Efficient Estimation Using the Characteristic Function

Marine Carrasco, Rachidi Kotchoni

Efficient Estimation using the Characteristic Function*

Marine Carrasco† University of Montreal

Rachidi Kotchoni‡ University of Montreal

First draft: January 2010 This version: July 2013

Abstract

The method of moments proposed by Carrasco and Florens (2000) permits to fully exploit the information contained in the characteristic function and yields an estimator which is asymptotically as efficient as the maximum likelihood estimator. However, this estimation procedure depends on a regularization or tuning parameter α that needs to be selected. The aim of the present paper is to provide a way to optimally choose α by minimizing the approximate mean square error (AMSE) of the estimator. Following an approach similar to that of Newey and Smith (2004), we derive a higher-order expansion of the estimator from which we characterize the finite sample dependence of the AMSE on α. We provide a data-driven procedure for selecting the regularization parameter that relies on parametric bootstrap. We show that this procedure delivers a root T consistent estimator of α. Moreover, the data-driven selection of the regularization parameter preserves the consistency, asymptotic normality and efficiency of the CGMM estimator. Simulation experiments based on a CIR model show the relevance of the proposed approach.

Keywords: Conditional moment restriction, Continuum of moment conditions, Generalized method of moments, Mean square error, Stochastic expansion, Tikhonov regularization. JEL Classification: C00, C13, C15

*An earlier version of this work was joint with Jean-Pierre Florens. We are grateful for his support. This paper has been presented at various conferences and seminars and we thank the participants for their comments, especially discussant Atsushi Inoue. Partial financial support from SSHRC is gratefully acknowledged.

†Université de Montreal; CIRANO; CIREQ. Fax (+1) 514 343 7221 E-mail: marine.carrasco@umontreal.ca

‡Corresponding author: Université de Montreal, Département de Sciences Economiques. E-mail: rachidi.kotchoni@umontreal.ca

1 Introduction

There is a one-to-one relationship between the characteristic function (henceforth, CF) and the probability distribution function of a random variable, the former being the Fourier transform of the latter. This implies that an inference procedure based on the empirical CF has the potential to be as efficient as another one that exploits the likelihood function. Paulson et al. (1975) used a weighted modulus of the difference between the theoretical CF and its empirical counterpart to estimate the parameters of the stable law. Feuerverger and Mureika (1977) studied the convergence properties of the empirical CF and suggested that "*it may be a useful tool in numerous statistical problems*". Since then, many interesting applications have been proposed, including Feuerverger and McDunnough (1981*b,c*), Koutrouvelis (1980), Carrasco and Florens (2000), Chacko and Viceira (2003) and Carrasco, Chernov, Florens, and Ghysels (2007) (henceforth, CCFG (2007)). For a quite comprehensive review of empirical CF-based estimation methods, see Yu (2004).

The CF provides a good alternative to econometricians when the likelihood function is not available in closed form. For example, some distributions in the α-stable family are naturally specified via their CFs while their densities are known in closed form only at isolated points of the parameter space (see e.g. Nolan, 2009). The density of the Variance-Gamma model of Madan and Seneta (1990) has an integral representation whereas its CF has a simple closed form expression. The transition density of a discretely sampled continuous time process is not available in closed form, except when its parameterization coincides with that of a square-root diffusion (Singleton, 2001). Even in this special case, the transition density takes the form of an infinite mixture of Gamma densities with Poisson weights. The same type of density arises in the autoregressive Gamma model studied in Gouriéroux and Jasiak (2005). Ait-Sahalia and Kimmel (2006) propose closed form approximations for the log-likelihood function of various continuous-time stochastic volatility models. But their method cannot be applied to other situations without solving a complicated Kolmogorov forward and backward equation. Interestingly, the conditional CF can be derived in closed form for all continuous-time stochastic volatility models.

The CF, $\varphi(\tau,\theta)$, of a random vector $x_t \in \mathbb{R}^p$ $(t = 1,...,T)$ is nothing but the expectation of $e^{i\tau' x_t}$ with respect the distribution of x_t, where θ is the parameter that characterizes the distribution of x_t, $\tau \in \mathbb{R}^p$ is the Fourier index and i is the imaginary number such that $i^2 = -1$. Hence, a candidate moment condition for the estimation of θ_0 (i.e., the true value of θ) is given by $h_t(\tau,\theta) = e^{i\tau' x_t} - E(e^{i\tau' x_t})$. This moment condition is valid for all $\tau \in \mathbb{R}^p$ and hence, $h_t(\tau,\theta)$ is a moment function or a continuum of moment conditions. Feuerverger and McDunnough (1981b) propose an estimation procedure that consists of minimizing a norm of the sample average of the moment function. Their objective function involves an optimal weighting function that depends on the true unknown likelihood function. Feuerverger and McDunnough (1981c) apply the Generalized Method of Moments (GMM) to a discrete set of moment conditions obtained by restricting the continuous index $\tau \in \mathbb{R}^p$ to a discrete grid $\tau \in (\tau_1, \tau_2, ...\tau_N)$. They show that the asymptotic variance of the resulting estimator can be made arbitrarily close to the Cramer-Rao bound by selecting the grid for τ sufficiently fine and extended. Similar discretization approaches are used in Singleton (2001) and Chacko and Viceira

(2003). However, the number of points in the grid for τ must not be larger than the sample size for the covariance matrix of the discrete set of moment conditions to be invertible. In particular, the first order optimality conditions associated with the discrete GMM procedure becomes ill-posed as soon as the grid $(\tau_1, \tau_2, ...\tau_N)$ is too refined or too extended. Intuitively, the discrete set of moment conditions $\{h_t(\tau_i, \theta)\}_{i=1}^N$ converges to the moment function $\tau \mapsto h_t(\tau, \theta), \tau \in \mathbb{R}$ as this grid is refined and extended. As a result, it is necessary to apply operator methods in a suitable Hilbert space to be able to handle the estimation procedure at the limit.

Carrasco and Florens (2000) proposed a Continuum GMM (henceforth, CGMM) that permits to efficiently use the whole continuum of moment conditions. Similarly to the classical GMM, the CGMM is a two-step procedure that delivers a consistent estimator at the first step and an efficient estimator at the second step. The ideal (unfeasible) objective function of the second step CGMM is a quadratic form in an suitably defined Hilbert space with metrics K^{-1}, where K is the asymptotic covariance operator associated with the moment function $h_t(\tau, \theta)$. To obtain a feasible efficient CGMM estimator, one replaces the operator K by an estimator K_T obtained from a finite sample. However, the latter empirical operator is degenerate and not invertible while its theoretical counterpart is invertible only on a dense subset of the reference space. To circumvent these difficulties, Carrasco and Florens (2000) resorted to a Tikhonov-type regularized inverse of K_T, e.g. $K_{\alpha T} = \left(K_T^2 + \alpha I\right)^{-1} K_T$, where I is the identity operator and α is a regularization parameter. The CGMM estimator is root-T consistent and asymptotically normal for any fixed and reasonably small value of α. However, asymptotic efficiency is obtained only by letting $\alpha T^{1/2}$ go to infinity and α go to zero as T goes to infinity.

