_{\theta} h), BE^{\theta_0} (\nabla_{\theta} h)' \right\rangle^{-1}. \end{aligned}$$

(2) Among all admissible weighting operators  $B$ , there is one yielding an estimator with minimal variance. It is equal to  $K^{-1/2}$ , where  $K$  is the covariance operator defined in (7).

As discussed in Carrasco and Florens (2000), the operator  $K^{-1/2}$  does not exist on the whole space  $\mathbb{L}^2(\pi)$  but only on a subset, denoted  $\mathcal{H}(K)$ , which corresponds to the so-called reproducing kernel Hilbert space (RKHS) associated with  $K$  (see Parzen, 1970, and Carrasco and Florens, 2004, for details). The inner product defined on  $\mathcal{H}(K)$  is denoted  $\langle f, g \rangle_K \equiv \langle K^{-1/2} f, K^{-1/2} g \rangle$  where  $f, g \in \mathcal{H}(K)$ . Since the inverse of  $K$  is not bounded, the regularized version of the inverse, involving a penalizing term  $\alpha_T$ , is considered. Namely, the operator  $K$  is replaced by some nearby operator that has a bounded inverse. For  $\alpha_T > 0$ , the equation:

$$(K^2 + \alpha_T I) g = K f \tag{8}$$

has a unique stable solution for each  $f \in \mathbb{L}^2(\pi)$ . The Tikhonov regularized inverse to  $K$  is given by:

$$(K^{\alpha_T})^{-1} = (K^2 + \alpha_T I)^{-1} K$$

In order to implement the C-GMM estimator with the optimal weighting operator, we have to estimate  $K$ , which can be done via a sequence of bounded operators  $K_T$  approaching  $K$  as the sample size grows because  $K$  is a Hilbert-Schmidt operator (see Lemma 3.1). We

postpone the explicit construction of  $K_T$  until the next subsection and establish first the asymptotic properties of the optimal C-GMM operator for a given  $K_T$ .

**Proposition 3.2** *Assume the regularity conditions A.1 to A.5 hold. Let  $K_T$  denote a consistent estimator of  $K$  that satisfies  $\|K_T - K\| = O_p(T^{-a})$  for some  $a \geq 0$  and  $(K_T^{\alpha_T})^{-1} = (K_T^2 + \alpha_T I)^{-1} K_T$  denote the regularized estimator of  $K^{-1}$ . The optimal GMM estimator of  $\theta$  is obtained by:*

$$\hat{\theta}_T = \underset{\theta \in \Theta}{\operatorname{argmin}} \left\| (K_T^{\alpha_T})^{-1/2} \hat{h}_T(\theta) \right\| \quad (9)$$

and satisfies  $\hat{\theta}_T \xrightarrow{P} \theta_0$  and

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \left(\left\langle E^{\theta_0}(\nabla_{\theta} h), E^{\theta_0}(\nabla_{\theta} h)' \right\rangle_K\right)^{-1}\right) \quad (10)$$

as  $T$  and  $T^a \alpha_T$  go to infinity and  $\alpha_T$  goes to zero.<sup>9</sup>

A simple estimator of the asymptotic variance of  $\sqrt{T}(\hat{\theta}_T - \theta_0)$  will be discussed in Subsection 3.3. Proposition 3.2 gives a rate of convergence of  $\alpha_T$  but does not indicate how to choose  $\alpha_T$  in practice. Recall that the estimator will be consistent for any  $\alpha_T > 0$  but its variance will be the smallest for the  $\alpha_T$  decreasing to zero at the right rate. Simulations in Carrasco and Florens (2002) and in this paper (see section 6) show that the estimator is not very sensitive to the choice of  $\alpha_T$ .

Of course a data-driven selection method of  $\alpha_T$  would be preferable. Ideally,  $\alpha_T$  should be selected so that it minimizes the mean-square error (MSE) of  $\hat{\theta}_T$ . Let  $\hat{\theta}_T^{\alpha}$  be the C-GMM estimator for a given  $\alpha$ . Then we look for  $\alpha_T$  such that

$$\alpha_T = \arg \min_{\alpha} E \left[ \left( \hat{\theta}_T^{\alpha} - \theta_0 \right)^2 \right].$$

There are two ways to estimate the unknown MSE. The first method consists in computing the MSE analytically using a second order expansion as in Donald and Newey (2001). This is the approach taken in Carrasco and Florens (2002) in an iid context and may be very

<sup>9</sup> Let  $\theta = (\theta_1, \dots, \theta_q)'$ . By a slight abuse of notation,  $\langle E^{\theta_0}(\nabla_{\theta} h), E^{\theta_0}(\nabla_{\theta} h)' \rangle_K$  in (10) denotes the  $q \times q$ -matrix with  $(i, j)$  element  $\langle E^{\theta_0}(\nabla_{\theta_i} h), E^{\theta_0}(\nabla_{\theta_j} h) \rangle_K$ .

tedious to compute in time-series. The second approach consists in approximating the MSE by block bootstrap along the line of Hall and Horowitz (1996). This second approach avoids the analytical derivation and is easier to implement.

### 3.2 Convergence rate of the estimator of the covariance operator

Note that the covariance operator defined in (7) is an integral operator that can be written as

$$Kf(\tau_1) = \int k(\tau_1, \tau_2) f(\tau_2) \pi(\tau_2) d\tau_2 \quad (11)$$

with

$$k(\tau_1, \tau_2) = \sum_{j=-\infty}^{\infty} E^{\theta_0} \left( h_t(\tau_1; \theta_0) \overline{h_{t-j}(\tau_2; \theta_0)} \right) \quad (12)$$

The function  $k$  is called the kernel of the integral operator  $K$ . We are interested in estimating the operator  $K$ .

There are two cases of interest. In the first case,  $\{h_t\}$  are martingale difference sequences of the form (5). Then the kernel of  $K$  is particularly simple and can be estimated via

$$\hat{k}_T(\tau_1, \tau_2) = \frac{1}{T} \sum_{t=1}^T h_t(\tau_1; \hat{\theta}_T^1) \overline{h_t(\tau_2; \hat{\theta}_T^1)} \quad (13)$$

given the first step estimator  $\hat{\theta}_T^1$ . The resulting estimator will satisfy  $\|K_T - K\| = O_p(T^{-1/2})$ .

In the second case, moment conditions are based on the characteristic function of  $Y_t$ . Typically, we have

$$h_t(\tau; \theta) = e^{\tau Y_t} - \psi_\theta(\tau).$$

Moment functions of this type enter in a general class where

$$h_t(\tau; \theta) = \vartheta(\tau, Y_t) - E^\theta[\vartheta(\tau, Y_t)] \quad (14)$$

where  $\vartheta$  is an arbitrary function. To estimate  $K$ , we use a kernel estimator of the type studied by Andrews (1991) but we do not need a first step estimator  $\hat{\theta}_T^1$  because  $E^\theta[\vartheta(\tau, Y_t)]$  can be

estimated by a sample mean  $\vartheta_T(\tau) = \frac{1}{T} \sum_{t=1}^T \vartheta(\tau, Y_t)$  (Parzen, 1957).

We define

$$\hat{k}_T(\tau_1, \tau_2) = \frac{T}{T-q} \sum_{j=-T+1}^{T-1} \omega\left(\frac{j}{S_T}\right) \hat{\Gamma}_T(j) \quad (15)$$

with

$$\hat{\Gamma}_T(j) = \begin{cases} \frac{1}{T} \sum_{t=j+1}^T (\vartheta(\tau_1, Y_t) - \vartheta_T(\tau_1)) \overline{(\vartheta(\tau_2, Y_{t-j}) - \vartheta_T(\tau_2))}, & j \geq 0 \\ \frac{1}{T} \sum_{t=-j+1}^T (\vartheta(\tau_1, Y_{t+j}) - \vartheta_T(\tau_1)) \overline{(\vartheta(\tau_2, Y_t) - \vartheta_T(\tau_2))}, & j < 0 \end{cases} \quad (16)$$

where  $\omega$  is a kernel and  $S_T$  is a bandwidth that will be allowed to diverge at a certain rate.

The kernel  $\omega$  is required to satisfy the regularity conditions A.6(i), which are based on the assumptions of Andrews (1991). Denote  $f(\lambda)$  the spectral density of  $Y_t$  at frequency  $\lambda$  and  $f^{(\nu)}$  its  $\nu$ th derivative at  $\lambda = 0$ . Denote  $\omega_\nu = (1/\nu!) (d^\nu \omega(x)/dx^\nu)|_{x=0}$ .

**Proposition 3.3** (i) *Let  $\{h_t\}$  be a martingale difference sequence and  $K_T$  be the integral operator with kernel (13) that depends on a first step estimator  $\hat{\theta}^1$  so that  $\|\hat{\theta}^1 - \theta_0\|_E = O_p(T^{-1/2})$  where  $\|\cdot\|_E$  denotes the Euclidean norm. Suppose that assumptions A.1 to A.5, and A.6(ii) hold. Then*

$$\|K_T - K\| = O_p(T^{-1/2}).$$

(ii) *Let  $h_t$  be given by (14) with  $|\vartheta(\tau, Y_t)| < C$  for some constant  $C$  independent of  $\tau$ . Assume that the regularity conditions A.1 to A.6(i) hold and that  $S_T^{2\nu+1}/T \rightarrow \gamma \in (0, +\infty)$  for some  $\nu \in (0, +\infty)$  for which  $\omega_\nu, \|f^{(\nu)}\| < \infty$ . Then the covariance operator with kernel (15) satisfies*

$$\|K_T - K\| = O_p(T^{-\nu/(2\nu+1)}).$$

For the Bartlett kernel,  $\nu = 1$  and for the Parzen, Tuckey-Hanning and QS kernels,  $\nu = 2$ . To obtain the result of Proposition 3.3, we have selected the value of  $S_T$  that delivers the fastest rate for  $K_T$ . For this  $S_T$ , we then select  $\alpha_T$  such that  $T^a \alpha_T$  goes to infinity according to Proposition 3.2. Instead, we could have chosen  $S_T$  and  $\alpha_T$  simultaneously. However, from Proposition 3.2, it seems that the faster the rate for  $K_T$ , the faster the rate for  $\alpha_T$ . So our approach seems to guarantee the fastest rate for  $\alpha_T$ . Note that if  $\{h_t\}$  are uncorrelated,

$a = 1/2$ . When  $\{h_t\}$  are correlated, the convergence rate of  $K_T$  is slower and accordingly the rate of convergence of  $\alpha_T$  to zero is slower.

### 3.3 Simplified Computation of the C-GMM Estimator

Carrasco and Florens (2000) propose to write the objective function in terms of the eigenvalues and eigenfunctions of the operator  $K_T^{\alpha_T}$ . The computation of eigenvalues and eigenfunctions can be burdensome, particularly in large samples. We propose here a simple expression of the objective function in terms of vectors and matrices.

Note that  $\hat{k}_T$  is a degenerate kernel that can be rewritten as

$$\hat{k}_T(\tau_1, \tau_2) = \frac{1}{T-q} \sum_{t=1}^T h_t(\tau_1; \hat{\theta}_T^1) U h_t(\tau_2; \hat{\theta}_T^1)$$

where

$$U h_t(\tau; \hat{\theta}_T^1) = \omega(0) \overline{h_t(\tau; \hat{\theta}_T^1)} + \sum_{j=1}^T \omega\left(\frac{j}{S_T}\right) \left( \overline{h_{t-j}(\tau; \hat{\theta}_T^1)} + \overline{h_{t+j}(\tau; \hat{\theta}_T^1)} \right)$$

using the convention that  $h_t(\tau; \hat{\theta}_T^1) = 0$  if  $t \leq 0$  or  $t > T$ .

**Proposition 3.4** *Solving (9) is equivalent to solving*

$$\min_{\theta} \underline{w}'(\theta) \left[ \alpha_T I_T + C^2 \right]^{-1} \underline{v}(\theta) \quad (17)$$

where  $C$  is a  $T \times T$ -matrix with  $(t, l)$  element  $c_{tl}/(T-q)$ ,  $t, l = 1, \dots, T$ ,  $I_T$  is the  $T \times T$  identity matrix,  $\underline{v} = [v_1, \dots, v_T]'$  and  $\underline{w} = [w_1, \dots, w_T]'$  with

$$\begin{aligned} v_t(\theta) &= \int U h_t(\tau; \hat{\theta}_T^1) \hat{h}_T(\tau; \theta) \pi(\tau) d\tau, \\ w_t(\theta) &= \langle h_t(\tau; \hat{\theta}_T^1), \hat{h}_T(\tau; \theta) \rangle, \\ c_{tl} &= \int U h_t(\tau; \hat{\theta}_T^1) h_l(\tau; \hat{\theta}_T^1) \pi(\tau) d\tau. \end{aligned}$$

Note that in the case where the  $\{h_t\}$  are uncorrelated, the former formulas simplify:  $Uh_t = \bar{h}_t$ ,  $v_t = \bar{w}_t$ ,  $c_{tl} = \langle h_l(\tau; \hat{\theta}_T^1), h_t(\tau; \hat{\theta}_T^1) \rangle$ .

Similarly, an estimator of the asymptotic variance of  $\sqrt{T}(\hat{\theta}_T - \theta_0)$  given in (10) can be computed in a simple way.

**Proposition 3.5** *Suppose that the assumptions of Proposition 3.3 hold and  $T$ ,  $T^{\nu/(2\nu+1)}\alpha_T^{3/4}$  go to infinity and  $\alpha_T$  goes to zero. Then a consistent estimator of the  $q \times q$ -matrix  $\langle E^{\theta_0}(\nabla_{\theta}h), E^{\theta_0}(\nabla_{\theta}h) \rangle_K$  is given by*

$$\begin{aligned} & \langle \nabla_{\theta} \hat{h}_T(\hat{\theta}_T), (K_T^{\alpha_T})^{-1} \nabla_{\theta} \hat{h}_T(\hat{\theta}_T) \rangle \\ &= \frac{1}{(T-q)} \underline{w}'(\hat{\theta}_T) [\alpha_T I_T + C^2]^{-1} \underline{v}(\hat{\theta}_T) \end{aligned}$$

where  $C$  is the  $T \times T$ -matrix defined in Proposition 3.4,  $I_T$  is the  $T \times T$  identity matrix,  $\underline{v} = [v_1, \dots, v_T]'$  and  $\underline{w} = [w_1, \dots, w_T]'$  are  $T \times q$ -matrices with  $(t, j)$  element

$$\begin{aligned} (v_t(\theta))_j &= \int Uh_t(\tau; \hat{\theta}_T^1) \nabla_{\theta_j} \hat{h}_T(\tau; \theta) \pi(\tau) d\tau, \\ (w_t(\theta))_j &= \langle h_t(\tau; \hat{\theta}_T^1), \nabla_{\theta_j} \hat{h}_T(\tau; \theta) \rangle. \end{aligned}$$

### 3.4 Efficiency

In Proposition 3.2, we saw that the asymptotic variance of  $\hat{\theta}_T$  is  $\left( \langle E^{\theta_0}(\nabla_{\theta}h), E^{\theta_0}(\nabla_{\theta}h) \rangle_K \right)^{-1}$ . Using results on RKHS (see Carrasco and Florens, 2004, and references therein), it is possible to compute this term and hence to establish conditions under which this variance coincides with the Cramer Rao efficiency bound. We consider arbitrary functions  $h(\tau, Y_t; \theta_0)$  that satisfy the identification Assumption A.3 and where, as usual,  $Y_t$  is the  $(L+1)$ -vector of r.v.:  $Y_t = (X_t, X_{t+1}, \dots, X_{t+L})'$ . Let  $L^2(Y_t)$  be the set of random variables of the form  $g(Y_t)$  with  $E^{\theta_0}[|g(Y_t)|^2] < \infty$ . It is assumed that  $h(\tau, Y_t; \theta_0)$  belongs to  $L^2(Y_t)$ . Let  $S$  be the set of all random variables that may be written as  $\sum_{j=1}^n c_j h(\tau_j, Y_t; \theta_0)$  for arbitrary integer  $n$ , real constants  $c_1, c_2, \dots, c_n$  and points  $\tau_1, \dots, \tau_n$  of  $I$ . Denote  $\bar{S}$  its closure,  $\bar{S}$  contains all the

elements of  $S$  and their limits in  $L^2(Y_t)$ -norm.

**Proposition 3.6** *Assume that the results of Proposition 3.1 hold. Then,  $\hat{\theta}_T$  is asymptotically as efficient as MLE if and only if*

$$\nabla_{\theta} \ln f_{\theta}(x_{t+L}|x_{t+L-1}, \dots, x_t; \theta)|_{\theta=\theta_0} \in \bar{S}.$$

A proof of this proposition is given in Carrasco and Florens (2004). It states that the GMM estimator is efficient if and only if the score belongs to the span of the moment conditions. This result is close to that of Gallant and Long (1997) who show that if the auxiliary model is rich enough to encompass the DGP, then the efficient method of moments estimator is asymptotically efficient. It is important to remark that  $\pi$  does not affect the efficiency as long as  $\pi > 0$  on  $\mathbb{R}^d$ . In small samples however, the choice of  $\pi$  might play a role.

## 4 C-GMM based on the characteristic function

This section studies the properties of moment conditions (6) based on the conditional or joint characteristic function. The first subsection will focus on Markov processes while the second subsection will discuss mainly the nonmarkovian case.

### 4.1 Using the conditional characteristic function

Suppose an econometrician observes realizations of a Markov process  $X \in \mathbb{R}^p$ . The conditional characteristic function of  $X_{t+1}$ ,  $\psi_{\theta}(s|X_t; \theta)$ , defined in (1) is assumed to be known. We denote  $\psi_{\theta}(s|X_t; \theta_0)$  by  $\psi_{\theta}(s|X_t)$ . Let  $Y_t = (X_t, X_{t+1})'$ . The next proposition establishes that the GMM estimator based on a well-chosen double index (**DI**) moment function achieves ML efficiency.



**Proposition 4.1** *Consider*

$$h(\tau, Y_t; \theta) = e^{irX_t} \left( e^{isX_{t+1}} - \psi_\theta(s|X_t) \right), \quad (18)$$

with  $\tau = (r, s)' \in \mathbb{R}^{2p}$  and denote  $K$  the covariance operator of  $\{h(\cdot, Y_t; \theta)\}$ . Suppose that Assumptions A.2, A.3, A.7, and A.8 hold. Then the optimal GMM estimator based on (18) satisfies  $\hat{\theta}_T \xrightarrow{P} \theta_0$  and

$$\sqrt{T} (\hat{\theta}_T - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N} \left( 0, I_{\theta_0}^{-1} \right)$$

as  $T, T^{1/2}\alpha_T$  go to infinity and  $\alpha_T$  goes to zero.  $I_{\theta_0}$  denotes the Information matrix.

The efficiency resulting from moment functions (18) can be proved from Proposition 3.6. Indeed  $\bar{S}$  the closure of the span of  $\{h_t\}$  includes all functions in  $\mathbb{L}^2(Y_t)$  hence it also includes the score function. Alternatively, one can prove this result directly by computing the asymptotic variance of the GMM estimator and comparing it with the information matrix, see Equation (B.21) in Appendix.

The intuition for the efficiency result is as follows. For the GMM estimator to be as efficient as the MLE, the moment conditions need to be sufficiently rich to permit to recover the score. The **DI** moment functions with instruments defined in (4) span all functions in  $L^2(Y_t)$  and the unknown score in particular.