The main objective of this paper is to characterize the optimal rate of convergence for α as T goes to infinity. To this end, we derive a Nagar (1959) type stochastic expansion of the CGMM estimator. This type of expansion has been used in Newey and Smith (2004) to study the higher order properties of empirical likelihood estimators. We use our expansion to find the convergence rates of the higher order terms of the MSE of the CGMM estimator. These rates depend on both α and T. We find that the higher order bias of the CGMM estimator is dominated by two higher order variance terms. By equating the rates of these dominant term, we find an expression of the form $\alpha_T = c(\theta_0) T^{-g(\beta)}$, where $c(\theta_0)$ does not depend on T and $g(\beta)$ inherits some properties from the covariance operator K. To implement the optimal selection of α empirically, we advocate a naive estimator of α_T obtained by minimizing an approximate MSE of the CGMM estimator obtained by parametric bootstrap. Even though the CGMM estimator is consistent, there is a concern that its variance be infinite in finite sample for certain data generating processes. This concern seems unfounded for the CIR model on which our Monte Carlo simulations are based. If applicable, this difficulty is avoided by truncating the AMSE similarly as in Andrews (1991).

The remainder of the paper is organized as follows. In Section 2, we review the properties of the CGMM estimator in IID and Markov cases. In Section 3, we derive a higher-order expansion for the MSE of the CGMM estimator and use this expansion to obtain the optimal rate of convergence for the regularization parameter α_T. In Section 4, we describe a simulation-based method to estimate α_T and show the consistency of the resulting estimator. Section 5 presents a simulation study based on the

CIR term structure model and Section 6 concludes. The proofs are collected in appendix.

2 Overview of the CGMM

This section is essentially a summary of known results about the CGMM estimator. The first subsection present a general framework for implementing the CF-based CGMM procedure whilst the second subsection presents the basic properties of the resulting estimator.

2.1 The CGMM Based on Characteristic function

Let $x_t \in \mathbb{R}^p$ be a random vector process whose distribution is indexed by a finite dimensional parameter θ with true value θ_0. When the process x_t is IID, Carrasco and Florens (2000) propose to estimate θ_0 based on the moment function given by:

$$h_t(\tau, \theta) = e^{i\tau' x_{t+1}} - \varphi(\tau, \theta), \tag{1}$$

where $\varphi(\tau, \theta) = E^\theta \left(e^{i\tau' x_{t+1}} \right)$ is the CF of x_t and E^θ is the expectation operator with respect to the data generating process indexed by θ.

CCFG (2007) extend the scope of the CGMM procedure to Markov and weakly dependent models. In this paper, we restrict our attention to IID and Markov cases. The moment function used in CCFG (2007) for the Markov case is:

$$h_t(\tau, \theta) = \left(e^{is' x_{t+1}} - \varphi_t(s, \theta) \right) e^{ir' x_t}. \tag{2}$$

where $\varphi_t(s, \theta) = E^\theta(e^{is' x_{t+1}} | x_t)$ is the conditional CF of x_t and $\tau = (s, r) \in \mathbb{R}^{2p}$. In equation (2), the set of basis functions $\{e^{ir' x_t}\}$ is being used as instruments. CCFG (2007) show that these instruments are optimal given the Markovian structure of the model. Moment conditions defined by (1) are IID whereas equation (2) describes a martingale difference sequence.

Note that a standard conditional moment restriction (i.e., non CF-based) can be converted into a continuum of moment unconditional moment restrictions featuring (2). In this case, the CGMM estimator may be viewed as an alternative to the estimator proposed by Dominguez and Lobato (2004) and to the smooth minimum distance estimator of Lavergne and Patilea (2008). Subsequently, we use the generic notation $h_t(\tau, \theta)$, $\tau \in \mathbb{R}^d$ to denote a moment function defined by either (1) or (2), where $d = p$ for (1) and $d = 2p$ for (2).

Let π be a probability distribution function on $\mathbb{R}^d$ and $L^2(\pi)$ be the Hilbert space of complex valued functions that are square integrable with respect to π, i.e.:

$$\mathbf{L}^2(\pi) = \{ f : \mathbb{R}^d \to \mathbf{C} | \int f(\tau) \overline{f(\tau)} \pi(\tau) d\tau < \infty \}, \tag{3}$$

where $\overline{f(\tau)}$ denotes the complex conjugate of $f(\tau)$. As $|h_t(., \theta)|^2 \leq 2$ for all $\theta \in \Theta$, the function $h_t(., \theta)$

belongs to $L^2(\pi)$ for all $\theta \in \Theta$ and for any finite measure π. Hence, we consider the following scalar product on $\mathbf{L}^2(\pi) \times \mathbf{L}^2(\pi)$:

$$\langle f, g \rangle = \int f(\tau)\overline{g(\tau)}\pi(\tau)d\tau. \tag{4}$$

Based on this notation, the efficient CGMM estimator is given by

$$\widehat{\theta} = \arg\min_{\theta} \left\langle K^{-1}\widehat{h}_T(.,\theta), \widehat{h}_T(.,\theta) \right\rangle.$$

where K is the asymptotic covariance operator associated with the moment conditions. K is an integral operator and satisfies:

$$Kf(\tau_1) = \int_{-\infty}^{\infty} k(\tau_1,\tau) f(\tau)\, \pi(\tau)\, d\tau, \text{ for any } f \in L^2(\pi), \tag{5}$$

where $k(\tau_1, \tau_2)$ is the kernel given by:

$$k(\tau_1,\tau_2) = E\left(h_t(\tau_1,\theta)\overline{h_t(\tau_2,\theta)}\right). \tag{6}$$

Some basic properties of the operator K are discussed in Appendix A.

With a sample of size T and a consistent first step estimator $\widehat{\theta}^1$ in hand, one estimates $k(\tau_1, \tau_2)$ by:

$$k_T(\tau_1,\tau_2,\widehat{\theta}^1) = \frac{1}{T}\sum_{t=1}^{T} h_t(\tau_1,\widehat{\theta}^1)\overline{h_t(\tau_2,\widehat{\theta}^1)}. \tag{7}$$

In the specific case of IID data, an estimator of the kernel that does not use a first step estimator is given by:

$$k_T(\tau_1,\tau_2) = \frac{1}{T}\sum_{t=1}^{T}\left(e^{i\tau_1' x_t} - \widehat{\varphi}_T(\tau_1)\right)\overline{\left(e^{i\tau_2' x_t} - \widehat{\varphi}_T(\tau_2)\right)}. \tag{8}$$

where $\widehat{\varphi}_T(\tau_1) = \frac{1}{T}\sum_{t=1}^{T} e^{i\tau_1' x_t}$. Unfortunately, an empirical covariance operator K_T with kernel function given by either (7) or (8) is degenerate and not invertible. Indeed, the inversion of K_T raises a problem similar to one of the Fourier inversion of an empirical characteristic function. This problem is worsened by the fact that the inverse of K which K_T is aimed at estimating exists only on a dense subset of $L^2(\pi)$. Moreover, when $K^{-1}f = g$ exists for a given function f, a small perturbation in f may give rise to a large variation in g.

To circumvent these difficulties, we consider estimating K^{-1} by:

$$K_{\alpha T}^{-1} = \left(K_T^2 + \alpha I\right)^{-1} K_T,$$

where the hyperparameter α plays two roles. First, it is a smoothing parameter as it allows $K_{\alpha T}^{-1} f$ to exist for all f in $L^2(\pi)$. Second, it is a regularization parameter as it dampens the sensitivity of $K_{\alpha T}^{-1} f$ to perturbations in the input f. For any function f in the range of K and any consistent estimator

$\widehat{f}_T$ of f, $K_{\alpha T}^{-1}\widehat{f}_T$ converges to $K^{-1}f$ as T goes to infinity and α goes to zero at appropriate rate. The expression for $K_{\alpha T}^{-1}$ uses a Tikhonov regularization, also called ridge regularization. Other forms of regularization could have been used, see e.g. Carrasco, Florens and Renault (2007).

The feasible CGMM estimator is given by:

$$\widehat{\theta}_T(\alpha) = \arg\min_{\theta} \widehat{Q}_T(\alpha, \theta), \tag{9}$$

where $\widehat{Q}_T(\alpha, \theta) = \left\langle K_{\alpha T}^{-1}\widehat{h}_T(., \theta), \widehat{h}_T(., \theta) \right\rangle$. An expression of the objective function $\widehat{Q}_T(\alpha, \theta)$ in matrix form is given in CCFG (2007a, Section 3.3). An alternative expression and a numerical algorithm for the numerical evaluation of this objective function based on Gauss-Hermite quadratures is described in Appendix D.

2.2 Consistency and Asymptotic Normality

In order to study the properties of a CGMM estimator obtained within the framework described previously, the following assumptions are posited:

Assumption 1: The probability density function π is strictly positive on $\mathbb{R}^d$ and admits all its moments.

Assumption 2: 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 θ_0 which is an interior point of a compact set Θ.

Assumption 3: $h_t(\tau, \theta)$ is three times continuously differentiable with respect to θ. Furthermore, the first two derivatives satisfy:

$$Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial h_t(\tau, \theta)}{\partial \theta_j}\right) < \infty \text{ and } Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial^2 h_t(\tau, \theta)}{\partial \theta_j \partial \theta_k}\right) < \infty,$$

for all j, k and T.

Assumption 4: $E^{\theta_0}(h_T(., \theta)) \in \Phi_\beta$ for all $\theta \in \Theta$ and for some $\beta \geq 1$, and the first two derivatives of $E^{\theta_0}(h_T(., \theta))$ w.r.t. θ belong to Φ_β for all θ in a neighborhood of θ_0 and for the same β as previously, where:

$$\Phi_\beta = \left\{ f \in L^2(\pi) \text{ such that } \left\| K^{-\beta} f \right\| < \infty \right\} \tag{10}$$

Assumption 5: The random variable x_t is stationary Markov and satisfies $x_t = r(x_{t-1}, \theta_0, \varepsilon_t)$ where $r(x_{t-1}, \theta_0, \varepsilon_t)$ is three times continuously differentiable with respect to θ_0 and ε_t is a IID white noise whose distribution is known and does not depend on θ_0.

Assumption 1 and 2 are quite standard and they have been used in Carrasco and Florens (2000). The first part of Assumption 3 ensures some smoothness properties for $\widehat{\theta}_T(\alpha)$ while the second part is always satisfied for IID models. The largest real β such that $f \in \Phi_\beta$ in Assumption 4 may be

called the level of regularity of f with respect to K: the larger β is, the better f is approximated by a linear combination of the eigenfunctions of K associated with the highest eigenvalues. Because $Kf(.)$ involve a d-dimensional integration, β may be affected by both the dimensionality of the index τ and the smoothness of f. CCFG (2007) have shown that we always have $\beta \geq 1$ if $f = E^{\theta_0}(h_t(\tau, \theta))$. Assumption 5 implies that the data can be simulated upon knowing how to draw from the distribution of ε_t. It is satisfied for all random variables that can be written as a location parameter plus a scale parameter time a standardized representative of the family of distribution. Examples include the exponential family and the stable distribution. The IID case is a special case of Assumption 5 where $r(x_{t-1}, \theta_0, \varepsilon_t)$ takes the simpler form $r(\theta_0, \varepsilon_t)$. Further discussions on this type of model can be found in Gourieroux, Monfort, and Renault (1993) in the indirect inference context. Note that the function $r(x_{t-1}, \theta_0, \varepsilon_t)$ may not be available in analytical form. In particular, the relation $x_t = r(x_{t-1}, \theta_0, \varepsilon_t)$ can be the numerical solution of a general equilibrium asset pricing model (e.g., as in Duffie and Singleton, 1993).

We have the following results:

Theorem 1 *Under Assumptions 1 to 5, the CGMM estimator is consistent and satisfies:*

$$T^{1/2}\left(\widehat{\theta}_T(\alpha) - \theta_0\right) \xrightarrow{L} N(0, I_{\theta_0}^{-1}).$$

as T and $\alpha T^{1/2}$ go to infinity and α goes to zero, where $I_{\theta_0}^{-1}$ denotes the inverse of the Fisher Information Matrix.

See Proposition 3.2 of CCFG (2007) for a more general statement of the consistency and asymptotic normality result. A nice feature about the CGMM estimator is that its asymptotic distribution does not depend on the probability density function π.

3 Stochastic expansion of the CGMM estimator

The conditions required for the asymptotic efficiency result stated by Theorem 1 allow for a wide range of convergence rates for α. Indeed, any sequence of type $\alpha_T = cT^{-a}$ (with $c > 0$) satisfies these conditions as soon as $0 < a < 1/2$. Among the admissible convergence rates, we would like to find the one that minimizes the mean square error of the CGMM estimator for a given sample size T. To achieve this, we consider deriving the stochastic expansion of the CGMM estimator. The higher order properties of GMM-type estimators have been studied by Rothenberg (1983, 1984), Koenker et al. (1994), Rilstone et al. (1996) and Newey and Smith (2004). For estimators derived in the linear simultaneous equation framework, examples include Nagar (1959), Buse (1992) and Donald and Newey (2001). The approach followed here is similar to Nagar (1959) and Newey and Smith (2004), which tries to approximate the MSE of an estimator analytically based on the leading terms of its stochastic expansion.