Notice that since the moment functions are uncorrelated and the optimal instrument is known to have an exponential form, the computation of the terms  $C$  and  $v$  in the objective function (17) is simplified and all elements involving the index  $r$  can be computed analytically. Therefore, using the **DI** instrument does not introduce computational complications. We outline these computations here. Let  $y_t = (x_{t+1}, x_t)$  and  $\hat{\pi}$  be the Fourier transform of  $\pi$  defined as

$$\hat{\pi}(x_t, x_{t+1}) = \int e^{i(rx_t + sx_{t+1})} \pi(\tau) d\tau. \quad (19)$$

Taking a product measure on  $r$  and  $s$ , we have

$$\pi(\tau) = \pi(r, s) = \pi_r(r) \pi_s(s). \quad (20)$$

If  $\pi$  is the p.d.f. of the bivariate normal variable  $y$  with zero mean and variance  $\Sigma$ , then  $\hat{\pi}(y) = \exp[-(y'\Sigma y/2)]$  where  $\Sigma$  is diagonal. Consider the moments of the type (18). An element of  $\underline{v}$  is computed as follows:

$$\begin{aligned}
v_t(\theta) &= \frac{1}{T} \sum_j \int \bar{h}(y_t, \tau; \hat{\theta}_T^1) h(y_j, \tau; \theta) \pi(\tau) d\tau \\
&= \frac{1}{T} \sum_j \int \overline{(e^{isx_{t+1}} - \psi_{\hat{\theta}_T^1}(s|x_t))} e^{irx_t} (e^{isx_{j+1}} - \psi_{\theta}(s|x_j)) e^{irx_j} \pi(\tau) d\tau \\
&= \frac{1}{T} \sum_j \int e^{is(x_{j+1}-x_{t+1})} e^{ir(x_j-x_t)} \pi(\tau) d\tau \\
&\quad - \frac{1}{T} \sum_j \int e^{i(sx_{j+1}+r(x_j-x_t))} \psi_{\hat{\theta}_T^1}(-s|x_t) \pi(\tau) d\tau \\
&\quad - \frac{1}{T} \sum_j \int e^{i(-sx_{t+1}+r(x_j-x_t))} \psi_{\theta}(s|x_j) \pi(\tau) d\tau \\
&\quad + \frac{1}{T} \sum_j \int \psi_{\hat{\theta}_T^1}(-s|x_t) \psi_{\theta}(s|x_j) e^{ir(x_j-x_t)} \pi(\tau) d\tau.
\end{aligned}$$

The first term is equal to  $1/T \sum_j \hat{\pi}(x_j - x_t, x_{j+1} - x_{t+1})$ . Given (20), the other terms involve:

$$I_r \equiv \int e^{ir(x_j-x_t)} \pi_r(r) dr = \hat{\pi}(x_j - x_t, 0). \quad (21)$$

Therefore, the second and third terms have the form

$$I_1 = I_r \cdot \int e^{-isv} \psi_{\theta}(s|w) \pi_s(s) ds \quad (22)$$

with opposite signs, and the last term is equal to

$$I_2 = I_r \cdot \int \psi_{\hat{\theta}_T^1}(-s|x_t) \psi_{\theta}(s|x_j) \pi_s(s) ds. \quad (23)$$

The remaining integrals, which have to be evaluated numerically, can be characterized as multidimensional integrals over infinite integration regions with a Gaussian weight function  $\pi$ . Evaluation of such integrals represents an important problem in the evaluation of quantum-mechanical matrix elements with gaussian wave functions in physics. Hence a plethora of fast and accurate numerical methods have been developed, see e.g. Genz and Keister (1996).

Note that integral  $I_1$  in (22) evaluated at  $(v, w) = (x_{t+1}, x_t)$  looks very similar to the Fourier inverse of the CF used in Singleton (2001, Equation (14)) to construct conditional density for MLE estimation. Presence of the density  $\pi$  turns out to be critical in the simplification of the numerical integration task. Figure 1 compares the integrand used in Singleton with  $I_1$  and  $I_2$ . It is clear that  $\pi$  dampens off all the oscillating behavior of the integrand needed for MLE.

The elements of the matrix  $C$  can be similarly computed by replacing  $\theta$  by  $\hat{\theta}_T^1$ .

#### 4.2 Using the joint characteristic function

Many important models in finance involve latent factors, the most prominent example being the stochastic volatility (SV) model. In this case, the full system can be described by a Markov vector  $(X_t, \mathcal{X}_t)'$  consisting of observable and latent components. As a result,  $X_t$  is most likely not Markov.<sup>10</sup>

For non-Markovian processes, the conditional characteristic function is usually unknown and difficult to estimate.<sup>11</sup> On the other hand, the joint characteristic function (JCF), if not known, can be computed by simulations.<sup>12</sup><sup>13</sup> Denote the JCF as:

$$\psi_{\theta}^L(\tau) = E^{\theta} \left( e^{i\tau'Y_t} \right) \quad (24)$$

where  $\tau = (\tau_0, \tau_1, \dots, \tau_L)'$ , and  $Y_t = (X_t, X_{t+1}, \dots, X_{t+L})'$ .

<sup>10</sup> Florens, Mouchard, and Rolin (1993) give necessary and sufficient conditions for the marginal of a jointly Markov process to be itself Markov.

<sup>11</sup> Bates (2003) provides an elegant way to compute conditional likelihood exploiting the fact that analytical form of affine CCF allows for filtering in the frequency domain. However, it appears that his method is limited to cases where  $\dim(X_t) = \dim(\mathcal{X}_t) = 1$  due to computational burdens.

<sup>12</sup> Jiang and Knight (2002) discuss examples of diffusion models for which JCF is available in analytical form. Yu (2001) derives JCF of the Merton model generalization to self-exciting jump component.

<sup>13</sup> Simulations are discussed in Section 5.

Feuerverger (1990) has considered this problem. His estimator is the solution to

$$\int \left( \psi_{\theta}^L(\tau) - \psi_T^L(\tau) \right) \varpi(\tau) d\tau = 0. \quad (25)$$

where  $\psi_T^L(\tau)$  denotes the empirical JCF. For a special weighting function  $\varpi$ , which is very similar to (2), Feuerverger shows that the estimator is as efficient as the estimator which solves

$$\frac{1}{T} \sum_{t=1}^T \nabla_{\theta} \ln f_{\theta}(X_{t+L} | X_{t+L-1}, \dots, X_t; \theta) = 0 \quad (26)$$

where  $f_{\theta}(X_{t+L} | X_{t+L-1}, \dots, X_t)$  is the true distribution of  $X_{t+L}$  conditional on  $X_t, \dots, X_{t+L-1}$ . This result holds even if the process  $X_t$  is not Markovian of order  $L$  (or less).

If  $X_t$  is Markovian of order  $L$  then the variance of the resulting estimator is  $I_{\theta}^{-1}(L)$  with

$$I_{\theta}(L) = E_{\theta} \left( \nabla_{\theta} \ln f_{\theta}(X_{t+L} | X_{t+L-1}, \dots, X_t; \theta)^2 \right) \quad (27)$$

which is the Cramér-Rao efficiency bound. If  $X_t$  is not Markovian of order  $L$  then the variance of the estimator has the usual sandwich form because  $\nabla_{\theta} \ln f_{\theta}(X_{t+L} | X_{t+L-1}, \dots, X_t; \theta_0)$  is not a martingale difference sequence with respect to  $\{X_{t+L}, \dots, X_1\}$ . This variance differs from  $I_{\theta}^{-1}(L)$  and is greater than the Cramér-Rao efficiency bound. Note that (26) should not be confused with quasi-maximum likelihood estimation because  $f_{\theta}(X_{t+L} | X_{t+L-1}, \dots, X_t; \theta)$ , is the exact distribution conditional on a restricted information set.

Feuerverger (1990) notes that the estimator based on the JCF can be made arbitrarily efficient provided that “ $L$  (fixed) is sufficiently large” although no proof is provided. This argument is clearly valid when the process is Markovian of order  $L$ . However, in the non-Markovian case, the only feasible way to achieve the efficiency would be to let  $L$  go to infinity with the sample size at a certain (slow) rate, the question of the optimal speed of convergence has to the best of our knowledge not been addressed in the literature. The implementation of such approach might be problematic since for  $L$  too large, the lack of data to estimate consistently the characteristic function might result in an  $\hat{\theta}_T$  with undesirable properties.

The approach of Feuerverger based on the joint characteristic function of basically the full

vector  $(X_1, X_2, \dots, X_T)$  is not realistic because only one observation of this vector is available. Instead, we can avoid using the unknown instrument  $A$  in (25) by considering a moment condition based on the JCF of  $Y_t$  :

$$h(\tau, Y_t, \theta) = e^{i\tau Y_t} - \psi_\theta^L(\tau). \quad (28)$$

for some small  $L = 0, 1, 2, \dots$ <sup>14</sup> Assume that the JCF is sufficient to identify the parameters. Now the moments  $h(\tau, Y_t, \theta_0)$  are not a martingale difference sequence (even if  $X_t$  is Markovian) and the kernel of  $K$  is given by

$$k(\tau_1, \tau_2) = \sum_{j=-\infty}^{\infty} E^{\theta_0} \left[ h(\tau_1, Y_t, \theta_0) \overline{h(\tau_2, Y_{t-j}, \theta_0)} \right].$$

When  $X_t$  is Markov of order  $L$ , the optimal GMM estimator is efficient as stated below.

**Proposition 4.2** *Assume that  $X_t$  is Markov of order  $L$  and that the assumptions of Proposition 3.3 hold and  $T, T^{\nu/(2\nu+1)}\alpha_T$  go to infinity and  $\alpha_T$  goes to zero. Then the optimal GMM estimator using the moments (28) is as efficient as the MLE.*

As the closure of the span of  $\{h_t\}$  contains the score  $\nabla_\theta \ln f_\theta(X_{t+L}|X_{t+L-1}, \dots, X_t; \theta_0)$ , the efficiency follows from Proposition 3.6.

Note that if  $X_t$  is Markov, it makes more sense to use moment conditions based on the CCF because the resulting estimator, while being efficient, is easier to implement (as  $\{h_t\}$  are m.d.s.). If  $X_t$  is not Markov, the GMM-JCF estimator will not be efficient. However, it might still have some good properties if the temporal dependence dies out quickly. As the computation of the optimal  $K_T$  may be burdensome (it involves two smoothing parameters  $S_T$  and  $\alpha_T$ ), one may decide to use a suboptimal weighting operator obtained by inverting the covariance operator without the autocorrelations.

One interesting question is then: What is the resulting loss of efficiency? We can answer this question only partially because we are not able to compute the variance of the optimal

<sup>14</sup> Jiang and Knight (2002), in a particular case of an affine stochastic volatility model, arbitrary base the instrument  $m$  on the normal density and experiment with values of  $L$  from 1 to 5.

GMM-JCF estimator when  $X_t$  is not Markov. However, we have a full characterization of the variance of the suboptimal GMM-JCF estimator.

Assume that one ignores the autocorrelations and uses as weighting operator the inverse of the operator  $\widetilde{K}$  associated with the kernel:

$$\widetilde{k}(\tau_1, \tau_2) = E^{\theta_0} \left[ h(\tau_1, Y_t, \theta_0) \overline{h(\tau_2, Y_t, \theta_0)} \right]. \quad (29)$$

**Proposition 4.3** *Assume that the assumptions of Proposition 4.2 hold. The asymptotic variance of the suboptimal GMM-JCF estimator  $\hat{\theta}_T$  using (28) and (29) is the same as that of the estimator  $\widetilde{\theta}_T$  which is the solution of*

$$\frac{1}{T} \sum_t \nabla_{\theta} \ln f_{\theta}(Y_t; \theta) = 0 \quad (30)$$

where  $\ln f_{\theta}(Y_t; \theta)$  is the exact joint distribution of  $Y_t$ .

Since using the efficient weighting matrix should result in a gain of efficiency, the asymptotic variance of  $\widetilde{\theta}_T$  (given in Appendix B) can be considered as an upper bound for the variance of the estimator obtained by using the optimal weighting operator that is  $K^{-1}$ . To illustrate the results of Proposition 4.3, consider first the case where  $\{X_t\}$  is i.i.d. and  $L = 1$ . Then solving (30) is basically (for  $T$  large) equivalent to solving

$$2 \frac{1}{T} \sum_t \nabla_{\theta} \ln f_{\theta}(X_t; \theta) = 0$$

so that the resulting estimator  $\hat{\theta}_T$  is efficient. Now turn to the case where  $\{X_t\}$  is Markov of order 1 and again  $L = 1$  then (30) is equivalent to

$$\frac{1}{T} \sum_t \nabla_{\theta} \ln f_{\theta}(X_t | X_{t-1}; \theta) + \frac{1}{T} \sum_t \nabla_{\theta} \ln f_{\theta}(X_{t-1}; \theta) = 0$$

which will not deliver an efficient estimator in general.

## 5 Case where the CF is unknown

As pointed out by Singleton (2001), the CF is not always available in closed form, especially if the model involves unobserved latent variables, like in the stochastic volatility model. To deal with this case, he suggests using the Simulated Method of Moments (SMM) along the line of Duffie and Singleton (1993). See also Gouriéroux and Monfort (1996), for a review on SMM. In this section, we consider two ways to estimate the CF via simulations depending on whether the observable variable is Markov or not.

- Assume that the observable random variable  $X_t$  is Markov and that it is possible to draw data from the conditional distribution of  $X_{t+1}$  given  $X_t$ . This simulation scheme is called conditional simulation. The conditional characteristic function is then estimated by an average over the simulated data.
- Assume now that the observable variable  $X_t$  is not Markov because of e.g. the presence of unobserved state variables in the model. In this case, it is usually impossible to draw in the conditional distribution. However, it may be possible to simulate a full sequence of random variables that have the same joint distribution as  $(X_1, \dots, X_T)$ . This simulation scheme is called path simulation. The joint characteristic function is then estimated using the simulated data.

The main difference in the properties of the two estimators is that in the first case, the estimator is as efficient as the MLE when the number of simulated paths,  $J$ , goes to infinity while in the second case, as  $X_t$  is not Markov, the estimator will never reach the efficiency bound even if  $J$  goes to infinity. A subsection will be devoted to each case.

### 5.1 Conditional simulation

In this subsection, we assume that  $X_t$  is a Markov process satisfying

$$X_{t+1} = H(X_t, \varepsilon_t, \theta) \tag{31}$$

where  $\varepsilon_t$  is an i.i.d. sequence independent of  $X_t$  with known distribution. For instance,  $X_t$  may be the solution of a dynamic asset pricing model as that presented by Duffie and Singleton (1993). If  $X_t$  is a discretely sampled diffusion process then  $H$  in (31) can be obtained from an Euler discretization.<sup>15</sup>

Moments based on the unknown CCF are used to estimate  $\theta$ . For a given  $\theta$  and conditionally on  $X_t$ , we generate a sequence  $\{\tilde{X}_{t+1|t}^{\theta,j}, j = 1, 2, \dots, J\}$  from

$$\tilde{X}_{t+1|t}^{\theta,j} = H(X_t, \tilde{\varepsilon}_{j,t+1}, \theta)$$

where  $\{\tilde{\varepsilon}_{j,t}\}_{j,t}$  are identically and independently distributed as  $\{\varepsilon_t\}$ . Note that  $\{\tilde{X}_{t+1|t}^{\theta,j}\}_j$  are i.i.d. conditionally on  $X_t$  and distributed as  $X_{t+1}|X_t$  when  $\theta = \theta_0$ . The moment conditions become

$$\tilde{h}_T^J(\tau; \theta) = \frac{1}{T} \sum_{t=1}^T e^{irX_t} \left( e^{isX_{t+1}} - \frac{1}{J} \sum_{j=1}^J e^{is\tilde{X}_{t+1|t}^{\theta,j}} \right)$$

where  $\tau = (r, s)$ . To facilitate the discussion, we introduce the following notations:

$$\begin{aligned} Y_t &= (X_t, X_{t+1})' \\ h_t(\tau; \theta) &= e^{irX_t} [e^{isX_{t+1}} - \psi_\theta(s|X_t)] \\ \tilde{h}(Y_t, \tilde{X}_{t+1|t}^{\theta,j}, \tau) &= e^{irX_t} [e^{isX_{t+1}} - e^{is\tilde{X}_{t+1|t}^{\theta,j}}] \\ \tilde{h}_t^J(\tau; \theta) &= \frac{1}{J} \sum_{j=1}^J \tilde{h}(Y_t, \tilde{X}_{t+1|t}^{\theta,j}, \tau) = e^{irX_t} \left( e^{isX_{t+1}} - \frac{1}{J} \sum_{j=1}^J e^{is\tilde{X}_{t+1|t}^{\theta,j}} \right) \end{aligned}$$

The resulting  $\{\tilde{h}_t^J(\tau; \theta_0)\}$  are martingale difference sequences with respect to  $\{X_t, X_{t-1}, \dots, X_1\}$

<sup>15</sup> However, there is a pitfall with this approach. When the number of discretization intervals per unit of time,  $N$ , is fixed, none of the  $J$  simulated paths,  $\tilde{X}_t^j$ , is distributed as  $X_t$  and the estimator  $\hat{\theta}_T$  is biased. Broze, Scaillet and Zakoian (1998) document the discretization bias of the Indirect Inference estimator and show that it vanishes when  $N \rightarrow \infty$  and  $J$  is fixed. In a recent paper, Detemple, Garcia, and Rindisbacher (2002) study estimators of the conditional expectation of diffusions. They show that if  $J$  is allowed to diverge too fast relative to  $N$ , then the bias of their estimator blows up. The same is likely to be true here. However as there is no limitation on how fine we can discretize (besides the computer precision), we assume that  $N$  is chosen sufficiently large for the discretization bias to vanish.



and therefore are uncorrelated. Moreover,

$$E^\theta \left[ \tilde{h} \left( Y_t, \tilde{X}_{t+1|t}^{\theta,j}, \tau \right) | Y_t \right] = h_t(\tau; \theta).$$

Let  $K$  be the covariance operator associated with the kernel

$$k(\tau_1, \tau_2) = E^{\theta_0} \left[ h_t(\tau_1; \theta_0) \overline{h_t(\tau_2; \theta_0)} \right] \quad (32)$$

and let  $U$  be the operator with kernel

$$u(\tau_1, \tau_2) = E^{\theta_0} \left[ \left( \tilde{h}_t^J - h_t \right) (\tau_1; \theta_0) \overline{\left( \tilde{h}_t^J - h_t \right) (\tau_2; \theta_0)} \right].$$

Denote  $\tilde{K}$  the covariance operator of  $\{\tilde{h}_t^J\}$ . Let  $\tilde{K}_T^{\alpha_T}$  be the regularized estimator of  $\tilde{K}$ . The GMM estimator associated with the moments  $\tilde{h}^J$  is defined as

$$\tilde{\theta}_T = \underset{\theta}{\operatorname{argmin}} \left\| \tilde{h}_T^J(\cdot, \theta) \right\|_{\tilde{K}_T^{\alpha_T}}^2.$$

Now we can state the efficiency result:

**Proposition 5.1** *Suppose that Assumptions A.2, A.3, A.7, A.8(i), A.9, and A.10 hold for  $\tilde{h}_t^J$  and a fixed  $J$ . We have  $\tilde{\theta}_T \xrightarrow{P} \theta_0$  and*

$$\sqrt{T} \left( \tilde{\theta}_T - \theta_0 \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left( 0, \left( \left\langle E^{\theta_0}(\nabla_\theta h), E^{\theta_0}(\nabla_\theta h)' \right\rangle_{\tilde{K}} \right)^{-1} \right)$$

as  $T$  and  $T^{1/2}\alpha_T$  go to infinity and  $\alpha_T$  goes to zero. Moreover,  $\tilde{K} = K + \frac{1}{J}U$  and we have the inequality

$$\left( \left\langle E^{\theta_0}(\nabla_\theta h), E^{\theta_0}(\nabla_\theta h)' \right\rangle_{\left(K + \frac{1}{J}U\right)} \right) \leq \left( \left\langle E^{\theta_0}(\nabla_\theta h), E^{\theta_0}(\nabla_\theta h)' \right\rangle_K \right). \quad (33)$$

For  $J$  large, the SMM estimator will be as efficient as the GMM-CCF estimator which itself

has been shown to reach the Cramér-Rao Efficiency bound because we have

$$\left\langle E^{\theta_0} (\nabla_{\theta} h), E^{\theta_0} (\nabla_{\theta} h)' \right\rangle_K = I_{\theta_0}.$$

## 5.2 Path simulation

Assume that one can generate a sequence of r.v.  $(\tilde{X}_1^{\theta}, \dots, \tilde{X}_{J(T)}^{\theta})$  such that the joint distribution of  $(\tilde{X}_1^{\theta}, \dots, \tilde{X}_{J(T)}^{\theta})$  given  $\theta$  and conditional on a starting value  $\tilde{X}_0 = x_0$  is the same as that of  $(X_1, \dots, X_{J(T)})$  given  $\theta$  and a starting value  $X_0 = x_0$ . This simulation scheme advocated by Duffie and Singleton (1993) is typically used when  $X_t$  is the marginal of a Markov process  $Z_t$ . For instance,  $Z_t = (X_t, X_t^*)$  where  $X_t^*$  is a latent variable, e.g. the volatility in a stochastic volatility model and  $X_t$  only is observable. In such cases, it is usually unknown how to draw from the conditional distribution of  $X_t$ . Moreover, even though the full system  $Z_t$  is Markov,  $X_t$  itself is usually not Markov. Therefore there is no hope to reach the Cramér-Rao efficiency bound using the JCF when  $L$  is fixed, as discussed in Section 4.2. We briefly explain how one can implement a path simulation. Assume, for instance, that  $Z_t$  is the solution of the recursion (31). For a given  $\theta$ , we generate a sequence  $\{Z_j^{\theta}, j = 1, 2, \dots, J(T)\}$  from

$$\begin{aligned} Z_{j+1}^{\theta} &= H(Z_j^{\theta}, \tilde{\varepsilon}_{j+1}, \theta) \\ Z_0^{\theta} &= z_0 \end{aligned} \tag{34}$$

where  $\{\tilde{\varepsilon}_j\}$  are identically and independently distributed as  $\{\varepsilon_t\}$ ,  $z_0$  is some arbitrary starting value, and the number of simulations  $J(T)$  goes to infinity with  $T$ . A simulator,  $\tilde{X}_j^{\theta}$ , of  $X$  is the first component of  $Z_j^{\theta}$ .