Two difficulties arise when analyzing the terms of the expansion of the CGMM estimator. First, when the rate of α as a function of T is unknown, it is not always possible to write the terms of the

expansion in decreasing order. The second difficulty stems from a result that dramatically differs from the case with a finite number of moment conditions. Indeed, when the number of moment conditions is finite, the quadratic form $T\widehat{h}_T(\theta_0)' K^{-1}\widehat{h}_T(\theta_0)$ is $O_p(1)$ and follows asymptotically a chi-square distribution with degrees of freedom given by the number of moment conditions. However, the analogue of the previous quadratic form, $\left\|K^{-1/2}\sqrt{T}\widehat{h}_T(\theta_0)\right\|^2$, is not well defined in the presence of a continuum of moment conditions. Its regularized version, $\left\|K_\alpha^{-1/2}\sqrt{T}\widehat{h}_T(\theta_0)\right\|^2$, exists but diverges as T goes to infinity and α goes to zero. Indeed, we have

$$\begin{aligned}\left\|K_\alpha^{-1/2}\sqrt{T}\widehat{h}_T(\theta_0)\right\| &\leq \underbrace{\left\|\left(K^2+\alpha I\right)^{-1/4}\right\|}_{\leq \alpha^{-1/4}}\underbrace{\left\|\left(K^2+\alpha I\right)^{-1/4}K^{1/2}\right\|}_{\leq 1}\underbrace{\left\|\sqrt{T}\widehat{h}_T(\theta_0)\right\|}_{=O_p(1)} \\ &= O_p\left(\alpha^{-1/4}\right).\end{aligned} \tag{11}$$

The expansion that we derive for $\widehat{\theta}_T(\alpha)-\theta_0$ is of the same form for both the IID and Markov cases. Namely:

$$\widehat{\theta}_T(\alpha)-\theta_0 = \Delta_1+\Delta_2+\Delta_3+o_p\left(\alpha^{-1}T^{-1}\right)+o_p\left(\alpha^{\min(1,\frac{2\beta-1}{2})}T^{-1/2}\right) \tag{12}$$

where $\Delta_1 = O_p\left(T^{-1/2}\right)$, $\Delta_2 = O_p\left(\alpha^{\min(1,\frac{2\beta-1}{2})}T^{-1/2}\right)$ and $\Delta_3 = O_p\left(\alpha^{-1}T^{-1}\right)$. Appendix B provides details about the above expansion whose validity is ensured by the consistency result of Theorem 1. In deriving the expansion above, we wish to find the rate of convergence of the α which minimizes the leading terms of the MSE:

$$MSE(\alpha,\theta_0) = TE\left[T\left(\widehat{\theta}_T(\alpha)-\theta_0\right)\left(\widehat{\theta}_T(\alpha)-\theta_0\right)'\right] \tag{13}$$

We have the following results on the higher order MSE matrix and on the optimal convergence rate for the regularization parameter.

Theorem 2 *Assume that Assumptions 1 to 5 hold. Then we have:*

(i) The approximate MSE matrix of $\widehat{\theta}_T(\alpha)$ up to order $O\left(\alpha^{-1}T^{-1/2}\right)$ (henceforth, AMSE) is decomposed as the sum of the squared bias and variance:

$$AMSE(\alpha,\theta_0) = TBias*Bias' + TVar$$

where

$$\begin{aligned}TBias*Bias' &= O\left(\alpha^{-2}T^{-1}\right), \\ TVar &= I_{\theta_0}^{-1}+O\left(\alpha^{\min\left(2,\frac{2\beta-1}{2}\right)}\right)+O\left(\alpha^{-1}T^{-1/2}\right).\end{aligned}$$

as $T\to\infty$, $\alpha^2T\to\infty$ and $\alpha\to 0$.

(ii) The α that minimizes the trace of $AMSE(\alpha, \theta_0)$, denoted $\alpha_T \equiv \alpha_T(\theta_0)$, satisfies:

$$\alpha_T = O\left(T^{-\max(\frac{1}{6}, \frac{1}{2\beta+1})}\right).$$

Remarks.

1. We have the usual trade-off between a term that is decreasing in α and another that is increasing in α. Interestingly, the squared bias term is dominated by two higher order variance terms whose rates are equated to obtain the optimal rate for the regularization parameter. The same situation happens for the Limited Information Maximum Likelihood estimator for which the bias is also dominated by variance terms (see Donald and Newey, 2001).

2. The rate for the $O\left(\alpha^{\min\left(2, \frac{2\beta-1}{2}\right)}\right)$ variance term does not improve for $\beta > 2.5$. This is due to a property of Tikhonov regularization that is well documented in the literature on inverse problems, see e.g. Carrasco, Florens and Renault (2007). The use of another regularization such as spectral cut-off or Landweber-Fridman would permit to improve the rate of convergence for large values of β. However, this improvement comes at the cost of a greater complexity in the proofs (e.g. in the spectral cut-off, we lose the differentiability of the estimator with respect to α).

3. Our expansion is consistent with the condition of Theorem 1, since the optimal regularization parameter α_T satisfies $\alpha_T^2 T \to \infty$.

4. It follows from Theorem 2 that the optimal regularization parameter α_T is necessarily of the form:

$$\alpha_T = c(\theta_0) T^{-g(\beta)}, \tag{14}$$

for some positive function $c(\theta_0)$ that does not depend on T and a positive function $g(\beta)$ that satisfies $\max\left(\frac{1}{6}, \frac{1}{2\beta+1}\right) \leq g(\beta) < 1/2$. An expression of the form (14) is often used as starting point for optimal bandwidth selection in nonparametric density estimation. Examples in the semiparametric context include Linton (2002) and Jacho-Chavez (2010).

4 Estimation of the Optimal Regularization parameter

Our purpose is to select the regularization parameter α so as to minimize the trace of the MSE matrix of $\widehat{\theta}_T(\alpha)$ for a given sample of size T, i.e.:

$$\alpha_T(\theta_0) = \arg \min_{\alpha \in [0,1]} \Sigma_T(\alpha, \theta_0),$$

where $\Sigma_T(\alpha, \theta_0) = TE\left(\left\|\widehat{\theta}_T(\alpha) - \theta_0\right\|^2\right)$. This raises at least three problems. First, the MSE $\Sigma_T(\alpha, \theta_0)$ might be infinite in finite samples even though $\widehat{\theta}_T(\alpha)$ is consistent.[1] Second, the true parameter value θ_0 is unknown. Third, the finite sample distribution of $\widehat{\theta}_T(\alpha) - \theta_0$ is not known even

[1] This is due to the fact that $\widehat{\theta}_T(\alpha)$ is a GMM-type estimator. The large sample properties of such estimators are well-known whilst their finite sample properties can be established only a special cases.

when θ_0 is known. Each of these problems is examined below.

The variance of $\widehat{\theta}_T(\alpha)$ may be infinite for some data generating processes. To hedge against such situations, one may consider a truncated MSE of the form:

$$\Sigma_T(\alpha, \theta_0, \nu) = TE\left[\xi_T(\alpha, \theta_0) \left| \xi_T(\alpha, \theta_0) < n_\nu \right.\right], \tag{15}$$

where $\xi_T(\alpha, \theta) \equiv \left\| \widehat{\theta}_T(\alpha) - \theta \right\|^2$ and n_ν satisfies $\nu = \Pr\left(\xi_T(\alpha, \theta_0) > n_\nu\right)$. A similar approach has been used in Andrews (1991, p. 826). Given that the finite sample distribution of $\widehat{\theta}_T(\alpha)$ is unknown in practice, it is convenient to first select the probability of truncation ν (e.g., $\nu = 1\%$) and then deduce the corresponding quantile n_ν by simulation. To account for the possibility of the pair (ν, n_ν) depending on α, one may consider instead:

$$\Sigma_T(\alpha, \theta_0, \nu) = (1-\nu) TE\left[\xi_T(\alpha, \theta_0) \left| \xi_T(\alpha, \theta_0) < n_\nu \right.\right] + \nu n_\nu T, \tag{16}$$

which accounts for the probability mass at the truncation boundary. Note that the truncation will play no role if the MSE of $\widehat{\theta}_T(\alpha)$ is finite. In this case, we simply let:

$$\Sigma_T(\alpha, \theta_0, 0) \equiv \Sigma_T(\alpha, \theta_0) = TE\left[\xi_T(\alpha, \theta_0)\right]. \tag{17}$$

As $\widehat{\theta}_T(\alpha)$ is asymptotically normal, its second moment exists for large enough T. Hence, the truncation disappears (i.e., each of the expressions (15) and (16) converges to (17)) if one let ν go to zero as T goes to infinity.[2]

We define the optimal regularization parameter as:

$$\alpha_T(\theta_0) = \arg \min_{\alpha \in [0,1]} \Sigma_T(\alpha, \theta_0, \nu), \tag{18}$$

where $\Sigma_T(\alpha, \theta_0, \nu)$ is given by either (15), (16) or (17).

Our strategy for estimating $\alpha_T(\theta_0)$ relies on approximating the unknown MSE by parametric bootstrap. Let $\widehat{\theta}_T^1$ be the CGMM estimator of θ_0 obtained by replacing the covariance operator with the identity operator. This estimator is consistent and asymptotically normal albeit inefficient. We use $\widehat{\theta}_T^1$ to simulate M independent samples of size T, denoted $X_T^{(j)}(\widehat{\theta}^1)$ for $j = 1, 2, ..., M$. It should be emphasized that we have adopted a fully parametric approach from the beginning by assuming that the model of interest is fully specified. Indeed, it would not be possible to obtain MLE efficiency otherwise. The model can be simulated by exploiting Assumption 5, which stipulates that the data generating process satisfies $x_t = r(x_{t-1}, \theta, \varepsilon_t)$. To start with, one first generates MT IID draws $\varepsilon_t^{(j)}$ (for $j = 1, ..., M$ and $t = 1, ..., T$) from the known distribution of the errors. Next, M time-series of size T are obtained by applying the recursion $x_t^{(j)} = r\left(x_{t-1}^{(j)}, \widehat{\theta}_T^1, \varepsilon_t^{(j)}\right)$, $t = 1, ..., T$, from M arbitrary starting values $x_0^{(j)}$.

Using the simulated samples, one computes M IID copies of the CGMM estimator for any given

[2] As $n_\nu \to \infty$ as $\nu \to 0$, one might be concerned by the limiting behavior of νn_ν as $\nu \to 0$ when $\Sigma_T(\alpha, \theta_0)$ is infinite. However, this is not an issue as long as n_ν is finite for all finite T.

α. We let $\widehat{\theta}_T^j(\alpha, \widehat{\theta}_T^1)$ denote the CGMM estimator computed from the j^{th} sample. The truncated MSE given by (15) is estimated by:

$$\widehat{\Sigma}_{TM}\left(\alpha, \widehat{\theta}_T^1, \nu\right) = \frac{T}{(1-\nu)\,M} \sum_{j=1}^{M} \xi_{j,T}(\alpha, \widehat{\theta}_T^1) 1\left(\xi_{j,T}(\alpha, \widehat{\theta}_T^1) \leq \widehat{n}_\nu\right), \tag{19}$$

where $\xi_{j,T}(\alpha, \widehat{\theta}_T^1) = \left\|\widehat{\theta}_T^j(\alpha, \widehat{\theta}_T^1) - \widehat{\theta}_T^1\right\|^2$, ν is a probability selected by the econometrician and $\widehat{n}_\nu$ satisfies:

$$\frac{1}{M} \sum_{j=1}^{M} 1\left(\xi_{j,T}(\alpha, \widehat{\theta}_T^1) \leq \widehat{n}_\nu\right) = 1 - \nu,$$

The truncated MSE based on the alternative Formula (16) is estimated by:

$$\widehat{\Sigma}_{TM}\left(\alpha, \widehat{\theta}_T^1, \nu\right) = \frac{T}{M} \sum_{j=1}^{M} \xi_{j,T}(\alpha, \widehat{\theta}_T^1) 1\left(\xi_{j,T}(\alpha, \widehat{\theta}_T^1) \leq \widehat{n}_\nu\right) + \nu \widehat{n}_\nu T. \tag{20}$$

With no truncation, (21) and (19) are identical to the naive MSE estimator given by:

$$\widehat{\Sigma}_{TM}\left(\alpha, \widehat{\theta}_T^1\right) = \frac{T}{M} \sum_{j=1}^{M} \xi_{j,T}(\alpha, \widehat{\theta}_T^1), \tag{21}$$

which is aimed at estimating (17). Finally, we select the optimal regularization parameter according to:

$$\widehat{\alpha}_{TM}\left(\widehat{\theta}^1\right) = \underset{\alpha \in [0,1]}{\arg\min} \widehat{\Sigma}_{TM}\left(\alpha, \widehat{\theta}^1, \nu\right), \tag{22}$$

where $\widehat{\Sigma}_{TM}\left(\alpha, \widehat{\theta}^1, \nu\right)$ is either (19), (20) or (21).

Let $\Sigma_T\left(\alpha, \widehat{\theta}^1, \nu\right)$ be the limit of $\widehat{\Sigma}_{TM}\left(\alpha, \widehat{\theta}^1, \nu\right)$ as the number of replications M goes to infinity and define:

$$\alpha_T\left(\widehat{\theta}^1\right) = \underset{\alpha \in [0,1]}{\arg\min} \Sigma_T\left(\alpha, \widehat{\theta}^1, \nu\right).$$

Note that $\alpha_T\left(\widehat{\theta}^1\right)$ is a deterministic function of a stochastic argument while $\widehat{\alpha}_{TM}\left(\widehat{\theta}^1\right)$ is doubly random, being a stochastic function of a stochastic argument. The estimator $\alpha_T\left(\widehat{\theta}^1\right)$ is not feasible. However, its properties are the key ingredients for establishing the consistency of its feasible counterpart $\widehat{\alpha}_{TM}\left(\widehat{\theta}^1\right)$. To pursue, we need the following assumption:

Assumption 6: The regularization parameter α that minimizes (the possibly truncated criterion) $\Sigma_T(\alpha, \theta_0, \nu)$ is of the form $\alpha_T(\theta_0) = c(\theta_0) T^{-g(\beta)}$, for some continuous positive function $c(\theta_0)$ that does not depend on T and a positive function $g(\beta)$ that satisfies $\max\left(\frac{1}{6}, \frac{1}{2\beta+1}\right) \leq g(\beta) < 1/2$.

Basically, Assumption 6 requires that the optimal rate found for the regularization parameter at (14)

be insensitive to the MSE truncation scheme. This assumption ensures that $\alpha_T\left(\widehat{\theta}^1\right) = c\left(\widehat{\theta}^1\right) T^{-g(\beta)}$ and is necessarily satisfied as T goes to infinity and ν goes to zero. The following result can further be proved.

Theorem 3 *Let $\widehat{\theta}^1$ be a $\sqrt{T}-$consistent estimator of θ_0. Then under Assumptions 1 to 5, $\frac{\alpha_T(\widehat{\theta}^1)}{\alpha_T(\theta_0)} - 1$ converges in probability to zero as T goes to infinity.*

In Theorem 3, the function $\alpha_T(.)$ is deterministic and continuous but the argument $\widehat{\theta}^1$ is stochastic. As T goes to infinity, $\widehat{\theta}^1$ gets closer and closer to θ_0, but at the same time $\alpha_T(\theta_0)$ converges to zero at some rate that depends on T. This prevents us from claiming without caution that $\frac{\alpha_T(\widehat{\theta}^1)}{\alpha_T(\theta_0)} - 1 = o_p(1)$ since the denominator is not bounded away from zero. The next theorem characterizes the rate of convergence of $\frac{\widehat{\alpha}_{TM}(\theta_0)}{\alpha_T(\theta_0)}$.