Contrary to the simulation scheme in the previous subsection, the sequence  $\{\tilde{X}_j^{\theta}\}$  is completely independent of the observations  $\{X_t\}$ . Note that, as the starting value  $x_0$  is not drawn from the stationary distribution of  $X_t$ , the sequence  $\{\tilde{X}_j^{\theta}\}$  is in general not stationary. We assume that  $X_t$  and consequently  $\{X_t^{\theta}\}$  are  $\beta$ -mixing with exponential decay, which guarantees that  $\tilde{X}_j^{\theta}$  becomes stationary exponentially fast. Hence the initial starting value

will not affect the distribution of our estimator.

The JCF of  $Y_t = (X_t, X_{t+1}, \dots, X_{t+L})'$ , as defined in (24), is assumed to be unknown and will be estimated via simulations. Let  $\tilde{Y}_j = (\tilde{X}_j^\theta, \dots, \tilde{X}_{j+L}^\theta)'$ . The estimation procedure is based on

$$\tilde{h}_T(\tau; \theta) = \frac{1}{T} \sum_{t=1}^T e^{i\tau Y_t} - \frac{1}{J(T)} \sum_{j=1}^{J(T)} e^{i\tau \tilde{Y}_j} \equiv \frac{1}{T} \sum_{t=1}^T \tilde{h}_t(\tau; \theta).$$

If  $\psi_\theta^L$  were known, the following moment conditions would be used

$$h_T(\tau; \theta) = \frac{1}{T} \sum_{t=1}^T e^{i\tau Y_t} - \psi_\theta^L(\tau) \equiv \frac{1}{T} \sum_{t=1}^T h_t(\tau; \theta)$$

Note that  $\{h_t(\tau; \theta)\}$  are not a martingale difference sequence and are autocorrelated. Therefore,  $K$ , the covariance operator associated with  $\{h_T(\tau; \theta)\}$ , has a more complicated expression than in the previous subsection:

$$k(\tau_1, \tau_2) = \sum_{i=-\infty}^{\infty} E^{\theta_0} \left[ \left( e^{i\tau_1' Y_t} - \psi_\theta^L(\tau_1) \right) \overline{\left( e^{i\tau_2' Y_{t-i}} - \psi_\theta^L(\tau_2) \right)} \right]$$

We estimate  $K$  using the kernel estimator  $K_T$  described in (15) and (16) where  $\psi_\theta^L(\tau_1)$  is estimated using the observations  $Y_t$ . Let  $K_T^{\alpha_T}$  be the regularized version of  $K_T$ . The GMM estimator associated with moments  $\tilde{h}$  is defined as

$$\tilde{\theta}_T = \underset{\theta}{\operatorname{argmin}} \left\| \tilde{h}_T(\cdot, \theta) \right\|_{K_T^{\alpha_T}}^2.$$

Note that  $\tilde{X}_{j+1}^\theta$  depends on  $\theta$  through the past history of  $\{\tilde{X}_j^\theta\}$ . Sufficient conditions for the uniform weak law of large numbers of  $\tilde{h}_T(\cdot, \theta)$  are discussed in Duffie and Singleton (1993). Let  $T/J(T)$  converge to  $\zeta$  as  $T$  goes to infinity. Then, under the additional mixing property of  $X_t$  (Assumption A.11), we have the following result:

**Proposition 5.2** *Suppose that Assumptions A.2 to A.6(i) (for  $\tilde{h}_t$  replacing  $h_t$  and  $E^{\theta_0}$  denotes the expectation with respect to the stationary distribution of  $Y_t$ ), and A.11 hold. Let  $K_T$  be the kernel estimator of  $K$  with kernel  $\omega$  and bandwidth  $S_T$  satisfying the conditions*

of Proposition 3.3(ii). Then,  $\tilde{\theta}_T \xrightarrow{P} \theta_0$  and

$$\sqrt{T} (\tilde{\theta}_T - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N} \left( 0, (1 + \zeta) \left( \langle E^{\theta_0} (\nabla_{\theta} h), E^{\theta_0} (\nabla_{\theta} h)' \rangle_K \right)^{-1} \right)$$

as  $T$  and  $T^{\nu/(2\nu+1)} \alpha_T$  go to infinity and  $\alpha_T$  goes to zero.

It should be noted that the variance of  $\tilde{\theta}_T$  can be made as close as possible to that of  $\hat{\theta}_T$  in Proposition 3.2 by letting  $T/J(T)$  go to 0. Because of the autocorrelations, the estimation of the optimal weighting operator  $K$  is burdensome. To simplify this computation we could use the covariance operator that ignores the autocorrelations but the resulting estimator would be less efficient. Its variance is given by Proposition 4.3 for the non-simulated case. The variance of the C-SMM estimator is again equal to  $(1 + \zeta)$  times the variance obtained in the non-simulated context.

## 6 Monte-Carlo Study

In this section we evaluate the performance of the CF-based estimators via Monte-Carlo analysis. For this purpose we consider an example of the CIR, or square-root, interest rate model from financial economics. The conditional CF is available in closed-form for this model. We compare the performance of two CF-based estimators – one is using the **SI** instrument, and the other is using the **DI** instrument – with that of MLE, QMLE, EMM.<sup>16</sup>

The CIR square-root process

$$dr_t = (\gamma - \kappa r_t) dt + \sigma \sqrt{r_t} dW_t \quad (35)$$

has the following conditional characteristic function (see e.g. Singleton, 2001):

---

<sup>16</sup> We are grateful to Ken Singleton for providing his code for the approximately efficient GMM-CCF estimator based on the **SI** instrument.

$$\psi(\tau|r_t) = \left(1 - \frac{i\tau}{c}\right)^{-2\gamma/\sigma^2} \exp\left\{\frac{i\tau e^{-\kappa}}{1 - \frac{i\tau}{c}} r_t\right\} \quad (36)$$

$$c = \frac{2\kappa}{\sigma^2(1 - e^{-\kappa})}$$

We assume that  $\kappa$ ,  $\gamma$  and  $\sigma$  are all strictly positive and  $\sigma^2 \leq 2\gamma$ . Under these conditions, the square root process is known to have a unique fundamental solution and its marginal density is *Gamma* and its transition density is a type I Bessel function distribution or noncentral  $\chi^2$  with a fractional order (see e.g. Cox et al., 1985). The following lemma, proved in the appendix B, guarantees that the assumptions needed for the consistency and asymptotic normality (see proposition 4.1) of the C-GMM estimator hold.

**Lemma 6.1** (1) *The process  $\{r_t\}$  is  $\beta$ -mixing with geometric decay and therefore is  $\alpha$ -mixing with geometric decay.* (2) *Assumptions A7 and A8 are satisfied.*

The simulation design is identical to Zhou (2001). This is done on purpose as it allows us to compare our results with the MLE, QMLE and EMM results reported in Zhou (2001). We consider a sample size  $T = 500$  with a weekly sampling frequency in mind. The parameter estimates are obtained from Gallant and Tauchen (1998).

For the CF-based estimator based on the continuum of moment conditions, which we term C-GMM-**DI**, we have experimented with different values of the penalization term  $\alpha_T = 0.05, 0.1, 0.2$ , and different volatility values of the Gaussian integrating density  $\pi(\tau)$  : we tried the values of 1 (standard normal) and the inverse of the standard deviation of the data. The standard normal density produces slightly better results.

For the Singleton's approximately efficient estimator, which we term GMM-**SI**, we have experimented with different values of  $(M, \delta)$  which control the number of grid points and the distance between the grid points respectively. We have considered the following pairs  $(3, 0.5)$ ,  $(6, 1)$ ,  $(6, 0.75)$ ,  $(9, 1)$ . We found the combination  $(6, 1)$  to be the most successful as coarser grid did not contain enough information about the distribution, and the finer grid led to too many moment conditions generating an ill-behaved weighting matrix.

For brevity, Table 1 reports the C-GMM-**DI** results only for  $\alpha_T = 0.02$ , and for the standard normal integrating density  $\pi(\tau)$ ; and the GMM-**SI** results only for  $(M, \delta) = (6, 1)$ . We computed the results for 1000 Monte-Carlo paths.<sup>17</sup> Results for other configurations are available upon request. We report the Mean Bias, Median Bias and Root Mean Squared Error of the following estimators: MLE, QMLE, EMM, GMM-**SI**, and C-GMM-**DI**. The first three estimators appeared in Zhou (2001) and we report his results only for the purpose of comparison.

In terms of bias, the performance of C-GMM-**DI** and GMM-**SI** for  $\gamma$  and  $\kappa$  is comparable to MLE and vastly better than QMLE and EMM. However, performance of both CF-based estimators is worse for  $\sigma$ . In terms of RMSE, MLE's efficiency is triple of that of C-GMM-**DI** for  $\gamma$  and  $\kappa$ . Both CF-based estimators dominates the other two methods by far, but underperform for  $\sigma$  again.

If we compare the two CF-based estimators to each other, the C-GMM-**DI** fares better. The key improvement of this estimator over GMM-**SI** is that the distribution of the estimated parameters is far less skewed and leptokurtic. For example, in the case of the parameter  $\kappa$ , the skewness and excess kurtosis are 2 and 5, respectively, while in the case of the GMM-**SI** estimator the numbers are -5 and 38.

We illustrate the differences in the distributions by plotting histograms of the square root of the sum of squared errors across all three parameters,  $\sqrt{(\gamma - \hat{\gamma}_i)^2 + (\kappa - \hat{\kappa}_i)^2 + (\sigma - \hat{\sigma}_i)^2}$ , computed along each simulated path  $i$  for both methods in Figure 2. The rationale for combining errors across parameters is that one method could be claimed more efficient than the other only if the overall error is smaller. We observe the effect similar to the one noted with regard to the parameter  $\kappa$ . The GMM-**SI** estimator tends to produce more extreme errors than the C-GMM-**DI** one does.

---

<sup>17</sup> The GMM-**SI** method did not converge in seven cases, i.e., all the results are based on 993 paths.

## 7 Conclusion

This paper showed how to construct maximum likelihood efficient estimators in settings where the maximum likelihood estimation itself is not feasible. The solution is to use GMM and to select moment functions, which are based on the characteristic function, and optimal instruments, which form a basis spanning the unknown likelihood score. The efficiency is achieved by using the whole continuum of possible moment conditions resulting from this approach. We provide practical results allowing to construct such an estimator as well as auxiliary results pertaining to the cases when data are not Markov (estimation based on the joint characteristic function) and when characteristic functions are not available in analytical form (simulated method of moments estimation). Our Monte-Carlo study shows that the method indeed performs on par with MLE, and fares better than other methods.

The methodology is applicable to estimation of a wide range of non-linear time series models. It has particular relevance for empirical work in finance. Asset pricing models are frequently formulated in terms of stochastic differential equations, which have no closed form solution for the conditional density based on discrete-time observations. Motivated by these avenues of application, the future work will have to refine our results on estimation of non-Markovian processes and latent states as well as develop tests in the framework of characteristic function based continuum of moment conditions.

## References

- Altissimo, F. and A. Mele, 2005. Simulated Nonparametric Estimation of Dynamic Models with Applications to Finance. Working paper, London School of Economics.
- Andrews, D., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59, 817-858.
- Bates. D., 2003. Maximum Likelihood estimation of latent affine processes. Forthcoming *Review of Financial Studies*.
- Broze, L., O. Scaillet, and J-M. Zakoian, 1998. Quasi-indirect inference for diffusion processes. *Econometric Theory*, 14, 161-186.
- Carrasco, M. and J. P. Florens, 2000. Generalization of GMM to a continuum of moment conditions. *Econometric Theory*, 16, 797-834.
- Carrasco, M. and J. P. Florens, 2002. Efficient GMM estimation using the empirical characteristic function. Working paper, CREST, Paris.
- Carrasco, M. and J. P. Florens, 2004. On the asymptotic efficiency of GMM. Working paper, University of Rochester.
- Carrasco, M., J. P. Florens, and E. Renault, 2003. Linear inverse problems in structural econometrics: estimation based on spectral decomposition and regularization. Forthcoming in the *Handbook of Econometrics*, Vol. 6, edited by J.J. Heckman and E.E. Leamer.
- Carrasco, M., L. P. Hansen, and X. Chen, 1999. Time Deformation and Dependence. Working paper, University of Chicago.
- Chacko, G. and L. Viceira, 2005. Spectral GMM estimation of continuous-time processes. *Journal of Econometrics*, 116, 259-292.
- Chen, X., L. P. Hansen, and M. Carrasco, 1999. Nonlinearity and Temporal Dependence. Working paper, University of Chicago.
- Chen, X., L. P. Hansen, and J. Scheinkman, 1998. Shape Preserving Estimation of Diffusions. Working paper, University of Chicago.
- Chen, X. and H. White, 1998. Central Limit and Functional Central Limit Theorems for Hilbert space-valued dependent processes. *Econometric Theory*, 14, 260-284.
- Conley, T., L. P. Hansen, E., Luttmmer, and J. Scheinkman, 1997. Short term interest rates



- as subordinated diffusions. *Review of Financial Studies*, 10, 525-578.
- Cox, J.C., J. Ingersoll and S.A. Ross, 1985. A theory of the term structure of interest rates. *Econometrica*, 53, 385-408.
- Darolles, S., J.P. Florens, and E. Renault, 2002. Nonparametric instrumental regression. Working paper 05-2002, CRDE.
- Detemple, J., R. Garcia, and M. Rindisbacher, 2002. Asymptotic Properties of Monte Carlo Estimators of Diffusion Processes. Working paper, University of Montreal.
- Donald, S. and W. Newey, 2001. Choosing the number of instruments. *Econometrica*, 69, 1161-1191.
- Duffie, D. and K. Singleton, 1993. Simulated moments estimation of Markov models of asset prices. *Econometrica*, 61, 929-952.
- Dunford, N. and J. Schwartz, 1988. *Linear operators, Part II: Spectral Theory*. Wiley, New York.
- Fermanian, J-D. and B. Salanie, 2004. A nonparametric simulated maximum likelihood estimation Method. *Econometric Theory*, 20, 701-734.
- Feuerverger, A., 1990. An efficiency result for the empirical characteristic function in stationary time-series models. *Canadian Journal of Statistics*, 18, 155-161.
- Feuerverger, A. and P. McDunnough, 1981. On the Efficiency of Empirical Characteristic Function Procedures. *J. R. Statist. Soc. B*, 43, 20-27.
- Florens, J.-P., Mouchard, and Rolin, 1993. Noncausality and Marginalization of Markov Processes. *Econometric Theory*, 9, 239-260.
- Gallant, A.R. and J. R. Long, 1997. Estimating Stochastic Differential Equations Efficiently by Minimum Chi-Square. *Biometrika*, 84, 125-141.
- Gallant, A.R. and G. Tauchen, 1996. Which Moments to Match? *Econometric Theory*, 12, 657-681.
- Gallant, A.R. and G. Tauchen, 1998. Reprojecting partially observed systems with application to interest rate diffusions. *Journal of American Statistical Association*, 93, 10-24.
- Genz, A. and B.D. Keister, 1996. Fully symmetric interpolatory rules for multiple integrals over infinite regions with gaussian weight. *Journal of Computational and Applied Mathematics*, 71, 299-309.

- Gouriéroux, C., A. Monfort, 1996. Simulation based econometric methods. CORE Lectures, Oxford University Press, New York.
- Gouriéroux, C., A. Monfort and E. Renault, 1993. Indirect Inference. *Journal of Applied Econometrics*, 8, S85-S118.
- Hall, P. and J. Horowitz, 1996. Bootstrap critical values for tests based on Generalized-Method-of-Moments estimators. *Econometrica*, 64, 891-916.
- Hansen, L., 1982. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, 50, 1029-1054.
- Hansen, L., 1985. A method of calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators. *Journal of Econometrics*, 30, 203-238.
- Hansen, L. and J. Scheinkman, 1995. Back to the future: Generating moment implications for continuous-time Markov processes. *Econometrica*, 63, 767-804.
- Horn, R. and C. Johnson, 1985. *Matrix Analysis*. Cambridge University Press, Cambridge, UK.
- Jiang, G. and J. Knight, 2002. Estimation of Continuous Time Processes via the Empirical Characteristic Function. *Journal of Business and Economic Statistics*, 20, 198-212.
- Parzen, E., 1957. On consistent estimates of the spectrum of a stationary time series. *Annals of Mathematical Statistics*, 28, 329-348.
- Parzen, E., 1970. Statistical inference on time series by RKHS methods. 12th Biennial Seminar Canadian Mathematical Congress Proc., R. Pyke, ed., Canadian Mathematical Society, Montreal.
- Politis, D. and J. Romano, 1994. Limit theorems for weakly dependent Hilbert space valued random variables with application to the stationary bootstrap. *Statistica Sinica*, 4, 451-476.
- Singleton, K., 2001. Estimation of affine pricing models using the empirical characteristic function. *Journal of Econometrics*, 102, 111-141.
- van der Vaart, A., 1998. *Asymptotic Statistics*. Cambridge University Press, Cambridge, UK.
- White, H., 1994. *Estimation, inference and specification analysis*. Cambridge University Press, Cambridge, UK.

Yu, J., 2001. Empirical characteristic function estimation and its applications. Working paper, University of Auckland.

Zhou, H., 2001. Finite sample properties fo EMM, GMM, QMLE, and MLE for a square-root interest rate diffusion model. *Journal of Computational Finance*, 5, 89-122.

## A Regularity Conditions

**Assumption A.1** *The stochastic process  $X_t$  is a  $p \times 1$ -vector of random variables.  $X_t$  is stationary and  $\alpha$ -mixing with coefficients  $\alpha_j$  that satisfy  $\sum_{j=1}^{\infty} j^2 \alpha_j < \infty$ . The distribution of  $(X_1, X_2, X_3, \dots)$  is indexed by a finite dimensional parameter  $\theta \in \Theta \subset \mathbb{R}^q$  and  $\Theta$  is compact.*

The condition on the mixing numbers is satisfied if  $X_t$  is  $\alpha$ -mixing of size -3.<sup>18</sup> Sufficient conditions for  $\rho$ - and  $\beta$ - mixing (and, therefore,  $\alpha$ -mixing) of univariate diffusions can be found in Chen et al. (1999). For subordinated diffusions, they can be found in Carrasco et al. (1999) with many examples. The condition in Assumption A.1 is relatively weak and is expected to be satisfied for a large class of processes. The following assumption introduces the Hilbert space of reference.

**Assumption A.2**  *$\pi$  is the p.d.f. of a distribution that is absolutely continuous with respect to Lebesgue measure on  $\mathbb{R}^d$  and admits all its moments.  $\pi(\tau) > 0$  for all  $\tau \in \mathbb{R}^d$ .  $\mathbb{L}^2(\pi)$  is the Hilbert space of complex-valued functions that are square integrable with respect to  $\pi$ :*

$$\mathbb{L}^2(\pi) = \left\{ g : \mathbb{R}^d \rightarrow \mathbb{C} \mid \int |g(\tau)|^2 \pi(\tau) d\tau < \infty \right\}$$

Denote  $\langle \cdot, \cdot \rangle$  and  $\| \cdot \|$  the inner product and the norm defined on  $\mathbb{L}^2(\pi)$ . The inner product is  $\langle f, g \rangle = \int f(\tau) \overline{g(\tau)} \pi(\tau) d\tau$  where  $\overline{g(\tau)}$  denotes the complex conjugate of  $g(\tau)$ . If  $f = (f_1, \dots, f_m)'$  and  $g = (g_1, \dots, g_m)'$  are vectors of functions of  $\mathbb{L}^2(\pi)$ , we denote  $\langle f, g' \rangle$  the  $m \times m$ -matrix with  $(i, j)$  element  $\int f_i(\tau) \overline{g_j(\tau)} \pi(\tau) d\tau$ .