Theorem 4 *Under assumptions 1 to 5, $\frac{\widehat{\alpha}_{TM}(\theta_0)}{\alpha_T(\theta_0)} - 1$ converges in probability to zero at rate $M^{-1/2}$ as M goes to infinity and T is fixed.*

In Theorem 4, $\widehat{\alpha}_{TM}(\theta_0)$ is the minimum of the empirical MSE simulated with the true θ_0. In the proof, one first shows that the conditions of the uniform convergence in probability of the empirical MSE are satisfied. Next, one uses Theorem 2.1 of Newey and McFadden (1994) and the fact that $\alpha_T(\theta_0)$ is bounded away from zero for any finite T to establish the consistency of $\frac{\widehat{\alpha}_{TM}(\theta_0)}{\alpha_T(\theta_0)}$. In the next theorem, we revisit the previous results when θ_0 is replaced by a consistent estimator $\widehat{\theta}^1$.

Theorem 5 *Let $\widehat{\theta}^1$ be a $\sqrt{T}-$consistent estimator of θ_0. Then under assumptions 1 to 5, $\frac{\widehat{\alpha}_{TM}(\widehat{\theta}^1)}{\alpha_T(\theta_0)} - 1 = O_p(T^{-1/2}) + O_p(M^{-1/2})$ as M goes to infinity first and T goes to infinity second.*

The result of Theorem 5 is obtained by using a sequential limit in M and T, which is needed here because Theorem 4 has been derived for fixed T. Such sequential approach is often used in panel data econometrics, see for instance Phillips and Moon (1999). It is also used implicitly in the theoretical analysis of bootstrap.[3] Theorem 5 implies that $\widehat{\alpha}_{TM}(\widehat{\theta}^1)$ benefits from an increase in both M and T. The last theorem compares the feasible CGMM estimator based on $\widehat{\alpha}_{TM}$ to the unfeasible estimator $\widehat{\theta}\left(\alpha_T\right)$, where α_T is defined in (14).

Theorem 6 *Let $\widehat{\alpha}_{TM} = \widehat{\alpha}_{TM}(\widehat{\theta}^1)$ defined in (22). Then:*

$$\sqrt{T}\left(\widehat{\theta}\left(\widehat{\alpha}_{TM}\right) - \widehat{\theta}\left(\alpha_T\right)\right) = O_p(T^{-g(\beta)}),$$

provided that $M \geq T$.

[3] The properties of a bootstrap estimator are usually derived using its bootstrap distribution, hence letting M go to infinity before T.

Hence, theorem 6 implies that the distribution of $\sqrt{T}\left(\widehat{\theta}\left(\widehat{\alpha}_{TM}\right)-\theta_0\right)$ is the same as the distribution of $\sqrt{T}\left(\widehat{\theta}\left(\alpha_T\right)-\theta_0\right)$ to the order $T^{-g(\beta)}$. This ensures that replacing α_T by a consistent estimator $\widehat{\alpha}_{TM}$ such that $\frac{\widehat{\alpha}_{TM}}{\alpha_T}-1=o_p(1)$ does not affect the consistency, asymptotic normality and efficiency of the final CGMM estimator $\widehat{\theta}\left(\widehat{\alpha}_{TM}\right)$. The proof of this theorem relies mainly on the fact that $\widehat{\theta}\left(\alpha\right)$ is continuously differentiable with respect to α whilst the optimal α_T is bounded away from zero for any finite T. Overall, our selection procedure for the regularization parameter is optimal and adaptive as it does not require the a priori knowledge of the regularity parameter β.

5 Monte Carlo Simulations

The aim of this simulation study is to investigate the properties of the MSE function $\widehat{\Sigma}_{TM}\left(\alpha,\widehat{\theta}_T^1,\nu\right)$ as the regularization parameter (α), the sample size (T) and the number of replications (M) vary. For this purpose, we consider estimating the parameters of a square-root diffusion (also known as the CIR diffusion) by CGMM. Below, the first subsection describes the simulation design whilst the second subsection presents the simulation results.

5.1 Simulation Design

A continuous time process r_t is said to follows a CIR diffusion if it obeys the following stochastic differential equation:

$$dr_t=\kappa\left(\beta-r_t\right)dt+\sigma\sqrt{r_t}dW_t \tag{23}$$

where the parameter $\kappa>0$ is the strength of the mean reversion in the process, $\beta>0$ is the long run mean and $\sigma>0$ controls the volatility of r_t. This model has been widely used in the asset pricing literature, see e.g Heston (1993) or Singleton (2001). It is shown in Feller (1951) that Equation (23) admits a unique and positive fundamental solution if $\sigma^2\leq 2\kappa\beta$.

We assume that r_t is observed at regularly spaced discrete times t_1, t_2, ..., t_T such that $t_i-t_{i-1}=\Delta$. The conditional distribution of r_t given $r_{t-\Delta}$ is a noncentered chi-square with possibly fractional order. Its transition density is a Bessel function of type I, which can be represented as an infinite mixture of Gamma densities with Poisson weights:

$$f\left(r_t|r_{t-\Delta}\right)=\sum_{j=0}^{\infty}p_j\frac{r_t^{j+q_k-1}c^{j+q}}{\Gamma\left(j+q\right)}\exp\left(-cr_t\right)$$

where $c=\frac{2\kappa}{\sigma^2(1-e^{-\kappa\Delta})}$, $q=\frac{2\kappa\beta}{\sigma^2}$ and $p_j=\frac{\left(ce^{-\kappa\Delta}r_{t-\Delta}\right)^j\exp\left(-ce^{-\kappa\Delta}r_{t-\Delta}\right)}{j!}$. To implement a likelihood based inference for this model, one has to truncate the expression of $f\left(r_t|r_{t-\Delta}\right)$. However, the conditional CF of r_t has a simple closed form expression given by:

$$\varphi_t(s,\theta)\equiv E\left(e^{isr_t}|r_{t-\Delta}\right)=\left(1-\frac{is}{c}\right)^{-q}\exp\left(\frac{ise^{-\kappa\Delta}V_{t-1}}{1-\frac{is}{c}}\right) \tag{24}$$

with $\theta = (\kappa, \beta, \sigma)'$.