We also have to define a Hilbert-space analog of a random variable:

**Definition A.1** *An  $\mathbb{L}^2(\pi)$ -valued random element  $g$  has a Gaussian distribution on  $\mathbb{L}^2(\pi)$  with covariance operator  $K$  if, for all  $f \in \mathbb{L}^2(\pi)$ , the real-valued random variable  $\langle g, f \rangle$  has a Gaussian distribution on  $\mathbb{R}$  with variance  $\langle Kf, f \rangle$ .<sup>19</sup>*

<sup>18</sup> Note that a size -2 (instead of -3) is sufficient to establish the asymptotic normality of the estimator (Proposition 3.2). However we need a stronger condition (weaker dependency structure) to show the consistency of covariance operator estimate,  $K_T$  (Proposition 3.3).

<sup>19</sup> Background material on the Hilbert space - valued random elements can be found in, for instance,

We assume that the moment conditions (6) identify all the parameters of interest:

**Assumption A.3** *The equation*

$$E^{\theta_0}(h_t(\tau; \theta)) = 0 \text{ for all } \tau \in \mathbb{R}^d, \pi - \text{almost everywhere}$$

has a unique solution  $\theta_0$  which is an interior point of  $\Theta$ .  $E^{\theta_0}$  denotes the expectation with respect to the distribution of  $Y_t$  for  $\theta = \theta_0$ .

$\{h_t(\theta_0)\}$  is supposed to satisfy the set of assumptions:

**Assumption A.4** (i)  $h$  is a measurable function from  $\mathbb{R}^d \times \mathbb{R}^{\dim(Y)} \times \Theta$  into  $\mathbb{C}$ .

(ii)  $h_t(\tau; \theta)$  is continuously differentiable with respect to  $\theta$  and  $h_t(\tau; \theta) \in L_\infty(\pi \otimes P^\theta)$  where  $L_\infty(\pi \otimes P^\theta)$  is the set of measurable bounded functions of  $(\tau, Y_t)$ .

$$(iii) \sup_{\theta \in \Theta} \left\| \hat{h}_T(\theta) - E^{\theta_0} h_t(\theta) \right\| = O_p\left(\frac{1}{\sqrt{T}}\right)$$

$\sup_{\theta \in \mathfrak{N}} \left\| \nabla_\theta \hat{h}_T(\theta) - E^{\theta_0} \nabla_\theta h_t(\theta) \right\| = O_p\left(\frac{1}{\sqrt{T}}\right)$ , where  $\nabla_\theta$  denotes the derivative with respect to  $\theta$  and  $\mathfrak{N}$  is some neighborhood about  $\theta_0$ .

Note that we do not try to provide minimal assumptions and that A.4(ii) could certainly be relaxed. However, as our moment conditions are based on the conditional CF and on the joint CF, they will be necessarily bounded. Note that when  $h_t$  is based on the JCF, then  $\hat{h}_T(\theta) - E^{\theta_0} h_t(\theta)$  does not depend on  $\theta$  and  $\nabla_\theta \hat{h}_T(\theta) - E^{\theta_0} \nabla_\theta h_t(\theta)$  is identically zero. So that A.4(iii) is easy to check. On the other hand, when  $h_t$  is based on the CCF, the verification is less straightforward and will be undertaken in Proposition A.1 below.

The following assumption about the moment function  $h_t$  is required for establishing the properties of the optimal C-GMM estimator. We require the null space of  $K$  be reduced to zero for the following reason. If  $\mathcal{N}(K)$  is different from  $\{0\}$ , then 0 is an eigenvalue of  $K$  and the solution (in  $f$ ) to the equation  $Kf = g$  is not unique, hence  $K^{-1}$  and  $K^{-1/2}$  are not uniquely defined. It would be possible to define  $K^{-1}$  as the generalized inverse of  $K$ , that is

Chen and White (1998).

$K^{-1}$  would have for spectrum  $\{1/\lambda_j\}$  where  $\lambda_j$  are the nonzero eigenvalues of  $K$ . However in that case, the null space of  $K^{-1/2}$  (defined as  $(K^{-1})^{1/2}$ ) coincides with the null space of  $K$  and hence is not empty, as a result  $\theta$  is not identified. Indeed for  $\theta$  to be identified, we need

$$\|K^{-1/2}E^{\theta_0}h_t(\theta)\| = 0 \Rightarrow E^{\theta_0}h_t(\theta) = 0 \Rightarrow \theta = \theta_0,$$

which is true if  $\mathcal{N}(K^{-1/2}) = \{0\}$  and A3 holds.

**Assumption A.5** Let  $K$  be the asymptotic covariance operator of  $\sqrt{T}\hat{h}_T(\theta_0)$ . (i) The null space of  $K : \mathcal{N}(K) = \{f \in \mathbb{L}^2(\pi) : Kf = 0\} = \{0\}$ . (ii)  $E^{\theta_0}h_t(\theta) \in \mathcal{H}(K)$  for all  $\theta \in \Theta$  and (iii)  $E^{\theta_0}\nabla_{\theta}h_t(\theta) \in \mathcal{H}(K)$  for all  $\theta \in \mathfrak{N}$ .

The following conditions are used to establish the properties of the covariance estimator.

**Assumption A.6** (i) The kernel  $\omega$  satisfies  $\omega : \mathbb{R} \rightarrow [-1, 1]$ ,  $\omega(0) = 1$ ,  $\omega(x) = \omega(-x)$ ,  $x \in \mathbb{R}$ ,  $\int \omega^2(x) dx < \infty$ ,  $\int |\omega(x)| dx < \infty$ .  $\omega$  is continuous at 0 and at all but a finite number of points.

(ii)  $E^{\theta_0} \sup_{\theta \in \mathfrak{N}} \|\nabla_{\theta}h_t(\cdot; \theta)\| < \infty$ .

The following assumption is needed in Section 4.1. to use the conditional characteristic function in the markovian case.

**Assumption A.7** The stochastic process  $X_t$  is a  $p \times 1$ -vector of random variables.  $X_t$  is stationary, Markov, and  $\alpha$ -mixing with  $\sum_{j=1}^{\infty} j^2 \alpha_j < \infty$ . The conditional pdf of  $X_{t+1}$  given  $X_t$ ,  $f_{\theta}(x_{t+1}|x_t; \theta)$ , is indexed by a parameter  $\theta \in \Theta \subset R^q$  and  $\Theta$  is compact.  $f_{\theta}(x_{t+1}|x_t; \theta)$  is continuously differentiable w.r. to  $\theta$ . For brevity,  $f_{\theta}(x_{t+1}|x_t; \theta_0)$  and  $f_{\theta}(x_{t+1}|x_t; \theta)$  are denoted respectively  $f_{\theta_0}(x_{t+1}|x_t)$  and  $f_{\theta}(x_{t+1}|x_t)$ .

Now, we elaborate on the conditions to implement the efficient C-GMM estimator. Some of the assumptions, e.g. Assumption A.5, might seem to be difficult to verify. We can check these conditions using the properties of the RKHS. Below, we give a set of primitive assumptions under which the general Assumptions A.1, A.5, A.4 to A.6(ii) are satisfied.

**Assumption A.8** (i) For all  $\theta \in \Theta$ , the following inequality holds

$$E^{\theta_0} \left[ \left( 1 - \frac{f_\theta(r_t|r_{t-1})}{f_{\theta_0}(r_t|r_{t-1})} \right)^2 \right] < \infty$$

and there exists a neighborhood  $\aleph$  about  $\theta_0$  such that

$$E^{\theta_0} \left[ \frac{\nabla_\theta f_\theta(x_{t+1}|x_t) \nabla_\theta f_\theta(x_{t+1}|x_t)'}{f_{\theta_0}(x_{t+1}|x_t)^2} \right] < \infty$$

for all  $\theta \in \aleph$ .  $\psi_\theta$  is differentiable and  $\int \sup_{\theta \in \Theta} |\nabla_\theta \psi_\theta(s|x_t)| ds < \infty$ .

(ii)  $\psi_\theta(s|X_t; \theta)$  is twice continuously differentiable in  $\theta$ .  $E^{\theta_0} \|\nabla_\theta \psi_\theta(\cdot|X_t; \theta)\|^{2+\delta} < \infty$  for some  $\delta > 0$  and  $\sum_{j=1}^{\infty} \alpha_j^{\delta/(2+\delta)} < \infty$ .

(iii)  $E^{\theta_0} [(\sup_{\theta \in \Theta} \|\nabla_\theta \psi_\theta(\cdot|X_t; \theta)\|)^2] < \infty$  and  $E^{\theta_0} [(\sup_{\theta \in \aleph} \|\nabla_{\theta\theta} \psi_\theta(\cdot|X_t; \theta)\|)^2] < \infty$  where  $\nabla_{\theta\theta} \psi_\theta$  denotes the  $q \times q$  matrix of second derivatives of  $\psi_\theta$ .

Note that the first inequality in A8(i) (which corresponds to A5(ii)) may impose some restrictions on  $\Theta$  as illustrated in Section F3 of the unpublished appendix of Altissimo and Mele (2005). In Lemma 2, we verify that Assumption A7 and A8 are satisfied for the CIR model.

**Proposition A.1** Assumption A.7 implies Assumption A.1. If Assumption A.7 is satisfied and  $h_t$  is defined by

$$h(\tau, Y_t; \theta) = e^{irX_t} \left( e^{isX_{t+1}} - \psi_\theta(s|X_t) \right), \quad (\text{A.1})$$

with  $\tau = (r, s)' \in \mathbb{R}^{2p}$  then Assumption A.8 implies Assumptions A.4, A.5, and A.6(ii).

The proof is provided in Appendix B.

In the section on the simulated method of moments, our starting point is the following model.

**Assumption A.9**  $X_t$  satisfies

$$X_{t+1} = H(X_t, \varepsilon_t, \theta)$$

for some measurable transition function  $H : R^p \times R^N \times \Theta$  for some  $N > 0$ .  $\varepsilon_t$  is an i.i.d. sequence of  $R^N$ -valued random variables independent of  $X_t$  with a known distribution that does not depend on  $\theta$ .

We will need an assumption, which corresponds to Assumption A.8(ii) and (iii) for the particular moments  $\tilde{h}_t^J$  used in the conditional simulation case of the simulated method of moments (section 5.1).

**Assumption A.10** (i)  $H$  is twice continuously differentiable in  $\theta$ .

(ii)  $E^{\theta_0} E_\varepsilon \|\nabla_\theta H(X_t, \varepsilon_t, \theta)\|_E^{2+\delta} < \infty$  for some  $\delta > 0$  and  $\sum_{j=1}^\infty \alpha_j^{\delta/(2+\delta)} < \infty$ .  $E_\varepsilon$  denotes the expectation with respect to the distribution of  $\varepsilon_t$ .

(iii)  $E^{\theta_0} E_\varepsilon [(\sup_{\theta \in \Theta} \|\nabla_\theta H(X_t, \varepsilon_t, \theta)\|_E)^2] < \infty$  and  $E^{\theta_0} E_\varepsilon [(\sup_{\theta \in \mathbb{R}} \|\nabla_{\theta\theta} H(X_t, \varepsilon_t, \theta)\|_E)^2] < \infty$ .

The following assumption is required for the proof of asymptotic properties of the simulated estimator in case of path simulation (section 5.2).

**Assumption A.11**  $X_t$  is  $\beta$ -mixing with exponential decay and  $E^{\theta_0} \sup_{\theta \in \Theta} [\nabla_\theta e^{i\tau \tilde{Y}_j}] < \infty$ .  $E^{\theta_0}$  denotes the expectation with respect to the stationary distribution of  $Y_t$ .

## B Proofs

The following lemma is used in the proof of Propositions A.1 and 4.3. It gives an expression for the inner product/norm in a RKHS,  $\mathcal{H}(K)$ , that appears in Parzen (1970) and is further discussed in Carrasco and Florens (2004).

Let

$$C_i = \left\{ G : g_i(\tau) = E^{\theta_0}(h(\tau)G) \quad \forall \tau \in R^d \right\}$$

**Lemma B.1** Let  $K$  be a covariance operator from  $L^2(\pi)$  into  $L^2(\pi)$  with kernel  $k(\tau_1, \tau_2) =$



$E^{\theta_0} \left( h_t(\tau_1) \overline{h_t(\tau_2)} \right)$ . Let  $g$  be a  $L$  vector of elements of  $\mathcal{H}(K)$ . Then  $\Sigma = \langle g, g' \rangle_K$  is the “smallest”  $L \times L$  matrix with  $(i, j)$  element

$$\langle g_i, g_j \rangle_K = E^{\theta_0} (G_i G_j)$$

such that  $G_i \in C_i$ ,  $i = 1, 2, \dots, L$ . That is, for any  $\tilde{G}_i \in C_i$ , the matrix  $\tilde{\Sigma}$  with elements  $E^{\theta_0} (\tilde{G}_i \tilde{G}_j)$  satisfies the property that  $\tilde{\Sigma} - \Sigma$  is nonnegative definite.

**Proof of Lemma 3.1.** To prove this result, we need a functional central limit theorem for weakly dependent process. We use the results of Politis and Romano (1994). By Assumptions A.1 and A.4(i),  $\{h_t\}$  is stationary  $\alpha$ -mixing with  $\sum_{j=1}^{\infty} j^2 \alpha_j < \infty$ . Moreover by Assumption A.4(ii),  $\{h_t\}$  is bounded with probability one. The result follows directly from Theorem 2.2 of Politis and Romano (1994). Note that Politis and Romano require that the  $\alpha$  coefficient of  $\{h_t\}$  satisfies  $\sum_{i=1}^j i^2 \alpha(i) \leq K j^\mu$  for all  $1 \leq j \leq T$  and some  $\mu < 3/2$  which is satisfied.

Note that  $K$  is an integral operator with kernel  $k$  defined in Equation (12). An operator  $K : \mathbb{L}^2(\pi) \rightarrow \mathbb{L}^2(\pi)$  with kernel  $k$  is an operator of Hilbert Schmidt if

$$\int \int |k(\tau_1, \tau_2)|^2 \pi(\tau_1) \pi(\tau_2) d\tau_1 d\tau_2 < \infty.$$

As  $\pi$  is a pdf, it is enough to show that  $k(\tau_1, \tau_2) < \infty$ . As  $k(\tau_1, \tau_2)$  is the long-run covariance of  $\{h_t\}$ , it is well-known (see e.g. Politis and Romano, 1994, Remark 2.2) that a sufficient condition for  $k$  to be finite is that  $\{h_t\}$  is bounded with probability one and the  $\alpha$ -coefficients of  $\{h_t\}$  are summable i.e.  $\sum_j \alpha(j) < \infty$ . These two conditions are satisfied under our assumptions. Hence  $K$  is a Hilbert-Schmidt operator.

**Proof of Proposition 3.1.** The proof of Proposition 3.1(1) is similar to that of Theorem 2 in Carrasco and Florens (2000) and is not repeated here. The optimality argument follows from the proof of Theorem 8 in Carrasco and Florens (2000).

We need as preliminary result to the proof of Proposition 3.2 the following lemma. It generalizes Theorem 7 of Carrasco and Florens (2000) to the case where  $K_T$  has typically a

slower rate of convergence than  $T^{-1/2}$ .

**Lemma B.2** *Assume  $K_T$  is such that  $\|K_T - K\| = O_p(T^{-a})$ ,  $(K_T^{\alpha_T})^{-1} = (K_T^2 + \alpha_T I)^{-1} K_T$ , and  $\alpha_T$  goes to zero. We have*

$$\left\| (K_T^{\alpha_T})^{-1/2} - (K^{\alpha_T})^{-1/2} \right\| = O_p\left(\frac{1}{T^a \alpha_T^{3/4}}\right).$$

*Let  $\aleph$  be a subset of  $\Theta$  (or  $\Theta$  itself). Let  $f(\theta)$  and  $f_T(\theta)$  be such that  $\sup_{\theta \in \aleph} \|f_T(\theta) - f(\theta)\| = O_p(1/\sqrt{T})$ . Then, for  $f(\theta) \in \mathcal{H}(K)$  for all  $\theta \in \aleph$  and  $\sup_{\theta \in \aleph} \|f(\theta)\| < \infty$ , we have*

$$\sup_{\theta \in \aleph} \left\| (K_T^{\alpha_T})^{-1/2} f_T(\theta) - K^{-1/2} f(\theta) \right\| = O_p\left(\frac{1}{T^a \alpha_T^{3/4}}\right).$$

**Proof of Lemma B.2.** Note that

$$\begin{aligned} & \left\| (K_T^{\alpha_T})^{-1/2} - (K^{\alpha_T})^{-1/2} \right\| \\ &= \left\| (\alpha_T + K_T^2)^{-1/2} K_T^{1/2} - (\alpha_T + K^2)^{-1/2} K^{1/2} \right\| \\ &= \left\| (\alpha_T + K_T^2)^{-1/2} K_T^{1/2} - (\alpha_T + K_T^2)^{-1/2} K^{1/2} \right\| \\ & \quad + \left\| (\alpha_T + K_T^2)^{-1/2} K^{1/2} - (\alpha_T + K^2)^{-1/2} K^{1/2} \right\| \\ &\leq \underbrace{\left\| (\alpha_T + K_T^2)^{-1/2} \right\|}_{\leq \alpha_T^{-1/2}} \underbrace{\left\| K_T^{1/2} - K^{1/2} \right\|}_{=O_p(T^{-a})} \end{aligned} \tag{B.1}$$

$$+ \left\| \left[ (\alpha_T + K_T^2)^{-1/2} - (\alpha_T + K^2)^{-1/2} \right] K^{1/2} \right\| \tag{B.2}$$

Using  $A^{-1/2} - B^{-1/2} = A^{-1/2} [B^{1/2} - A^{1/2}] B^{-1/2}$ , we get

$$\begin{aligned} & (B.2) \\ &= \left\| (\alpha_T + K_T^2)^{-1/2} \left[ (\alpha_T + K_T^2)^{1/2} - (\alpha_T + K^2)^{1/2} \right] (\alpha_T + K^2)^{-1/2} K^{1/2} \right\| \\ &\leq \underbrace{\left\| (\alpha_T + K_T^2)^{-1/2} \right\|}_{\leq \alpha_T^{-1/2}} \underbrace{\left\| (\alpha_T + K_T^2)^{1/2} - (\alpha_T + K^2)^{1/2} \right\|}_{=O_p(T^{-a})} \underbrace{\left\| (\alpha_T + K^2)^{-1/4} \right\|}_{\leq \alpha_T^{-1/4}} \underbrace{\left\| (\alpha_T + K^2)^{-1/4} K^{1/2} \right\|}_{\leq 1}. \end{aligned}$$

Hence  $(B.2) = O_p\left(T^{-a}\alpha_T^{-3/4}\right)$ . The first equality of Lemma B.2 follows from the fact that  $(B.1)$  is negligible with respect to  $(B.2)$ . The second equality can be established from the following decomposition:

$$\begin{aligned} & \sup_{\theta \in \mathfrak{N}} \left\| (K_T^{\alpha_T})^{-1/2} f_T(\theta) - K^{-1/2} f(\theta) \right\| \\ \leq & \sup_{\theta \in \mathfrak{N}} \left\| (K^{\alpha_T})^{-1/2} f(\theta) - K^{-1/2} f(\theta) \right\| \end{aligned} \quad (B.3)$$

$$+ \sup_{\theta \in \mathfrak{N}} \left\| (K_T^{\alpha_T})^{-1/2} f(\theta) - (K^{\alpha_T})^{-1/2} f(\theta) \right\| \quad (B.4)$$

$$+ \sup_{\theta \in \mathfrak{N}} \left\| (K_T^{\alpha_T})^{-1/2} f_T(\theta) - (K_T^{\alpha_T})^{-1/2} f(\theta) \right\| \quad (B.5)$$

From the proof of Theorem 7 of Carrasco and Florens (2000), it follows that

$$\left\| (K^{\alpha_T})^{-1/2} f(\theta) - K^{-1/2} f(\theta) \right\|$$

goes to zero as  $\alpha_T$  goes to zero. Moreover, from the first part of Lemma B.2, we have

$$\begin{aligned} (B.4) & \leq \left\| (K_T^{\alpha_T})^{-1/2} - (K^{\alpha_T})^{-1/2} \right\| \sup_{\theta \in \mathfrak{N}} \|f(\theta)\| \\ & = O_p\left(T^{-a}\alpha_T^{-3/4}\right). \end{aligned}$$

$$\begin{aligned} (B.5) & \leq \left\| (K_T^{\alpha_T})^{-1/2} \right\| \sup_{\theta \in \mathfrak{N}} \|f_T(\theta) - f(\theta)\| \\ & = O_p\left(\alpha_T^{-1/4} T^{-1/2}\right). \end{aligned}$$

using the fact that  $\left\| (K_T^{\alpha_T})^{-1/2} \right\| \leq \left\| (\alpha_T + K^2)^{-1/4} \right\| \left\| (\alpha_T + K^2)^{-1/4} K^{1/2} \right\| \leq \alpha_T^{-1/4}$ . The result follows.

**Proof of Proposition 3.2.** First we prove consistency, second we prove asymptotic normality.

**Consistency.** The consistency follows from Theorem 3.4. of White (1994) under the following three conditions.

(a)  $Q_T(\theta) = - \left\| (K_T^{\alpha_T})^{-1/2} \hat{h}_T(\theta) \right\|^2$  is a continuous function of  $\theta$  for all finite  $T$ .

(b)  $Q_T(\theta) \xrightarrow{P} Q(\theta) = -\left\| (K)^{-1/2} E^{\theta_0} h_t(\theta) \right\|^2$  uniformly on  $\Theta$ .

(c)  $Q(\theta)$  has a unique maximizer  $\theta_0$  on  $\Theta$ .

We check successively (a), (b), and (c). (a)  $\hat{h}_T(\theta)$  is continuous in  $\theta$  by Assumption A.4 (ii). For  $T$  finite,  $(K_T^{\alpha_T})^{-1/2}$  is a bounded operator (because  $\alpha_T > 0$ ) and therefore  $\left\| (K_T^{\alpha_T})^{-1/2} \hat{h}_T(\theta) \right\|^2$  is a continuous function of  $\theta$ .

(b) The uniform convergence as  $T$  and  $T^a \alpha_T^{3/4}$  go to infinity follows from A.4 and Lemma B.2.

(c) Assumption A.5 implies that  $K$  is a positive definite operator. By the property of the norm, we have  $\left\| E^{\theta_0} h(\theta) \right\|_K^2 = 0 \Rightarrow E^{\theta_0} h(\theta) = 0$  which implies  $\theta = \theta_0$  by Assumption A.3.

**Asymptotic Normality.** Using a Taylor expansion of the first order condition

$$\left\langle \nabla_{\theta} \hat{h}_T(\hat{\theta}_T), \hat{h}_T(\hat{\theta}_T) \right\rangle_{K_T^{\alpha_T}} = 0$$

around  $\theta_0$ , we obtain

$$\begin{aligned} \sqrt{T}(\hat{\theta}_T - \theta_0) = & - \left[ \left\langle \nabla_{\theta} \hat{h}_T(\hat{\theta}_T), \nabla_{\theta} \hat{h}_T(\bar{\theta}) \right\rangle_{K_T^{\alpha_T}} \right]^{-1} \\ & \times \left\langle \nabla_{\theta} \hat{h}_T(\hat{\theta}_T), \sqrt{T} \hat{h}_T(\theta_0) \right\rangle_{K_T^{\alpha_T}} \end{aligned}$$

where  $\bar{\theta}$  is a mean value. We need to establish:

$$\text{N1} - \left\langle \nabla_{\theta} \hat{h}_T(\hat{\theta}_T), \nabla_{\theta} \hat{h}_T(\bar{\theta}) \right\rangle_{K_T^{\alpha_T}} \xrightarrow{P} \left\langle E^{\theta_0} \nabla_{\theta} h_t(\theta_0), E^{\theta_0} \nabla_{\theta} h_t(\theta_0) \right\rangle_K.$$

$$\text{N2} - \left\langle \nabla_{\theta} \hat{h}_T(\hat{\theta}_T), \sqrt{T} \hat{h}_T(\theta_0) \right\rangle_{K_T^{\alpha_T}} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \left\langle E^{\theta_0} \nabla_{\theta} h_t(\theta_0), E^{\theta_0} \nabla_{\theta} h_t(\theta_0) \right\rangle_K\right).$$

N1 -Note that  $E^{\theta_0} \nabla_{\theta} h_t(\theta) \in \mathcal{H}(K)$  by Assumption A.5. N1 follows directly from

$$\left\| (K_T^{\alpha_T})^{-1/2} \nabla_{\theta} \hat{h}_T(\hat{\theta}_T) - K^{-1/2} E^{\theta_0} \nabla_{\theta} h_t(\theta_0) \right\| \xrightarrow{P} 0, \quad (\text{B.6})$$

$$\left\| (K_T^{\alpha_T})^{-1/2} \nabla_{\theta} \hat{h}_T(\bar{\theta}) - K^{-1/2} E^{\theta_0} \nabla_{\theta} h_t(\theta_0) \right\| \xrightarrow{P} 0. \quad (\text{B.7})$$

We prove (B.6). The same proof applies to (B.7). We have

$$\begin{aligned} & \left\| (K_T^{\alpha_T})^{-1/2} \nabla_{\theta} \hat{h}_T(\hat{\theta}_T) - K^{-1/2} E^{\theta_0} \nabla_{\theta} h_t(\theta_0) \right\| \\ \leq & \left\| (K_T^{\alpha_T})^{-1/2} \nabla_{\theta} \hat{h}_T(\hat{\theta}_T) - K^{-1/2} E^{\theta_0} \nabla_{\theta} h_t(\hat{\theta}_T) \right\| \end{aligned} \quad (\text{B.8})$$

$$+ \left\| K^{-1/2} E^{\theta_0} \nabla_{\theta} h_t(\hat{\theta}_T) - K^{-1/2} E^{\theta_0} \nabla_{\theta} h_t(\theta_0) \right\| \quad (\text{B.9})$$

Let  $\aleph$  be some neighborhood of  $\theta_0$ . By Lemma B.2 and Assumption A.5(iii), we have for  $T$  sufficiently large

$$(B.8) \leq \sup_{\theta \in \aleph} \left\| (K_T^{\alpha_T})^{-1/2} \nabla_{\theta} \hat{h}_T(\theta) - K^{-1/2} E^{\theta_0} \nabla_{\theta} h_t(\theta) \right\| = O_p \left( T^{-a} \alpha_T^{-3/4} \right).$$

And by the continuity in  $\theta$  of  $E^{\theta_0} \nabla_{\theta} h_t(\theta)$  (Assumption A.5(ii)) and the consistency of  $\hat{\theta}_T$ , we have (B.9)  $\xrightarrow{P} 0$ .

N2 - We have

$$\begin{aligned} & \left\langle (K_T^{\alpha_T})^{-1/2} \nabla_{\theta} \hat{h}_T(\hat{\theta}_T), (K_T^{\alpha_T})^{-1/2} \sqrt{T} \hat{h}_T(\theta_0) \right\rangle \\ = & \left\langle (K_T^{\alpha_T})^{-1/2} \nabla_{\theta} \hat{h}_T(\hat{\theta}_T) - K^{-1/2} E^{\theta_0} \nabla_{\theta} h_t(\theta_0), (K_T^{\alpha_T})^{-1/2} \sqrt{T} \hat{h}_T(\theta_0) \right\rangle \end{aligned} \quad (\text{B.10})$$

$$+ \left\langle K^{-1/2} E^{\theta_0} \nabla_{\theta} h_t(\theta_0), (K_T^{\alpha_T})^{-1/2} \sqrt{T} \hat{h}_T(\theta_0) \right\rangle \quad (\text{B.11})$$

$$(B.10) \leq \left\| (K_T^{\alpha_T})^{-1/2} \nabla_{\theta} \hat{h}_T(\hat{\theta}_T) - K^{-1/2} E^{\theta_0} \nabla_{\theta} h_t(\theta_0) \right\| \left\| (K_T^{\alpha_T})^{-1/2} \right\| \left\| \sqrt{T} \hat{h}_T(\theta_0) \right\|.$$

Applying again Lemma B.2, we obtain

$$\begin{aligned} \left\| (K_T^{\alpha_T})^{-1/2} \nabla_{\theta} \hat{h}_T(\hat{\theta}_T) - K^{-1/2} E^{\theta_0} \nabla_{\theta} h_t(\theta_0) \right\| &= O_p \left( 1 / \left( T^a \alpha_T^{3/4} \right) \right), \\ \left\| (K_T^{\alpha_T})^{-1/2} \right\| &= O_p \left( 1 / \alpha_T^{1/4} \right). \end{aligned}$$

The term (B.10) is  $O_p(1 / (T^a \alpha_T)) = o_p(1)$  as  $T^a \alpha_T$  goes to infinity by assumption.

The term (B.11) can be decomposed as

$$\begin{aligned} & \left\langle K^{-1/2} E^{\theta_0} \nabla_{\theta} h_t(\theta_0), (K_T^{\alpha_T})^{-1/2} \sqrt{T} \hat{h}_T(\theta_0) \right\rangle \\ &= \left\langle K^{-1/2} E^{\theta_0} \nabla_{\theta} h_t(\theta_0), \left( (K_T^{\alpha_T})^{-1/2} - (K^{\alpha_T})^{-1/2} \right) \sqrt{T} \hat{h}_T(\theta_0) \right\rangle \end{aligned} \quad (\text{B.12})$$

$$+ \left\langle K^{-1/2} E^{\theta_0} \nabla_{\theta} h_t(\theta_0), (K^{\alpha_T})^{-1/2} \sqrt{T} \hat{h}_T(\theta_0) \right\rangle. \quad (\text{B.13})$$

We have

$$(B.12) \leq \left\| K^{-1/2} E^{\theta_0} \nabla_{\theta} h_t(\theta_0) \right\| \left\| (K_T^{\alpha_T})^{-1/2} - (K^{\alpha_T})^{-1/2} \right\| \left\| \sqrt{T} \hat{h}_T(\theta_0) \right\| = O_p \left( \frac{1}{T^{\alpha} \alpha_T^{3/4}} \right)$$

by Lemma B.2. It remains to show that (B.13) is asymptotically normal. Denote  $(\lambda_j, \phi_j : j = 1, 2, \dots)$  the eigenvalues and eigenfunctions of  $K$ .

$$(B.13) = \sum_{t=1}^T \sum_{j=1}^{\infty} \frac{1}{\sqrt{T}} \frac{1}{\sqrt{\lambda_j^2 + \alpha_T}} \left\langle E^{\theta_0} \nabla_{\theta} h_t, \phi_j \right\rangle \overline{\langle h_t, \phi_j \rangle} \equiv \frac{1}{\sqrt{T}} \sum_{t=1}^T \tilde{Z}_{Tt}.$$

where  $\theta_0$  is dropped to simplify the notation.  $\tilde{Z}_{Tt}$  is a  $q$ -vector. We apply Cramer Wold device (Theorem A.3.8. of White (1994)) to prove the asymptotic normality of  $\frac{1}{\sqrt{T}} \sum_{t=1}^T \tilde{Z}_{Tt}$ . Let  $\beta$  be a  $q$ -vector of constants so that  $\beta' \beta = 1$ . Denote  $Z_{Tt} = \beta' \tilde{Z}_{Tt}$ . By Theorem A.3.7. of White (1994), we have

$$\frac{1}{\sigma_T} \sum_{t=1}^T Z_{Tt} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

if the following assumptions are satisfied:

- (a)  $E^{\theta_0} (|Z_{Tt}|^{\mu}) \leq \Delta < \infty$  for some  $\mu > 2$
- (b)  $Z_{Tt}$  is near epoch dependent on  $\{V_t\}$  of size -1 where  $\{V_t\}$  is mixing of size  $-2\mu/(\mu - 2)$ .
- (c)  $\sigma_T^2 \equiv \text{var} \left( \sum_{t=1}^T Z_{Tt} \right)$  satisfies  $\sigma_T^{-2} = O(T^{-1})$ .

We verify Conditions (a) to (c) successively. (a) is satisfied for all  $\mu$  because  $Z_{Tt}$  is bounded with probability one. Indeed, we have

$$\begin{aligned}
Z_{Tt} &= \sum_{j=1}^{\infty} \frac{1}{\sqrt{\lambda_j^2 + \alpha_T}} \langle \beta' E^{\theta_0} \nabla_{\theta} h, \phi_j \rangle \overline{\langle h_t, \phi_j \rangle} \\
&\leq \left( \sum_{j=1}^{\infty} \frac{1}{\lambda_j^2 + \alpha_T} \left| \langle \beta' E^{\theta_0} \nabla_{\theta} h, \phi_j \rangle \right|^2 \right)^{1/2} \left( \sum_{j=1}^{\infty} |\langle h_t, \phi_j \rangle|^2 \right)^{1/2}
\end{aligned}$$

by Cauchy-Schwartz inequality. As  $\mu_j^2 + \alpha_T \geq \mu_j^2$  and

$$\begin{aligned}
\sum_{j=1}^{\infty} \frac{1}{\lambda_j^2} \left| \langle \beta' E^{\theta_0} \nabla_{\theta} h, \phi_j \rangle \right|^2 &= \beta' \langle E^{\theta_0} \nabla_{\theta} h, E^{\theta_0} \nabla_{\theta} h \rangle_K \beta, \\
\sum_{j=1}^{\infty} |\langle h_t, \phi_j \rangle|^2 &= \|h_t\|^2.
\end{aligned}$$

We have

$$Z_{Tt} \leq \beta' \langle E^{\theta_0} \nabla_{\theta} h, E^{\theta_0} \nabla_{\theta} h \rangle_K \beta \|h_t\|^2 \leq \Delta < \infty$$

with probability one. (b) It is easy to verify that  $Z_{Tt}$  is near epoch dependent on  $\{h_t\}$  of arbitrary size. (c) We have

$$\begin{aligned}
&\frac{1}{T} \text{var} \left( \sum_{t=1}^T Z_{Tt} \right) \\
&= \text{var} \left\{ \sum_{j=1}^{\infty} \frac{1}{\sqrt{\lambda_j^2 + \alpha_T}} \langle \beta' E^{\theta_0} \nabla_{\theta} h, \phi_j \rangle \overline{\langle \sqrt{T} \hat{h}_T, \phi_j \rangle} \right\} \\
&= \sum_{j=1}^{\infty} \frac{1}{\sqrt{\lambda_j^2 + \alpha_T}} \left| \langle \beta' E^{\theta_0} \nabla_{\theta} h, \phi_j \rangle \right|^2 \text{var} \left( \langle \sqrt{T} \hat{h}_T, \phi_j \rangle \right) \\
&\quad + \sum_{i \neq j} \frac{1}{\sqrt{\lambda_i^2 + \alpha_T} \sqrt{\lambda_j^2 + \alpha_T}} \langle \beta' E^{\theta_0} \nabla_{\theta} h, \phi_i \rangle \overline{\langle \beta' E^{\theta_0} \nabla_{\theta} h, \phi_j \rangle} \text{cov} \left( \langle \sqrt{T} \hat{h}_T, \phi_i \rangle, \langle \sqrt{T} \hat{h}_T, \phi_j \rangle \right).
\end{aligned}$$

Using as before  $\lambda_j^2 + \alpha_T \geq \lambda_j^2$ , both sums can be bounded by a term that does not depend on  $T$ , therefore we may, in passing at the limit as  $T \rightarrow \infty$ , interchange the limit and the summation. By Lemma 3.1, we have  $\sqrt{T} \hat{h}_T \xrightarrow{L} \mathcal{N}(0, K)$  and hence

$$\lim_{T \rightarrow \infty} \text{cov} \left( \langle \sqrt{T} \hat{h}_T, \phi_i \rangle, \langle \sqrt{T} \hat{h}_T, \phi_j \rangle \right) = \langle K \phi_i, \phi_j \rangle = \begin{cases} \lambda_j & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}.$$

Therefore

$$\frac{1}{T} \text{var} \left( \sum_{t=1}^T Z_{Tt} \right) \rightarrow \sum_j \frac{1}{\lambda_j} \left| \left\langle \beta' E^{\theta_0} \nabla_{\theta} h, \phi_j \right\rangle \right|^2 = \beta' \left\langle E^{\theta_0} \nabla_{\theta} h, E^{\theta_0} \nabla_{\theta} h' \right\rangle_K \beta$$

as  $T \rightarrow \infty$  proving that  $\sigma_T^{-2} = O(T^{-1})$ .

Hence, we have

$$(B.13) \xrightarrow{\mathcal{L}} \mathcal{N} \left( 0, \left\langle E^{\theta_0} \nabla_{\theta} h, E^{\theta_0} \nabla_{\theta} h' \right\rangle_K \right).$$

This completes the proof.

**Proof of Proposition 3.3.** Let  $\|A\|_{HS}$  denote the Hilbert Schmidt norm of the operator  $A$  (see Dautray and Lions, 1988, for a definition and the properties of the Hilbert-Schmidt norm). If  $\|A\|$  denotes the usual operator norm,  $\|A\| \leq \|A\|_{HS}$ . We have

$$\|K_T - K\|_{HS}^2 = \int \int \left| \hat{k}_T(\tau_1, \tau_2) - k(\tau_1, \tau_2) \right|^2 \pi(\tau_1) \pi(\tau_2) d\tau_1 d\tau_2.$$

(i) Here  $\hat{k}_T(\tau_1, \tau_2)$  depends on a first step estimator  $\hat{\theta}^1$ . Remark that

$$\|K_T - K\|_{HS}^2 = \left\| \hat{k}_T(\tau_1, \tau_2) - k(\tau_1, \tau_2) \right\|_{L^2 \times L^2}.$$

To simplify, we denote  $g_t(\theta) = h_t(\tau_1, \theta) h_t(\tau_2, \theta)$ . Applying the mean value theorem, we get

$$\begin{aligned} \hat{k}_T(\tau_1, \tau_2) &= \frac{1}{T} \sum_{t=1}^T g_t(\hat{\theta}^1) = \frac{1}{T} \sum_{t=1}^T g_t(\theta_0) + \frac{1}{T} \sum_{t=1}^T \nabla_{\theta} g_t(\tilde{\theta}) (\hat{\theta}^1 - \theta_0), \\ \left\| \hat{k}_T(\tau_1, \tau_2) - k(\tau_1, \tau_2) \right\|_{L^2 \times L^2} &\leq \left\| \frac{1}{T} \sum_{t=1}^T g_t(\theta_0) - E^{\theta_0}(g_t(\theta_0)) \right\|_{L^2 \times L^2} \\ &\quad + \left\| \frac{1}{T} \sum_{t=1}^T \nabla_{\theta} g_t(\tilde{\theta}) \right\|_{L^2 \times L^2} \left\| \hat{\theta}^1 - \theta_0 \right\|_E \end{aligned}$$

where  $\tilde{\theta}$  is between  $\theta_0$  and  $\hat{\theta}^1$ . As in Lemma 3.1, we apply Politis and Romano (1994) to establish that the process  $\frac{1}{\sqrt{T}} \sum_{t=1}^T [g_t(\theta_0) - E^{\theta_0}(g_t(\theta_0))]$  converges weakly in  $L^2(\pi) \times L^2(\pi)$



and hence

$$\left\| \frac{1}{T} \sum_{t=1}^T g_t(\theta_0) - E^{\theta_0}(g_t(\theta_0)) \right\|_{L^2 \times L^2} = O_p(T^{-1/2}).$$

Moreover we have

$$\left\| \frac{1}{T} \sum_{t=1}^T \nabla_{\theta} g_t(\tilde{\theta}) \right\|_{L^2 \times L^2} \leq \frac{1}{T} \sum_{t=1}^T \left\| \nabla_{\theta} g_t(\tilde{\theta}) \right\|_{L^2 \times L^2} \xrightarrow{P} E^{\theta_0} \left\| \nabla_{\theta} g_t(\theta_0) \right\|_{L^2 \times L^2} < \infty$$

by Theorem A.2.2 of White (1994) and Assumption A.6(ii). The result follows.