To start, we simulate one sample of size T from the CIR process assuming $\Delta = 1$ and the true value of θ is:

$$\theta_0 = (\kappa_0, \beta_0, \sigma_0) = (0.4, 6.0, 0.3)\,.$$

These parameter values are taken from Singleton (2001). We refer the reader to DeVroye (1986) and Zhou (2001) for details on how to simulate a CIR process. We treat this simulated sample as the actual data available to the econometrician and use it to estimate the first step CGMM estimator $\widehat{\theta}_T^1$ as:

$$\widehat{\theta}_T^1 = \arg\min_{\theta} \int_{\mathbb{R}^2} \widehat{h}_T(\tau, \theta)\overline{\widehat{h}_T(\tau, \theta)} e^{-\tau'\tau} d\tau,$$

where $\widehat{h}_T(\tau, \theta) = \frac{1}{T-1}\sum_{i=2}^{T}\left(e^{i\tau_1 r_{t_i}} - \varphi_t(s, \theta)\right) e^{i\tau_2 r_{t_{i-1}}}$, $\tau = (\tau_1, \tau_1) \in \mathbb{R}^2$ and $\varphi_t(s, \theta)$ is given by (24). Next, we simulate M samples of size T using $\widehat{\theta}_T^1$ as pseudo-true parameter value. Each simulated samples is used to compute the second step CGMM estimator $\widehat{\theta}_{T,j}(\alpha)$ as:

$$\widehat{\theta}_{T,j}(\alpha) = \arg\min_{\theta} \int_{\mathbb{R}^2} \left(K_{\alpha T}^{-1}\widehat{h}_T(\tau, \theta)\right) \overline{\widehat{h}_T(\tau, \theta)} e^{-\tau'\tau} d\tau \tag{25}$$

The objective function (25) is evaluated using a Gauss-Hermite quadrature with ten points. The regularization parameter α is selected on a thirty points grid that lies between 10^{-10} and 10^{-2}, that is:

$$\alpha \in [10^{-10}, 2.5 \times 10^{-10}, 5 \times 10^{-10}, 7.5 \times 10^{-10}, 1 \times 10^{-9}, ..., 1 \times 10^{-2}]$$

For each α in this grid, we compute the MSE using Equation (21) (i.e., no truncation of the distribution of $\left\|\widehat{\theta}_{T,j}(\alpha) - \widehat{\theta}_T^1\right\|^2$).

5.2 Simulations results

Table 1 shows the simulations $T = 251$, 501, 751 and 1001 for two different values of M. For a given sample size T, the scenarios with $M = 500$ and $M = 1000$ use common random numbers (i.e., the results for $M = 500$ are based on the first 500 replications of the scenarios with $M = 1000$). Curiously enough, the estimate of $\widehat{\alpha}_{TM}(\widehat{\theta}^1)$ is consistently equal to 2.5×10^{-6} across all scenarios except $(T = 251, M = 1000)$ and $(T = 1001, M = 500)$. This result might be suggesting that the grid on which α is selected is not refined enough. Indeed, the value that is immediately smaller than 2.5×10^{-6} on that grid is 1.0×10^{-6}, which is selected for the scenario $(T = 1001, M = 500)$. Arguably, the results suggest that the rate of convergence of $\alpha_T(\theta_0)$ to zero is quite slow for this particular data generating process. Overall, 2.5×10^{-6} seems a reasonable choice for the regularization parameter for all sample sizes for this data generating process. Note that our simulations results do not allow us to infer the behavior of $\alpha_T(\theta_0)$ as θ_0 vary in the parameter space.

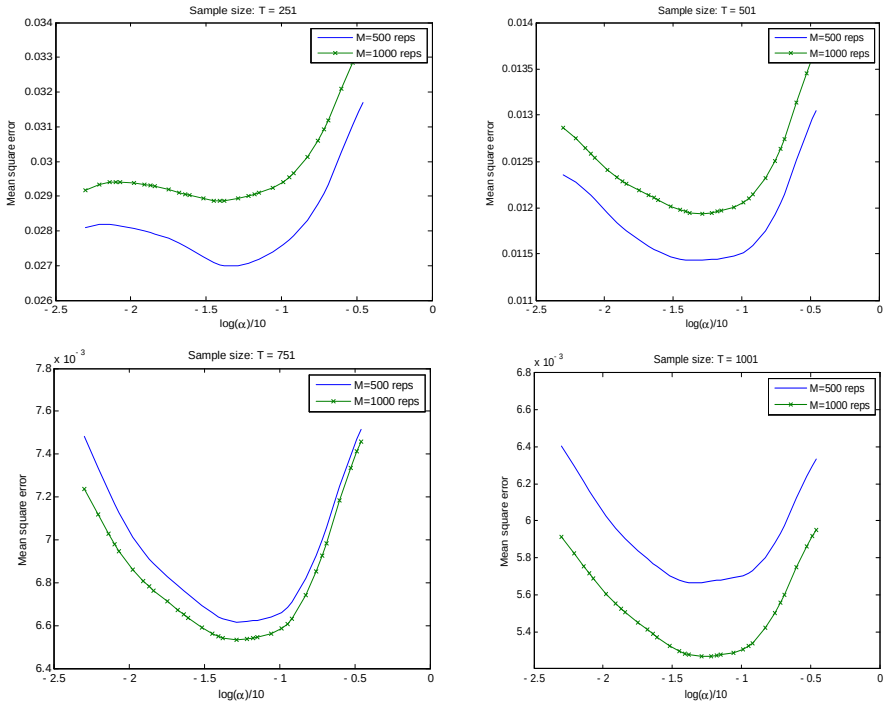
Figure 1 presents the simulated MSE curves. For all eight scenarios, these curves are convex and

have one minimum. The hump-shaped left tail of the MSE curves for $T = 251$ stems to the fact that the approximating matrix of the covariance operator (see Appendix D) is severely ill-posed. Hence, the shape of the MSE curve reflects the distortions inflicted to the eigenvalues of the regularized inverse of this approximating matrix as α varies. A smaller number of quadrature points should be used for smaller sample sizes in order to mitigate this ill-posedness and obtain perfectly convex MSE curves.

Table 1: Estimation of α_T for different sample size.

	$M = 500$		$M = 1000$	
	$\widehat{\alpha}_{TM}(\widehat{\theta}^1)$	$\frac{1}{T}\widehat{\Sigma}_{TM}$	$\widehat{\alpha}_{TM}(\widehat{\theta}^1)$	$\frac{1}{T}\widehat{\Sigma}_{TM}$
$T = 251$	2.5×10^{-6}	0.0270	7.5×10^{-7}	0.0289
$T = 501$	2.5×10^{-6}	0.0114	2.5×10^{-6}	0.0119
$T = 751$	2.5×10^{-6}	0.0066	2.5×10^{-6}	0.0065
$T = 1001$	1.0×10^{-6}	0.0057	2.5×10^{-6}	0.0053

Figure 1: MSE curves of the CGMM estimator for different M and T.
The vertical axis shows $\frac{1}{T}\widehat{\Sigma}_{TM} = \frac{1}{M}\sum_{j=1}^{M} \xi_{j,T}(\alpha, \widehat{\theta}_T^1)$ and the horizontal axis is scaled as $\frac{\log \alpha}{10}$.



6 Conclusion

The objective of this paper is to provide a method to optimally select the regularization parameter denoted α in the CGMM estimation. First, we derive a higher order expansion of the CGMM estimator that sheds light on how the finite sample MSE depends on the regularization parameter. We obtain the

convergence rate for the optimal regularization parameter α_T by equating the rates of two higher order variance terms. We find an expression of the form $\alpha_T = c(\theta_0) T^{-g(\beta)}$, where $c(\theta_0)$ does not depend of the sample size T and $0 \leq g(\beta) \leq 1/2$, where β is the regularity of the moment function with respect to the covariance operator (see Assumption 4).