(ii) Now  $\hat{k}_T(\tau_1, \tau_2)$  does not depend on a first step estimator. We use the following result. If  $X_T \geq 0$  is such that  $EX_T = O(1)$  then  $X_T = O_p(1)$ . This result is proved in Darolles, Florens, and Renault (2000, footnote 12). We can exchange the order of integration and expectation by Fubini's theorem to obtain

$$E^{\theta_0} \|K_T - K\|_{HS}^2 = \int \int E^{\theta_0} \left| \hat{k}_T(\tau_1, \tau_2) - k(\tau_1, \tau_2) \right|^2 \pi(\tau_1) \pi(\tau_2) d\tau_1 d\tau_2.$$

We have

$$E \left| \hat{k}_T(\tau_1, \tau_2) - k(\tau_1, \tau_2) \right|^2 \leq E^{\theta_0} \left| \hat{k}_T(\tau_1, \tau_2) - E^{\theta_0} \hat{k}_T(\tau_1, \tau_2) \right|^2 + \left| E_T^{\theta_0} \hat{k}_T(\tau_1, \tau_2) - k(\tau_1, \tau_2) \right|^2$$

Parzen (1957) and Andrews (1991) consider kernel estimators of the covariance of real-valued random variables. Here, we have complex-valued  $h_t$  but their results remain valid.

From Parzen (1957, Theorem 6) and Andrews (1991), we have

$$\begin{aligned} \lim_{T \rightarrow \infty} S_T^\nu \left( E_T^{\theta_0} \hat{k}_T(\tau_1, \tau_2) - k(\tau_1, \tau_2) \right) &= -2\pi\omega_\nu f^{(\nu)}, \\ \lim_{T \rightarrow \infty} \frac{T}{S_T} E^{\theta_0} \left| \hat{k}_T(\tau_1, \tau_2) - E_T^{\theta_0} \hat{k}_T(\tau_1, \tau_2) \right|^2 &= 8\pi^2 f^2 \int \omega^2(x) dx. \end{aligned}$$

To complete the proof, we need to be able to exchange the lim and integrals in

$\lim_{T \rightarrow \infty} \int \int E \left| \hat{k}_T(\tau_1, \tau_2) - k(\tau_1, \tau_2) \right|^2 \pi(\tau_1) \pi(\tau_2) d\tau_1 d\tau_2$ . This is possible if  $E \left| \hat{k}_T(\tau_1, \tau_2) - k(\tau_1, \tau_2) \right|^2$  is uniformly bounded in  $\tau_1, \tau_2$  and  $T$  by Theorem 5.4 of Billingsley (1995). The boundedness in  $\tau_1, \tau_2$  results from the fact that  $\vartheta$  is bounded. The boundedness in  $T$  follows from Andrews

(1981, page 853) using  $X_T = \left| \frac{T}{S_T} \left( \hat{k}_T(\tau_1, \tau_2) - k(\tau_1, \tau_2) \right)^2 \right|$ . He shows that  $\sup_{T \geq 1} EX_T^2 < \infty$  under assumptions that are satisfied here.

Hence if  $S_T^{2\nu+1}/T \rightarrow \gamma$ , we have

$$\lim_{T \rightarrow \infty} \frac{T}{S_T} E^{\theta_0} \|K_T - K\|_{HS}^2 = 4\pi^2 \left( \frac{\omega_\nu^2 f^{(\nu)2}}{\gamma} + 2f^2 \int \omega^2(x) dx \right).$$

and therefore,  $E^{\theta_0} \|K_T - K\|_{HS}^2 = O\left(T^{-2\nu/(2\nu+1)}\right)$ . This yields the result.

**Proof of Proposition 3.4.** The C-GMM estimator is solution of:

$$\hat{\theta}_T = \arg \min_{\theta} \left\| (K_T^{\alpha_T})^{-1/2} h_T(\theta) \right\|^2 \iff \hat{\theta}_T = \arg \min_{\theta} \left\langle (K_T^{\alpha_T})^{-1} \hat{h}_T(\cdot; \theta), \hat{h}_T(\cdot; \theta) \right\rangle \quad (\text{B.14})$$

Let  $g = (K_T^{\alpha_T})^{-1} \hat{h}_T(\theta)$  so that  $g$  satisfies:

$$\begin{aligned} (\alpha_T I_T + K_T^2) g &= K_T \hat{h}_T(\theta) \\ \alpha_T g(\tau) + \frac{1}{(T-q)^2} \sum_{t,l=1}^T h_t(\tau; \hat{\theta}^1) c_{tl} b_l &= \frac{1}{T-q} \sum_{t=1}^T h_t(\tau; \hat{\theta}^1) v_t(\theta) \end{aligned} \quad (\text{B.15})$$

with

$$b_l = \int U h_l(\tau; \hat{\theta}^1) g(\tau) \pi(\tau) d\tau.$$

First, we compute  $b_l, l = 1, \dots, T$ . We premultiply (B.15) by  $U h_k(\tau; \hat{\theta}^1) \pi(\tau)$  and integrate with respect to  $\tau$  to obtain:

$$\alpha_T b_k + \frac{1}{(T-q)^2} \sum_{t,l=1}^T c_{kt} c_{tl} b_l = \frac{1}{T-q} \sum_{t=1}^T c_{kt} v_t(\theta). \quad (\text{B.16})$$

Using the matrix notation, (B.16) can be rewritten

$$\left[ \alpha_T I_T + C^2 \right] \underline{b} = C \underline{v}(\theta).$$

where  $\underline{b} = [b_1, \dots, b_T]'$ . Solving in  $\underline{b}$ , we get

$$\underline{b} = [\alpha_T I_T + C^2]^{-1} C \underline{v}(\theta). \quad (\text{B.17})$$

Now we want to compute  $\langle g, h_T(\theta) \rangle$  that appears in (B.14). To do so, we multiply all terms of (B.15) with  $\overline{\hat{h}_T(\tau; \theta)} \pi(\tau)$  and integrate with respect to  $\tau$ :

$$\alpha_T \langle g, \hat{h}_T(\theta) \rangle + \frac{1}{(T-q)^2} \sum_{t,l=1}^T w_t(\theta) c_{tl} b_l = \frac{1}{T-q} \sum_{t=1}^T w_t(\theta) v_t(\theta).$$

So that

$$\langle g, h_T(\theta) \rangle = \frac{1}{\alpha_T (T-q)} [\underline{w}'(\theta) \underline{v}(\theta) - \underline{w}'(\theta) C \underline{b}]$$

and using (B.17), we obtain

$$\langle g, h_T(\theta) \rangle = \frac{1}{\alpha_T (T-q)} \underline{w}'(\theta) \left[ I_T - C [\alpha_T I_T + C^2]^{-1} C \right] \underline{v}(\theta).$$

Now we need to show that

$$\left[ I_T - C [\alpha_T I_T + C^2]^{-1} C \right] = \alpha_T [\alpha_T I_T + C^2]^{-1}$$

Note that  $C$  is hermitian that is  $C^* = C$  where  $C^*$  denotes  $\bar{C}'$ . From Horn and Johnson (1985), all hermitian matrices are normal and hence  $C$  can be written as  $C = U D U^*$  where  $D$  is a diagonal matrix and  $U$  satisfies  $U^* = U^{-1}$ . We have

$$\begin{aligned} I_T - \alpha_T [\alpha_T I_T + C^2]^{-1} &= I_T - \alpha_T U [\alpha_T I_T + D^2]^{-1} U^* \\ &= U \left[ U^* U - \alpha_T [\alpha_T I_T + D^2]^{-1} \right] U^* \\ &= U \left[ I_T - \alpha_T [\alpha_T I_T + D^2]^{-1} \right] U^* \\ &= U \left[ \frac{1}{\alpha_T} [\alpha_T I_T + D^2] - I_T \right] \alpha_T [\alpha_T I_T + D^2]^{-1} U^* \\ &= U D^2 [\alpha_T I_T + D^2]^{-1} U^* \\ &= C [\alpha_T I_T + C^2]^{-1} C. \end{aligned}$$

This yields the result.

**Proof of Proposition 3.5.** The proof is very similar to that of 3.4 and is not repeated here. The consistency follows from Lemma B.2.

For the proof of Proposition A.1, we need the following result.

**Lemma B.3** *Let  $\nu_T(\theta)$  be a process in  $L^2(\pi)$ . If  $\nu_T(\theta)$  is stochastically equicontinuous<sup>20</sup> in the following sense:  $\forall \varepsilon > 0, \eta > 0, \exists \delta > 0$  such that*

$$\overline{\lim}_{T \rightarrow \infty} \Pr \left[ \sup_{\|\theta_1, \theta_2\|, \|\theta_1 - \theta_2\| < \delta} \|\nu_T(\theta_1) - \nu_T(\theta_2)\| > \eta \right] < \varepsilon$$

and  $\|\nu_T(\theta)\| = O_p(1)$  for all  $\theta \in \Theta$ , then

$$\sup_{\theta \in \Theta} \|\nu_T(\theta)\| = O_p(1).$$

Proof of Lemma B.3:

Let  $\varepsilon > 0$ . There exists  $\delta > 0$  such that, as  $\Theta$  is compact, there is a finite open covering such that  $\Theta = \cup_{i=1}^J \Theta_j$  where  $\Theta_j$  are open balls of radius  $\delta$  and center  $\theta_j$ . There are  $\eta$  and  $T_0$  such that for all  $T > T_0$ , we have

$$\begin{aligned} & \Pr \left[ \sup_{\theta \in \Theta} \|\nu_T(\theta)\| > \eta \right] \\ &= \Pr \left[ \sup_{j, \theta \in \Theta_j} \|\nu_T(\theta) - \nu_T(\theta_j) + \nu_T(\theta_j)\| > \eta \right] \\ &\leq \Pr \left[ \sup_{j, \theta \in \Theta_j} \|\nu_T(\theta) - \nu_T(\theta_j)\| > \eta \right] + \Pr \left[ \sup_j \|\nu_T(\theta_j)\| > \eta \right] \\ &\leq \varepsilon/2 + \varepsilon/2 \end{aligned}$$

using the stochastic equicontinuity of  $\nu_T(\theta)$  and  $\|\nu_T(\theta)\| = O_p(1)$ . This completes the proof

<sup>20</sup> This is not the standard definition for stochastic equicontinuity because here  $\nu_T(\theta)$  is a function of  $\tau$  and  $\|\cdot\|$  denotes the norm in  $L^2(\pi)$ .

of the lemma.

**Proof of Proposition A.1.** Assumption A.7  $\Rightarrow$  Assumption A.1 is obvious. We check successively the conditions of Assumption A.4.

A.4(i) and (ii):

$$|h_t| = \left| e^{i(sx_{t+1}+rx_t)} - \psi_\theta(s|x_t) e^{irx_{t+1}} \right| \leq \left| e^{i(sx_{t+1}+rx_t)} \right| + \left| \psi_\theta(s|x_t) e^{irx_{t+1}} \right| \leq 2$$

as  $|\psi_\theta(s|x_t)| \leq 1$  for all  $s$ .  $h_t$  is continuously differentiable by A.8(ii).

A.4(iii): We want to establish that

$$\sup_{\theta \in \Theta} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^T \{h_t(\theta) - E^{\theta_0} h_t(\theta)\} \right\| = O_p(1).$$

Let us denote  $\nu_T(\theta) = \frac{1}{\sqrt{T}} \sum_{t=1}^T \{h_t(\theta) - E^{\theta_0} h_t(\theta)\}$ . The same way as we proved Lemma 3.1, we can prove that  $\nu_T(\theta)$  converges weakly to a Gaussian process with mean zero in  $L^2(\pi)$ . Hence  $\|\nu_T(\theta)\| = O_p(1)$  for all  $\theta \in \Theta$ . By Lemma B.3, it remains to prove the stochastic equicontinuity. We have

$$\begin{aligned} & \nu_T(\theta_1) - \nu_T(\theta_2) \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^T \left\{ -e^{irX_t} (\psi_{\theta_1}(s|X_t) - \psi_{\theta_2}(s|X_t)) + E^{\theta_0} \left[ e^{irX_t} (\psi_{\theta_1}(s|X_t) - \psi_{\theta_2}(s|X_t)) \right] \right\}. \end{aligned}$$

From van der Vaart (1998, Chapter 19), the equicontinuity follows from

$$|f(\theta_1) - f(\theta_2)| \leq B \|\theta_1 - \theta_2\|$$

where  $f(\theta) = e^{irX_t} \psi_\theta(s|X_t)$  and under the extra moment condition on  $E^{\theta_0}(B^2)$ . Using a mean-value theorem on  $\psi_\theta$ , we get  $B = \sup_{\theta \in \Theta} \|\nabla_\theta \psi_\theta(\cdot|X_t)\|$ .

Now, we turn to the term involving  $\nabla_{\theta} \hat{h}_T(\theta)$ . By Politis and Romano (1994, Theorem 2.3(i)),

$$\sqrt{T} \left( \nabla_{\theta} \hat{h}_T(\theta) - E^{\theta_0} \nabla_{\theta} h_t(\theta) \right)$$

converges weakly to a Gaussian process with mean zero under the Assumptions A.8(ii).

Hence

$$\left\| \nabla_{\theta} \hat{h}_T(\theta) - E^{\theta_0} \nabla_{\theta} h_t(\theta) \right\| = O_p \left( T^{-1/2} \right)$$

on  $\Theta$ . It remains to establish uniform convergence, which is satisfied using Lemma B.2 and condition A.8(iii).

A.5(i): Let  $\tau_1 = (r_1, s_1)$  and  $\tau_2 = (r_2, s_2)$ . We have

$$k(\tau_1, \tau_2) = E^{\theta_0} \left[ e^{i(r_1 - r_2)X_t} \{ \psi_{\theta}(s_1 - s_2 | X_t) - \psi_{\theta}(s_1 | X_t) \psi_{\theta}(-s_2 | X_t) \} \right].$$

By changing the order of integrations, we have for all  $\tau_1 = (r_1, s_1)$  :

$$(K\varphi)(\tau_1) = 0 \Leftrightarrow \int e^{ir_1 x} f_{\theta}(x) \left[ \int e^{-ir_2 x} \{ \psi_{\theta}(s_1 - s_2 | x) - \psi_{\theta}(s_1 | x) \psi_{\theta}(-s_2 | x) \} \varphi(r_2, s_2) \pi(\tau_2) d\tau_2 \right] dx = 0.$$

Applying the Fourier inversion formula, we obtain for all  $x, s_1$  :

$$\begin{aligned} \int e^{-ir_2 x} \{ \psi_{\theta}(s_1 - s_2 | x) - \psi_{\theta}(s_1 | x) \psi_{\theta}(-s_2 | x) \} \varphi(r_2, s_2) \pi(\tau_2) d\tau_2 = 0 &\Leftrightarrow \\ \int e^{-ir_2 x} \left\{ \int e^{i(s_1 - s_2)u} f_{\theta}(u | x) du - \left( \int e^{is_1 u} f_{\theta}(u | x) du \right) \psi_{\theta}(-s_2 | x) \right\} \varphi(r_2, s_2) \pi(\tau_2) d\tau_2 = 0 &\Leftrightarrow \\ \int e^{is_1 u} f_{\theta}(u | x) \left\{ \int e^{-ir_2 x} \left( e^{-is_2 u} - \psi_{\theta}(-s_2 | x) \right) \varphi(r_2, s_2) \pi(\tau_2) d\tau_2 \right\} du = 0. \end{aligned}$$

Again applying the Fourier inversion formula, we get for all  $x, u$  :

$$\int e^{-ir_2 x} \left( e^{-is_2 u} - \psi_{\theta}(-s_2 | x) \right) \varphi(r_2, s_2) \pi(\tau_2) d\tau_2 = 0.$$

We see that the second term on the left-hand side does not depend on  $u$ , the solution satisfies necessarily

$$\int e^{-ir_2x} e^{-is_2u} \varphi(r_2, s_2) \pi(\tau_2) d\tau_2 = 0 \text{ for all } x, u \Leftrightarrow \\ \varphi(r_2, s_2) = 0 \text{ for all } r_2 \text{ and } s_2.$$

Hence  $\mathcal{N}(K) = \{0\}$ .

A.5(ii): We check that  $E^{\theta_0} h_t(\theta) \in \mathcal{H}(K)$  for all  $\theta \in \Theta$ . Note that

$$E^{\theta_0} h_t(\theta) = E^{\theta_0} \left[ e^{irX_t} (\psi_{\theta_0}(s|X_t) - \psi_{\theta}(s|X_t)) \right] \equiv g(r, s) \quad (\text{B.18})$$

We apply Lemma B.1 to compute  $\|g\|_K^2$ . We need to find  $G$  such that

$$g(r, s) = E^{\theta_0} \left[ \left( e^{i(sX_{t+1} + rX_t)} - \psi_{\theta}(s|X_t) e^{irX_t} \right) G(X_t, X_{t+1}) \right] \\ = E^{\theta_0} \left[ e^{i(sX_{t+1} + rX_t)} \left\{ G(X_t, X_{t+1}) - E^{\theta_0} [G(X_t, X_{t+1}) | X_t] \right\} \right].$$

Let us denote  $\tilde{G} = G - E^{\theta_0} [G|X_t]$ . We want to solve in  $\tilde{G}$  the equation

$$g(\tau) = \int e^{i(sx_{t+1} + rx_t)} \tilde{G}(x_t, x_{t+1}) f_{\theta_0}(x_{t+1}|x_t) f_{\theta_0}(x_t) dx_{t+1} dx_t.$$

Applying twice the Fourier inversion formula, we obtain a unique solution

$$\tilde{G}(x_t, x_{t+1}) = \frac{1}{(2\pi)^2} \int \int \frac{g(r, s) e^{-i(sx_{t+1} + rx_t)}}{f_{\theta_0}(x_{t+1}|x_t) f_{\theta_0}(x_t)} ds dr. \quad (\text{B.19})$$

We now replace  $g(r, s)$  by its expression (B.18) into (B.19) to calculate  $\tilde{G}$ . Applying the Fourier inversion formula, we have

$$\frac{1}{2\pi} \int e^{-irx_t} \left( \int e^{iru} \psi_{\theta}(s|u) f_{\theta_0}(u) du \right) dr = \psi_{\theta}(s|x_t) f_{\theta_0}(x_t), \\ \frac{1}{2\pi} \int \psi_{\theta}(s|x_t) e^{-isx_{t+1}} ds = f_{\theta}(x_{t+1}|x_t).$$

Hence we have

$$\tilde{G}(x_t, x_{t+1}) = \frac{f_{\theta_0}(x_{t+1}|x_t) - f_{\theta}(x_{t+1}|x_t)}{f_{\theta_0}(x_{t+1}|x_t)}$$

and  $\|g\|_K^2 = E^{\theta_0} \tilde{G}^2 < \infty$  if and only if

$$\int \int \left[ \frac{f_{\theta_0}(x_{t+1}|x_t) - f_{\theta}(x_{t+1}|x_t)}{f_{\theta_0}(x_{t+1}|x_t)} \right]^2 f_{\theta_0}(x_t, x_{t+1}) dx_t dx_{t+1} < \infty$$

for all  $\theta \in \Theta$ . We recognize Pearson's chi-square distance.

A.5(iii): Now, we check that  $E^{\theta_0} \nabla_{\theta} h_t(\theta) \in \mathcal{H}(K)$  for all  $\theta \in \mathfrak{N}$ . We replace  $g(r, s)$  by

$$g(r, s) \equiv E^{\theta_0} \nabla_{\theta} h_t(\theta) = -E^{\theta_0} \left[ e^{irX_t} \nabla_{\theta} \psi_{\theta}(s|X_t) \right]$$

in Equation (B.19) to calculate  $\tilde{G}$ . We again apply the Fourier inversion formula to obtain

$$\begin{aligned} \frac{1}{2\pi} \int e^{-irx_t} \left( \int e^{iru} \nabla_{\theta} \psi_{\theta}(s|u) f_{\theta_0}(u) du \right) dr &= \nabla_{\theta} \psi_{\theta}(s|x_t) f_{\theta_0}(x_t), \\ \frac{1}{2\pi} \int \nabla_{\theta} \psi_{\theta}(s|x_t) e^{-isx_{t+1}} ds &= \frac{1}{2\pi} \nabla_{\theta} \int \psi_{\theta}(s|x_t) e^{-isx_{t+1}} ds \\ &= \nabla_{\theta} f_{\theta}(x_{t+1}|x_t). \end{aligned} \quad (\text{B.20})$$

We are allowed to interchange the order of integration and derivation in (B.20) because of  $\int \sup_{\theta \in \Theta} |\nabla_{\theta} \psi_{\theta}(s|x_t)| ds < \infty$  and by Lemma 3.6 of Newey and McFadden (1994). Hence we have

$$\tilde{G} = -\frac{\nabla_{\theta} f_{\theta}(x_{t+1}|x_t)}{f_{\theta_0}(x_{t+1}|x_t)}$$

and

$$\begin{aligned} \left\| E^{\theta_0} \nabla_{\theta} h_t(\theta) \right\|_K^2 &= E^{\theta_0} \tilde{G} \tilde{G}' \\ &= E^{\theta_0} \left[ \frac{\nabla_{\theta} f_{\theta}(x_{t+1}|x_t)}{f_{\theta_0}(x_{t+1}|x_t)} \left( \frac{\nabla_{\theta} f_{\theta}(x_{t+1}|x_t)}{f_{\theta_0}(x_{t+1}|x_t)} \right)' \right] \end{aligned} \quad (\text{B.21})$$

which is finite by assumption A8(i). When  $\theta = \theta_0$ , the term in (B.21) coincides with the information matrix  $I_{\theta_0}$  which proves the ML-efficiency without using Proposition 3.6.

Assumption A.6(ii) follows from A.8(iii) because



$$\begin{aligned}
\|\nabla_{\theta} h_t(\theta)\|^2 &= \int |e^{irX_t} \nabla_{\theta} \psi_{\theta}(s|X_t)|^2 ds \\
&\leq \int |\nabla_{\theta} \psi_{\theta}(s|X_t)|^2 ds \\
&= \|\nabla_{\theta} \psi_{\theta}(\cdot|X_t)\|^2.
\end{aligned}$$

**Proof of Proposition 4.1.** The asymptotic distribution of  $\hat{\theta}_T$  follows from Propositions 3.2 and A.1. The asymptotic efficiency follows from Equation (B.21).

**Proof of Proposition 4.3.** To simplify the notation, we omit  $\theta_0$  also all the terms in this proof are taken at  $\theta_0$ . Recall that the variance of  $\tilde{\theta}_T$  is given by  $J^{-1}\Sigma J^{-1}$  with

$$\begin{aligned}
J &= E^{\theta_0} (\nabla_{\theta\theta} \ln f_{\theta}(Y_0)) = -E^{\theta_0} [\nabla_{\theta} \ln f_{\theta}(Y_0) (\nabla_{\theta} \ln f_{\theta}(Y_0))'], \\
\Sigma &= \sum_{j=-\infty}^{\infty} E^{\theta_0} [\nabla_{\theta} \ln f_{\theta}(Y_0) (\nabla_{\theta} \ln f_{\theta}(Y_j))'].
\end{aligned}$$

The asymptotic variance of  $\hat{\theta}_T$  is given by Theorem 2 in Carrasco and Florens (2000) by replacing  $B$  by  $\tilde{K}^{-1/2}$ :

$$V = (\|g\|_{\tilde{K}}^2)^{-1} (\tilde{K}^{-1}g, K\tilde{K}^{-1}g) (\|g\|_{\tilde{K}}^2)^{-1}$$

where  $g = E^{\theta_0} (\nabla_{\theta} h)$ . Theorem 2 assumes that  $B$  is a bounded operator, here  $B$  is not bounded but a proof similar to that of Theorem 8 of Carrasco and Florens (2000) would show that the result is also valid for  $\tilde{K}^{-1/2}$ .

a - Calculation of  $\|g\|_{\tilde{K}}^2$ :

We apply results from Lemma B.1. First we check that

$$G_0 = \nabla_{\theta} \ln f_{\theta}(Y_t)$$

belongs to  $C(g)$  that is

$$\begin{aligned}
\nabla_{\theta} \psi_{\theta}(\tau) &= \int \nabla_{\theta} \ln f_{\theta}(y_t) \left( e^{i\tau y_t} - \psi_{\theta}(\tau) \right) f_{\theta}(y_t) dy_t \\
&= \int \nabla_{\theta} f_{\theta}(y_t) \left( e^{i\tau y_t} - \psi_{\theta}(\tau) \right) dy_t \\
&= \int \nabla_{\theta} f_{\theta}(y_t) e^{i\tau y_t} dy_t \\
&= \nabla_{\theta} \int f_{\theta}(y_t) e^{i\tau y_t} dy_t.
\end{aligned}$$

Now consider a general solution  $G = G_0 + G_1$ . The condition  $G \in C(g)$  implies

$$\begin{aligned}
\int G_1(y_t) \left( e^{i\tau y_t} - \psi_{\theta}(\tau) \right) f_{\theta}(y_t) dy_t &= 0 \quad \forall \tau \\
\Leftrightarrow \int (G_1(y_t) - EG_1) e^{i\tau y_t} f_{\theta}(Y_t) dY_t &= 0 \quad \forall \tau \\
&\Rightarrow G_1 - EG_1 = 0 \\
&\Rightarrow E^{\theta_0}(G_0 G_1) = 0.
\end{aligned}$$

This shows that the element of  $C(g)$  with minimal norm is  $G_0$ . Hence we have

$$\|g\|_{\tilde{K}}^2 = E^{\theta_0}(G_0 G_0') = E^{\theta_0} \left[ (\nabla_{\theta} \ln f_{\theta}(Y_t)) (\nabla_{\theta} \ln f_{\theta}(Y_t))' \right].$$

b - Calculation of  $\tilde{K}^{-1}g$  : We verify that  $g = \tilde{K}\omega$  with

$$\omega(\tau) = \int e^{-i\tau v} \nabla_{\theta} \ln f_{\theta}(v) dv$$

where  $v$  is a  $L$ -vector and  $f_{\theta}$  denotes the joint likelihood of  $Y_t$ . Because  $Y_t$  is assumed to be stationary,  $f_{\theta}$  does not depend on  $t$ . We have

$$\begin{aligned}
(\tilde{K}\omega)(\tau_1) &= \int (\psi_\theta(\tau_1 + \tau_2) - \psi_\theta(\tau_1)\psi_\theta(\tau_2)) \int e^{-i\tau_2 v} \nabla_\theta \ln f_\theta(v) \, dv d\tau_2 \\
&= \int \psi_\theta(\tau_1 + \tau_2) \int e^{-i\tau_2 v} \nabla_\theta \ln f_\theta(v) \, dv d\tau_2 - \psi_\theta(\tau_1) \int \nabla_\theta f_\theta(v) \, dv \\
&= \int e^{i\tau_1 y} \left[ \int e^{i\tau_2 y} e^{-i\tau_2 v} \nabla_\theta \ln f_\theta(v) \, dv d\tau_2 \right] f_\theta(y) \, dy \\
&= \int e^{i\tau_1 y} \nabla_\theta \ln f_\theta(y) f_\theta(y) \, dy = g(\tau_1).
\end{aligned}$$

The fourth equality follows from a property of the Fourier transform, see Theorem 4.11.12. in Debnath and Mikusinsky (1999).

c - Calculation of  $(\tilde{K}^{-1}g, K\tilde{K}^{-1}g)$  :

Note that  $(\tilde{K}^{-1}g, K\tilde{K}^{-1}g) = (\omega, K\omega)$ . The kernel of  $K$  is given by

$$k(\tau_1, \tau_2) = \sum_{j=-\infty}^{\infty} \left[ E^{\theta_0} \left( e^{i(\tau_1 Y_0 + \tau_2 Y_j)} \right) - \psi_\theta(\tau_1)\psi_\theta(\tau_2) \right] \equiv \sum_{j=-\infty}^{\infty} k_j(\tau_1, \tau_2).$$

Let us denote  $K_j$  the operator with kernel  $k_j(\tau_1, \tau_2)$ .

$$(K_j\omega)(\tau_1) = \int E^{\theta_0} \left( e^{i(\tau_1 Y_0 + \tau_2 Y_j)} \right) \int e^{-i\tau_2 v} \nabla_\theta \ln f_\theta(v) \, dv d\tau_2$$

because the second term equals zero.

$$\begin{aligned}
(K_j\omega)(\tau_1) &= \int e^{i\tau_1 y_0} \left[ \int \int e^{i\tau_2 y_j} e^{-i\tau_2 v} \nabla_\theta \ln f_\theta(v) \, dv d\tau_2 \right] f_\theta(y_0, y_j) \, dy_0 dy_j \\
&= \int e^{i\tau_1 y_0} \nabla_\theta \ln f_\theta(y_j) f_\theta(y_0, y_j) \, dy_0 dy_j.
\end{aligned}$$

We have

$$\begin{aligned}
(\omega, K_j\omega) &= \int \left[ \int \int e^{i\tau_1 y_0} e^{-i\tau_1 v} \nabla_\theta \ln f_\theta(v) \, dv d\tau_1 \right] \nabla_\theta \ln f_\theta(y_j)' f_\theta(y_0, y_j) \, dy_0 dy_j \\
&= \int \nabla_\theta \ln f_\theta(y_0) \nabla_\theta \ln f_\theta(y_j)' f_\theta(y_0, y_j) \, dy_0 dy_j \\
&= E^{\theta_0} \left[ \nabla_\theta \ln f_\theta(Y_0) (\nabla_\theta \ln f_\theta(Y_j))' \right].
\end{aligned}$$

It follows that

$$(\omega, K\omega) = \sum_{j=-\infty}^{\infty} (\omega, K_j\omega) = \sum_{j=-\infty}^{\infty} E^{\theta_0} \left[ \nabla_{\theta} \ln f_{\theta}(Y_0) (\nabla_{\theta} \ln f_{\theta}(Y_j))' \right]$$

which finishes the proof.

**Proof of Proposition 5.1.** We wish to apply Proposition 3.2 on  $\{\tilde{h}_t^J\}$ . To do this, we need to check that the conditions of this proposition are satisfied for  $\{\tilde{h}_t^J\}$ . The mixing properties of  $\{\tilde{h}_t^J\}$  are the same as those of  $\{X_t\}$ , moreover  $\{\tilde{h}_t^J\}$  is a martingale difference sequence. Hence by Assumption A.7 and Politis and Romano (1994), we have

$$\sqrt{T}\tilde{h}_t^J \Rightarrow \mathcal{N}(0, \tilde{K})$$

as  $T \rightarrow \infty$  in  $\mathbb{L}^2(\pi)$  where  $\tilde{K}$  is the operator with kernel  $\tilde{k}$  satisfying

$$\begin{aligned} \tilde{k}(\tau_1, \tau_2) &= \text{cov}(\tilde{h}_t^J(\tau_1), \tilde{h}_t^J(\tau_2)) \\ &= E \left[ \text{cov}(\tilde{h}_t^J(\tau_1), \tilde{h}_t^J(\tau_2) | Y_t) \right] \\ &\quad + \text{cov} \left[ E(\tilde{h}_t^J(\tau_1) | Y_t), E(\tilde{h}_t^J(\tau_2) | Y_t) \right] \\ &= \frac{1}{J} E_Y E_{\varepsilon} \left[ (\tilde{h}^J(\tau_1) - h(\tau_1)) \overline{(\tilde{h}^J(\tau_2) - h(\tau_2))} | Y_t \right] + \text{cov}(h(\tau_1), h(\tau_2)) \\ &= \frac{1}{J} u(\tau_1, \tau_2) + k(\tau_1, \tau_2). \end{aligned}$$

Note that we use  $E$  and  $\text{cov}$  for the expectation and covariance with respect to both  $\varepsilon_t$  and  $Y_t$ . Therefore  $\tilde{K} = K + U/J$ . Note that  $U$  is a positive definite operator. Assumption A.5 is satisfied under Assumption A.8(i) because

$$\begin{aligned} E h_t(\theta) &= E \tilde{h}_t^J(\theta), \\ E \nabla_{\theta} h_t(\theta) &= E \nabla_{\theta} \tilde{h}_t^J(\theta). \end{aligned}$$

The second equality follows from

$$\begin{aligned}
E\left(\nabla_{\theta}\tilde{h}_t\right) &= E_Y E_{\varepsilon}\left[\nabla_{\theta}\tilde{h}_t|Y_t\right] \\
&= E_Y\left[\nabla_{\theta}E_{\varepsilon}\left(\tilde{h}_t|Y_t\right)\right].
\end{aligned}
\tag{B.22}$$

The order of integration and differentiation in B.22 can be exchanged because the distribution of  $\tilde{\varepsilon}_{j,t}$  does not depend on  $\theta$  and  $E \sup_{\theta \in \Theta} |\nabla_{\theta} H| < \infty$  which is true under A.10(iii). Therefore  $E\left(\nabla_{\theta}\tilde{h}_t^J\right) = E\left(\nabla_{\theta}h\right)$ . Finally, using a proof very similar to that of Proposition A.1, we see that Assumptions A.4(iii), A.6(ii) and A.8(ii)-(iii) are satisfied under Assumption A.10. It is enough to notice that

$$\begin{aligned}
\left|\nabla_{\theta}\tilde{h}_t^J\right| &= \left|\frac{1}{J}\sum_{j=1}^J i s \nabla_{\theta} H\left(X_t, \varepsilon_{j,t+1}, \theta\right) e^{i s X_{t+1}^{\theta j}}\right| \\
&\leq \frac{1}{J}\sum_{j=1}^J \left|\nabla_{\theta} H\left(X_t, \varepsilon_{j,t+1}, \theta\right)\right|
\end{aligned}$$

and  $\left|\nabla_{\theta\theta}\tilde{h}_t^J\right| \leq \frac{1}{J}\sum_{j=1}^J \left|\nabla_{\theta\theta} H\left(X_t, \varepsilon_{j,t+1}, \theta\right)\right|$ . Hence, from Proposition 3.2, we have

$$\sqrt{T}\left(\tilde{\theta}_T - \theta_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \left(\left\langle E^{\theta_0}\left(\nabla_{\theta}\tilde{h}^J\right), E^{\theta_0}\left(\nabla_{\theta}\tilde{h}^J\right)\right\rangle_{\tilde{K}}\right)^{-1}\right).$$

We can rewrite the variance by using  $E\left(\nabla_{\theta}\tilde{h}_t^J\right) = E\left(\nabla_{\theta}h\right)$ .

Now, we show the inequality  $\|g\|_{\tilde{K}}^2 \leq \|g\|_K^2$  for any function  $g$  in the range of  $K$ . For sake of simplicity, we assume  $g$  scalar, the proof for  $g$  vector is very similar. Denote

$$\begin{aligned}
f &= \left(K + \frac{1}{J}U\right)^{-1} g \\
l &= K^{-1}g.
\end{aligned}$$

We have  $\|g\|_{\tilde{K}}^2 = \langle f, g \rangle$  and  $\|g\|_K^2 = \langle l, g \rangle$ . We want to show  $\langle l - f, g \rangle \geq 0$ .

$$\begin{aligned}
& \langle l - f, g \rangle \geq 0 \\
& \Leftrightarrow \langle K(l - f), g \rangle_K \geq 0 \\
& \Leftrightarrow \left\langle \frac{1}{J} Uf, Kf + \frac{1}{J} Uf \right\rangle_K \geq 0 \\
& \Leftrightarrow \frac{1}{J} \langle Uf, f \rangle + \|Uf\|_K^2 \geq 0
\end{aligned}$$

This last inequality is true because  $U$  is definite positive.

**Proof of Proposition 5.2** The consistency holds under Assumptions A.2-A.4. By the geometric ergodicity and the boundedness of  $e^{i\tau Y_t}$  and  $e^{i\tau \tilde{Y}_j}$ , the functional CLT of Chen and White (1998, Theorem 3.9) gives:

$$\begin{aligned}
& \frac{\sqrt{T}}{T} \sum_{t=1}^T \left( e^{i\tau Y_t} - \psi_\theta^L(\tau) \right) \Rightarrow \mathcal{N}(0, K), \\
& \frac{\sqrt{J(T)}}{J(T)} \sum_{j=1}^{J(T)} \left( e^{i\tau \tilde{Y}_j} - \psi_\theta^L(\tau) \right) \Rightarrow \mathcal{N}(0, K).
\end{aligned}$$

as  $T \rightarrow \infty$  in  $\mathbb{L}^2(\pi)$ . The asymptotic normality follows from

$$\begin{aligned}
\sqrt{T} \tilde{h}_T(\tau; \theta_0) &= \frac{\sqrt{T}}{T} \sum_{t=1}^T \left( e^{i\tau Y_t} - \psi_\theta^L(\tau) \right) - \frac{\sqrt{T}}{\sqrt{J(T)}} \frac{\sqrt{J(T)}}{J(T)} \sum_{j=1}^{J(T)} \left( e^{i\tau \tilde{Y}_j} - \psi_\theta^L(\tau) \right) \\
&\xrightarrow{\mathcal{L}} \mathcal{N}(0, (1 + \zeta) K)
\end{aligned}$$

because  $Y_t$  and  $\tilde{Y}_j$  are independent. Let  $\tilde{K} = (1 + \zeta) K$ . Minimizing  $\|\tilde{h}_T\|_{K_T^{\alpha T}}$  is equivalent to minimizing  $\|\tilde{h}_T\|_{\tilde{K}_T^{\alpha T}}$  where  $\tilde{K}_T^{\alpha T}$  denote a regularized estimator of  $\tilde{K}$ . By Proposition 3.2,  $\tilde{\theta}_T$  is asymptotically normal and the inverse of its variance is equal to

$$\left\langle E^{\theta_0}(\nabla_\theta \tilde{h}_t), E^{\theta_0}(\nabla_\theta \tilde{h}_t) \right\rangle_{\tilde{K}} = \frac{1}{(1 + \zeta)} \left\langle E^{\theta_0}(\nabla_\theta \tilde{h}_t), E^{\theta_0}(\nabla_\theta \tilde{h}_t) \right\rangle_K.$$

Now, we compute  $E^{\theta_0}(\nabla_\theta \tilde{h}_t)$ . By Assumption A.11, we have:

$$E^{\theta_0}(\nabla_\theta \tilde{h}_t) = -E^{\theta_0}(\nabla_\theta e^{i\tau \tilde{Y}_j}) = -\nabla_\theta E^{\theta_0}(e^{i\tau \tilde{Y}_j}) = -\nabla_\theta \psi_\theta^L(\tau) = E^{\theta_0}(\nabla_\theta h).$$

**Proof of Lemma 6.1** (1) follows from Chen, Hansen and Carrasco (1999, Theorem 7.1). This theorem states that a scalar diffusion with drift coefficient,  $\mu$ , and diffusion coefficient,  $\sigma$  and non attracting boundaries is  $\beta$ -mixing with geometric decay if  $(\mu/\sigma + 0.5(\partial\sigma/\partial x))$  is negative at the right boundary and positive at the left boundary. Here, we get

$$\begin{aligned}\lim_{r \uparrow \infty} \frac{\mu}{\sigma} - \frac{1}{2} \frac{\partial\sigma}{\partial r} &< 0, \\ \lim_{r \downarrow 0} \frac{\mu}{\sigma} - \frac{1}{2} \frac{\partial\sigma}{\partial r} &> 0.\end{aligned}$$

These conditions are satisfied provided that  $4\gamma - \sigma^2 > 0$ . Note that the stronger condition  $2\gamma - \sigma^2 \geq 0$  guarantees that neither boundary is attracting.

(2) A7: The mixing property follows from (1) where  $\alpha_i = \rho^i$  for some  $0 < \rho < 1$ . Denote  $\theta = (\gamma, \sigma^2, \kappa)$  and  $\Theta$  be a compact set of  $(\mathbf{R}_+^*)^3$ . To see that  $f_\theta(r_t|r_{t-1})$  is continuously differentiable, it suffices to examine the expression of the conditional likelihood (Zhou, 2001):

$$\begin{aligned}f_\theta(r_t|r_{t-1}) &= \sum_{j=0}^{\infty} \text{Gamma}(r_t|j + \lambda, 1) \text{Poisson}(j|cr_{t-1}e^{-\kappa}) \\ &= \sum_{j=0}^{\infty} \frac{r_t^{j+\lambda-1} e^{-r_t} (cr_{t-1}e^{-\kappa})^j e^{-cr_{t-1}e^{-\kappa}}}{\Gamma(j + \lambda) j!} \\ &\equiv \sum_{j=0}^{\infty} u_{\theta j} \text{ (notation)}\end{aligned}\tag{B.23}$$

with  $\lambda = 2\gamma/\sigma^2$ .

A8(i): By developing the square we obtain

$$E^{\theta_0} \left[ \left( 1 - \frac{f_\theta(r_t|r_{t-1})}{f_{\theta_0}(r_t|r_{t-1})} \right)^2 \right] = \int \int \frac{f_\theta(r_t|r_{t-1})^2}{f_{\theta_0}(r_t|r_{t-1})} f_{\theta_0}(r_{t-1}) dr_t dr_{t-1} - 1.$$

Using the notation introduced in (B.23), we have

$$\frac{f_{\theta}(r_t|r_{t-1})^2}{f_{\theta_0}(r_t|r_{t-1})} = \frac{(\sum u_{\theta j})^2}{\sum u_{\theta_0 j}} \leq \frac{\sum u_{\theta j}^2}{\sum u_{\theta_0 j}} \leq \sum \left( \frac{u_{\theta j}^2}{u_{\theta_0 j}} \right).$$

Replacing  $u_{\theta j}$  and  $u_{\theta_0 j}$  by their expressions, we obtain

$$\begin{aligned} & \sum \left( \frac{u_{\theta j}^2}{u_{\theta_0 j}} \right) \\ = & \sum \left\{ \frac{r_t^{j+\tilde{\lambda}-1} e^{-r_t}}{\Gamma(j+\tilde{\lambda})} \right\} \times \frac{\Gamma(j+\lambda_0) \Gamma(j+\tilde{\lambda})}{\Gamma(j+\lambda)^2} \times \frac{\left( r_{t-1} \left( \frac{c^2}{c_0} e^{-(2\kappa-\kappa_0)} \right) \right)^j e^{-r_{t-1}(2ce^{-\kappa}-c_0e^{-\kappa_0})}}{j!} \end{aligned}$$

where  $\tilde{\lambda} = 2\lambda - \lambda_0$ . Remark that the first element of the sum integrates to 1 with respect to  $r_t$  provided  $\tilde{\lambda} > 0$  (which imposes a restriction on  $\lambda$  and therefore  $\Theta$ ). The marginal pdf of  $r_{t-1}$  is a Gamma:

$$f_{\theta}(r_{t-1}) = \frac{\omega^{\lambda}}{\Gamma(\lambda)} r_{t-1}^{\lambda-1} e^{-\omega r_{t-1}}$$

where  $\omega = 2\kappa/\sigma^2$ . Regrouping the terms yields

$$\begin{aligned} \int \int \frac{f_{\theta}(r_t|r_{t-1})^2}{f_{\theta_0}(r_t|r_{t-1})} dr_t f_{\theta_0}(r_{t-1}) dr_{t-1} &= \sum \frac{\Gamma(j+\lambda_0) \Gamma(j+\tilde{\lambda})}{\Gamma(j+\lambda)^2 j!} \\ &\times \left( \frac{c^2}{c_0} e^{-(2\kappa-\kappa_0)} \right)^j \\ &\times \int r_{t-1}^{j+\lambda_0-1} e^{-r_{t-1}(2ce^{-\kappa}-c_0e^{-\kappa_0}+\omega)} dr_{t-1}. \end{aligned}$$

Remark that

$$\begin{aligned} \int r_{t-1}^{j+\lambda_0-1} e^{-r_{t-1}(2ce^{-\kappa}-c_0e^{-\kappa_0}+\omega)} dr_{t-1} &= \frac{\Gamma(j+\lambda_0)}{(2ce^{-\kappa}-c_0e^{-\kappa_0}+\omega)^{j+\lambda_0}} \int \frac{v^{j+\lambda_0-1} e^{-v}}{\Gamma(j+\lambda_0)} dv \\ &= \frac{\Gamma(j+\lambda_0)}{(2ce^{-\kappa}-c_0e^{-\kappa_0}+\omega)^{j+\lambda_0}}. \end{aligned}$$



Hence, it follows that

$$\int \int \frac{f_\theta(r_t|r_{t-1})^2}{f_{\theta_0}(r_t|r_{t-1})} f_{\theta_0}(r_{t-1}) dr_t dr_{t-1} = \sum \frac{\Gamma(j+\lambda_0)^2 \Gamma(j+\tilde{\lambda})}{\Gamma(j+\lambda)^2 j!} q^j \times const. \quad (\text{B.24})$$

where

$$q = \frac{c^2 e^{-(2\kappa-\kappa_0)}}{c_0 (2ce^{-\kappa} - c_0 e^{-\kappa_0} + \omega_0)}.$$

The sum (B.24) is finite provided  $|q| < 1$ . Let  $\Delta = c - c_0$  and  $\Delta' = \kappa - \kappa_0$  where  $\Delta$  and  $\Delta'$  may be positive or negative. We want to show that  $0 < q < 1$  for values of  $\Delta$  and  $\Delta'$  around 0.  $0 < q < 1$  holds if

$$\frac{c^2}{c_0} e^{-(2\kappa-\kappa_0)} < 2ce^{-\kappa} - c_0 e^{-\kappa_0} + \omega_0,$$

which is equivalent to

$$0 < c_0^2 (2 - e^{\Delta'} - e^{-\Delta'}) + 2\Delta c_0 (1 - e^{-\Delta'}) + \Delta^2 e^{-\Delta'} + \frac{c_0^2}{(1 - e^{-\kappa_0})} \equiv g(\Delta, \Delta').$$

Note that  $g(0, 0) = c_0^2 / (1 - e^{-\kappa_0}) > 0$ . By continuity,  $g(\Delta, \Delta')$  is positive on an interval around  $(0, 0)$ . This shows that there exists a compact  $\Theta$  that contains  $\theta_0$  as an interior point and such that the first inequality of A8(i) holds for all  $\theta \in \Theta$ . The proof of the second inequality follows the same line and is omitted here.

A8(ii) and (iii): Remark that the conditional characteristic function can be written as

$$\psi_\theta = b(\theta, \tau) e^{i\tau g(\theta, \tau) r_t}$$

where  $b$  and  $g$  are twice continuously differentiable.

$$\begin{aligned} |\nabla_\theta \psi_\theta| &\leq |\nabla_\theta b| + |\nabla_\theta g| |\tau| |r_t| |\psi_\theta|, \\ \nabla_{\theta\theta} \psi_\theta &= \nabla_{\theta\theta} b e^{i\tau g(\theta) r_t} + \nabla_\theta b \nabla_\theta g e^{i\tau g(\theta) r_t} + i\tau r_t \nabla_\theta \psi_\theta \nabla_\theta g + \psi_\theta i\tau r_t \nabla_{\theta\theta} g \\ &= \nabla_{\theta\theta} b e^{i\tau g(\theta) r_t} + \nabla_\theta b \nabla_\theta g e^{i\tau g(\theta) r_t} + i\tau r_t \nabla_\theta b \nabla_\theta g e^{i\tau g(\theta) r_t} \\ &\quad - (\tau r_t)^2 \nabla_\theta b (\nabla_\theta g)^2 \psi_\theta + \psi_\theta i\tau r_t \nabla_{\theta\theta} g \end{aligned}$$

Recall that  $|\psi_\theta| \leq 1$  and

$$b = \left(1 - \frac{i\tau}{c}\right)^{-\frac{2\gamma}{\sigma^2}}, \quad g = \frac{e^{-\kappa}}{1 - \frac{i\tau}{c}}.$$

Note that

$$\frac{1}{1 - \frac{i\tau}{c}} = \frac{1 + \frac{i\tau}{c}}{1 + \frac{\tau^2}{c^2}} \quad \text{and} \quad \left| \frac{1}{1 - \frac{i\tau}{c}} \right|^2 = \frac{1}{1 + \frac{\tau^2}{c^2}} \leq 1$$

and hence

$$|b| \leq 1.$$

$$\begin{aligned} \frac{\partial b}{\partial \gamma} &= -\frac{2}{\sigma^2} \ln \left(1 - \frac{i\tau}{c}\right) b, \\ \frac{\partial b}{\partial \kappa} &= -\frac{2\gamma}{\sigma^2} \left( \frac{i\tau}{c^2} \frac{\partial c}{\partial \kappa} \right) \frac{b}{1 - \frac{i\tau}{c}}, \\ \frac{\partial b}{\partial \sigma^2} &= \left\{ \frac{\frac{i\tau}{c}}{1 - \frac{i\tau}{c}} \frac{2\gamma}{\sigma^2} + \ln \left(1 - \frac{i\tau}{c}\right) \frac{2\gamma}{\sigma^4} \right\} b, \\ \frac{\partial g}{\partial \gamma} &= 0, \\ \frac{\partial g}{\partial \kappa} &= e^{-\kappa} \frac{\frac{i\tau}{c^2} \frac{\partial c}{\partial \kappa} - \left(1 - \frac{i\tau}{c}\right)}{\left(1 - \frac{i\tau}{c}\right)^2}, \\ \frac{\partial g}{\partial \sigma^2} &= e^{-\kappa} \frac{-\frac{i\tau}{c^2} \frac{\partial c}{\partial \sigma^2}}{\left(1 - \frac{i\tau}{c}\right)^2}. \end{aligned}$$

As  $r_t$  admit second moments,  $\tau$  admits fourth moments (that is  $\int \tau^4 \pi(\tau) d\tau < \infty$ ) and  $\Theta$  is compact bounded away from  $(0, 0, 0)$ , we have  $\|\nabla_\theta b\|$  and  $\|\nabla_\theta g\|$  bounded and the condition A8(ii) is satisfied. The expressions of the second derivatives are omitted here. A8(iii) requires the existence of the fourth moment of  $r_t$  and of the tenth moment of  $\tau$  which is true.

Table 1

Monte Carlo Comparison of Estimation Methods based on the CIR model of interest rates. We report three measures of estimation method performance – Mean Bias, Median Bias, and Root Mean Squared Error (RMSE) – for five different estimation methods: C-GMM with the optimal **DI** instrument (C-GMM-**DI**), CF-based estimator with Singleton’s approximation to the optimal **SI** instrument (GMM-**SI**), MLE, QMLE, and EMM (the results for the latter three methods are taken from Zhou, 2001). The simulations are performed based on the CIR model:

$$dr_t = (\gamma - \kappa r_t)dt + \sigma\sqrt{r_t}dW_t$$

with parameter values from Gallant and Tauchen (1998). All results are based on 1000 replications of samples with 500 observations. We use  $\alpha_T = 0.02$  and standard normal integrating density  $\pi(\tau)$  for C-GMM-**DI**;  $M = 6$  and  $\delta = 1$  for GMM-**SI**.

| True Value         | Mean Bias | Median Bias | RMSE   |
|--------------------|-----------|-------------|--------|
| <b>C-GMM-DI</b>    |           |             |        |
| $\gamma = 0.02491$ | 0.0090    | -0.0040     | 0.0374 |
| $\kappa = 0.00285$ | 0.0010    | -0.0004     | 0.0043 |
| $\sigma = 0.02750$ | 0.0064    | 0.0072      | 0.0130 |
| <b>GMM-SI</b>      |           |             |        |
| $\gamma = 0.02491$ | 0.0172    | 0.0235      | 0.0453 |
| $\kappa = 0.00285$ | 0.0013    | 0.0025      | 0.0079 |
| $\sigma = 0.02750$ | 0.0347    | 0.0257      | 0.0276 |
| <b>MLE</b>         |           |             |        |
| $\gamma = 0.02491$ | -0.0123   | -0.0119     | 0.0125 |
| $\kappa = 0.00285$ | -0.0014   | -0.0014     | 0.0014 |
| $\sigma = 0.02750$ | 0.0000    | 0.0000      | 0.0009 |
| <b>QMLE</b>        |           |             |        |
| $\gamma = 0.02491$ | 0.0994    | 0.0803      | 0.1343 |
| $\kappa = 0.00285$ | -0.0113   | -0.0091     | 0.0153 |
| $\sigma = 0.02750$ | 0.0000    | 0.0000      | 0.0009 |
| <b>EMM</b>         |           |             |        |
| $\gamma = 0.02491$ | 0.0451    | 0.0002      | 0.1252 |
| $\kappa = 0.00285$ | -0.0054   | 0.0000      | 0.0149 |
| $\sigma = 0.02750$ | -0.0015   | 0.0000      | 0.0076 |

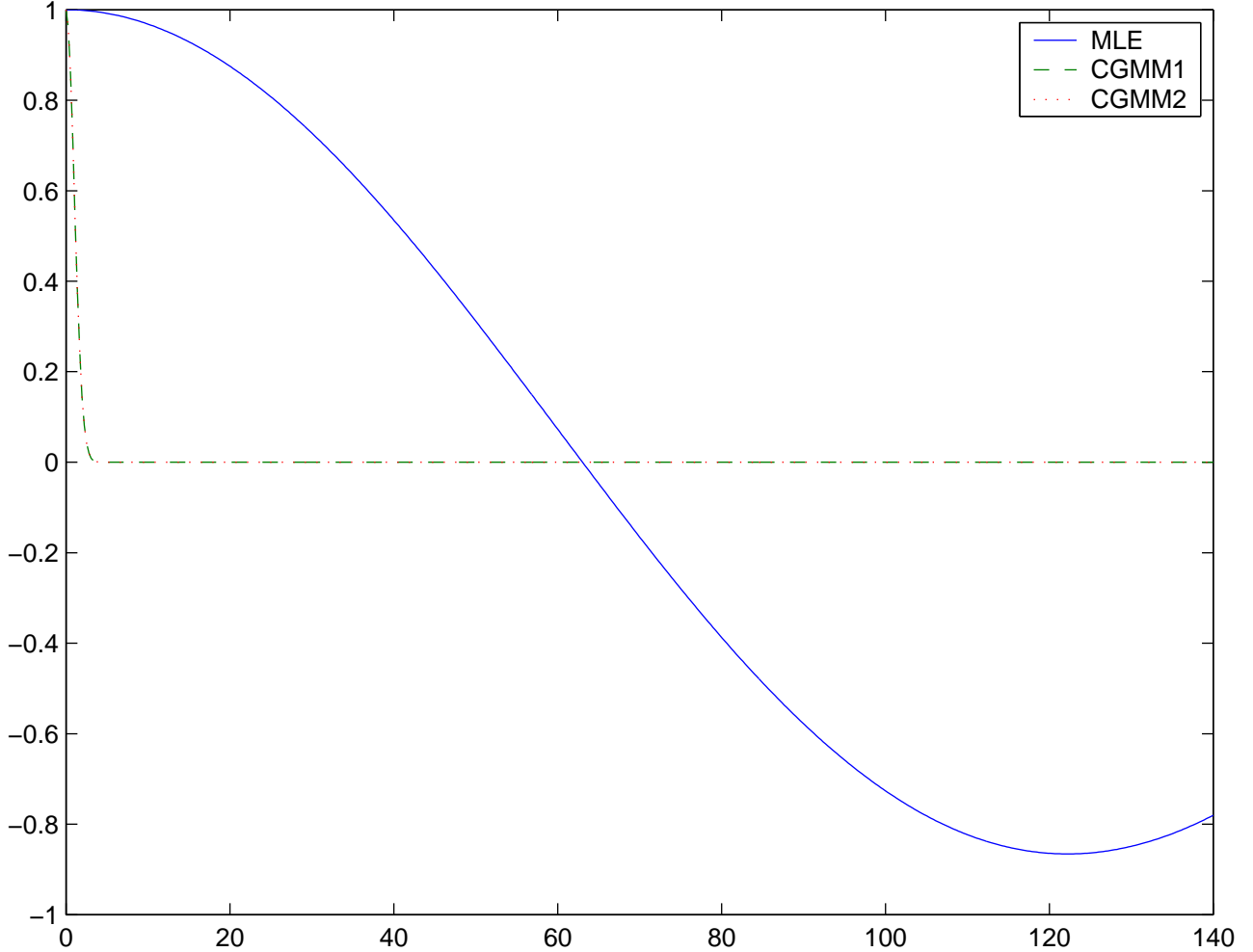


Fig. 1. Plot of real parts of integrands for computing the MLE and C-GMM estimators. We illustrate the degree of the numerical effort involved in computing the integrals necessary for the MLE estimation based on the Fourier inverse technique described in Singleton (2001) and the C-GMM estimation. The integrand is computed for the CIR model, studied in Section 6:

$$dr_t = \left( \begin{array}{c} \gamma \\ 0.02491 \end{array} - \begin{array}{c} \kappa \\ 0.00285 \end{array} r_t \right) dt + \begin{array}{c} \sigma \\ 0.0275 \end{array} \sqrt{r_t} dW_t$$

and evaluated at the point  $(r_{t+1}, r_t) = (\gamma/\kappa, 0.5 \cdot \gamma/\kappa) = (8.74, 4.37)$ . CGMM1 (CGMM2) denotes the integrand  $I_1$  in (22) ( $I_2$  in (23)).

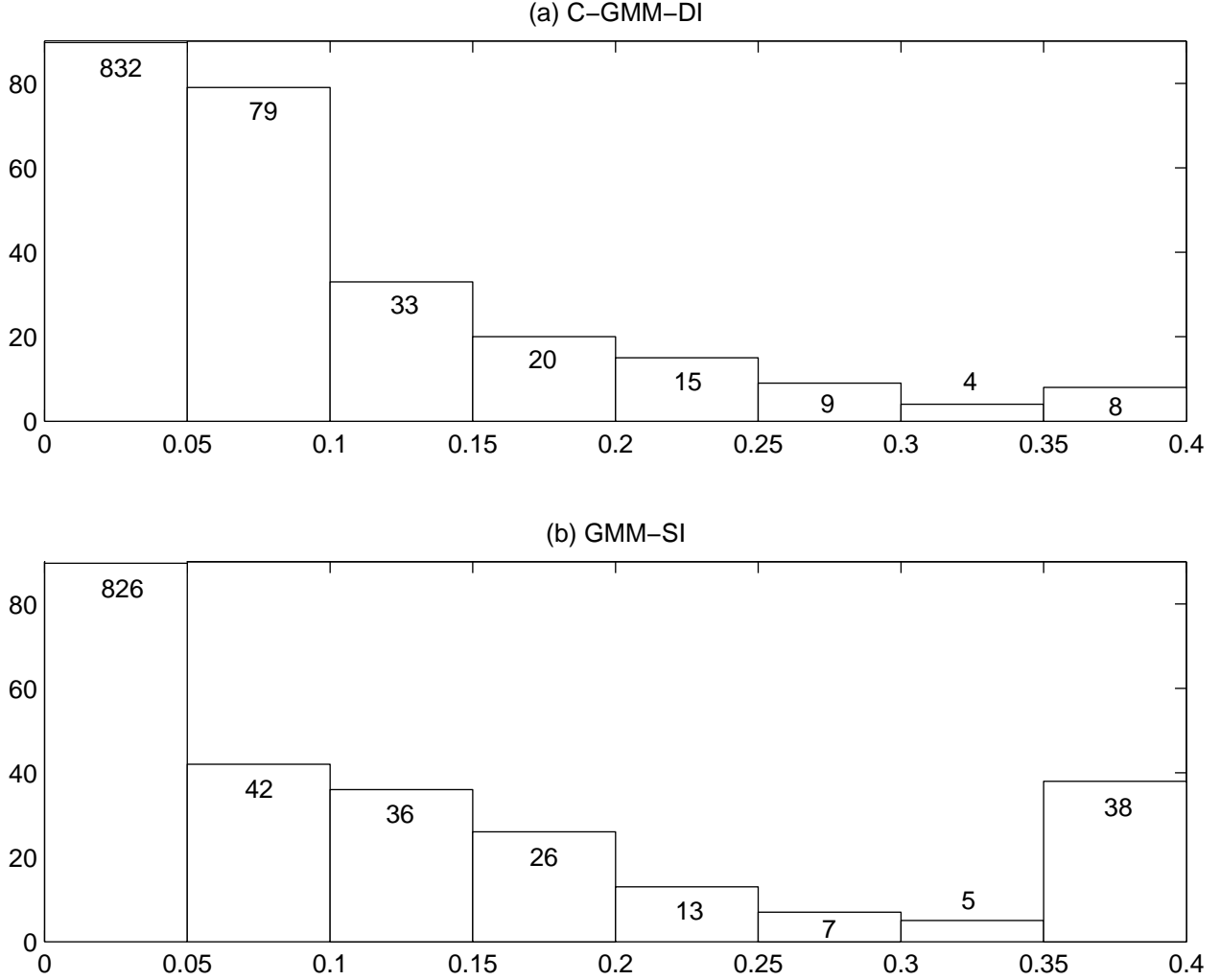


Fig. 2. Histogram of the estimation errors of the C-GMM-DI and GMM-SI estimators in the case of the CIR model. We compute, for each simulation path  $i$ , the the square root of the sum of squared errors across all three parameters,  $\sqrt{(\gamma - \hat{\gamma}_i)^2 + (\kappa - \hat{\kappa}_i)^2 + (\sigma - \hat{\sigma}_i)^2}$ . The plot shows the histogram of these errors. We cut off the first bin at 90, so that other bins could be seen clearly. The count corresponding to each error size category is reported in the corresponding bins.