Next, we propose an estimation procedure for α_T that relies on the minimization of an approximate MSE criterion obtained by Monte Carlo simulations. The proposed estimator, $\widehat{\alpha}_{TM}$, is indexed by the sample size T and the number of Monte Carlo replications M. To hedge against situations where the MSE is not finite, we propose to base the selection of α_T on a truncated MSE that is always finite. Under the assumption that the truncation scheme does not alter the rate of α_T, $\widehat{\alpha}_{TM}$ is consistent for α_T as T and M increase to infinity. Our simulation-based selection procedure has the advantage to be easily applicable to other estimators, for instance it could be used to select the number of polynomial terms in the efficient method of moments procedure of Gallant and Tauchen (1996). The optimal selection of the regularization parameter permits to devise a fully feasible CGMM estimator that is a real alternative to the maximum likelihood estimator.

References

[1] Ait-Sahalia, Y. and R. Kimmel (2007), "Maximum likelihood estimation of stochastic volatility models," *Journal of Financial Economics* 83, 413–452.

[2] Andews, D. W. K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59(3): 817-858

[3] Buse, A. (1992), "The Bias of Instrumental Variable Estimators", *Econometrica* 60:1, 173–180.

[4] Carrasco, M. (2012), "A regularization approach to the many instruments problem," *Journal of Econometrics*, 170:2, 383–398.

[5] Carrasco, M., Chernov, M., Florens J. P. and E. Ghysels (2007), "Efficient estimation of general dynamic models with a continuum of moment conditions," *Journal of Econometrics,* 140, 529–573.

[6] Carrasco, M. and J. P. Florens (2000), "Generalization of GMM to a continuum of moment conditions," *Econometric Theory*, 16, 797-834.

[7] Carrasco, M., J. P. Florens, and E. Renault (2007), "Linear Inverse Problems in Structural Econometrics: Estimation based on spectral decomposition and regularization," in the *Handbook of Econometrics*, Vol. 6, edited by J.J. Heckman and E.E. Leamer.

[8] Chacko, G. and L. Viceira (2003), "Spectral GMM estimation of continuous-time processes," *Journal of Econometrics*, 116, 259-292.

[9] Devroye, L. (1986), *"Non-Uniform Random Variate Generation,"* Spinger-Verlag.

[10] Dominguez, M. A. and I. N. Lobato (2004), "Consistent Estimation of Models Defined by Conditional Moment Restrictions," Econometrica 72:5, 1601-1615.

[11] Donald, S. and W. Newey (2001), "Choosing the number of instruments," *Econometrica*, 69, 1161-1191.

[12] Duffie, D. and K. Singleton (1993), "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, 61, 929-952.

[13] Feller, W. (1951), "Two Singular Diffusion Problems," *Annals of Mathematics*, 54, 173-182.

[14] Feuerverger, A. (1990), "An efficiency result for the empirical characteristic function in stationary time-series models," *Canadian Journal of Statistics*, 18, 155-161.

[15] Feuerverger, A. and P. McDunnough (1981a), "On Efficient Inference in Symmetry Stable Laws and Processes," *Csorgo, M., ed. Statistics and Related Topics.* New York: North Holland, 109-122.

[16] Feuerverger, A. and P. McDunnough (1981b), "On some Fourier Methods for Inference," *J. R. Statist. Assoc.* 76:379-387.

[17] Feuerverger, A. and P. McDunnough (1981c), "On the Efficiency of Empirical Characteristic Function Procedures," *J. R. Statist. Soc.* B, 43, 20-27.

[18] Feuerverger, A. and R. Mureika (1977), "The empirical characteristic function and its applications," *The annals of statistics*, 5, No 1, 88-97.

[19] Gallant, A.R. and G. Tauchen (1996), "Which Moments to Match?" *Econometric Theory*, 12, 657-681.

[20] Gouriéroux, C. and A. Monfort (1996), "Simulation Based Econometric Methods," CORE Lectures, *Oxford University Press*, New York.

[21] Gouriéroux, C., A. Monfort and E. Renault (1993), "Indirect Inference," *Journal of Applied Econometrics* 8, S85-S118 .

[22] Hansen, L. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.

[23] Heston, S. (1993), "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *The Review of Financial Studies*, 6:2, 327-343.

[24] Jacho-Chavez, D. T. (2010), "Optimal Bandwidth Choice for Estimation of Inverse Conditional–Density–Weighted Expectations," *Econometric Theory*, 26, 94-118.

[25] 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.

[26] Koenker, R., Machado, J. A. F., Skeels, C. L. and A. H. I. Welsh (1994), "Momentary Lapses: Moment Expansions and the Robustness of Minimum Distance Estimators," *Econometric Theory*, 10, 172-197.

[27] Koutrouvelis, I. A. (1980), "Regression-type estimation of the parameters of stable laws," *J. Am. Statist. Assoc.* 75:372, 918–928.

[28] Lavergne, P., and V. Patilea (2008), "Smooth minimum distance estimation and testing with conditional estimating equations: uniform in bandwidth theory," *Working paper available at ideas.repec.org.*

[29] Linton, O. (2002) "Edgeworth Approximation for Semiparametric Instrumental Variable and Test Statistics," *Journal of Econometrics*, 106, 325-368.

[30] Liu, Q. and D. A. Pierce (1994), "A Note on Gauss-Hermite Quadrature," *Biometrika*, 81:3, 624-629.

[31] Madan, D. B. and E. Senata (1990), "The Variance Gamma Model for Share Market Returns," *Journal of Business*, 63:4, 511-524.

[32] Nagar, A. L. (1959): "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations," *Econometrica*, 27, 573-595.

[33] Newey, W. K. and D. McFadden (1994), "Large Sample Estimation and Hypotheses Testing," *Handbook of Econometrics*, Vol IV, ed. by R.F. Engle and D.L. McFadden.

[34] Newey, W K. and R. J. Smith (2004), "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica,* 72:1, 219–255.

[35] Nolan, J. P. (2009), *"Stable Distributions: Models for Heavy Tailed Data,"* Birkhauser, Boston,. In preparation.

[36] Paulson, A. S., Holcomb W. E. and R. A. Leitch (1975), "The estimation of the parameters of the stable laws," *Biometrika*, 62:1, 163-170.

[37] Phillips, P. C. B. and H. R. Moon (1999), "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, 67:5, 1057–1112.

[38] Rilstone, P., Srivastava, V. K., and A. Ullah (1996), "The Second-Order Bias and Mean-Squared Error of Nonlinear Estimators," *Journal of Econometrics*, 75, 369-395.

[39] Rothenberg, T. J. (1983), "Asymptotic properties of Some Estimators in Structural Models," *Studies in Econometrics, Time Series and Multivariate Statistics*, edited by S. Karlin, T. Amemiya and L. A. Goodman, New York: Academic Press.

[40] Rothenberg, T. J. (1984), "Approximating the Distributions of Econometric Estimators and Test Statistics," *Handbook of Econometrics*, Vol. 2, ed. by Z. Griliches and M. D. Intriligator. New York: North-Holland.

[41] Singleton, K. (2001), "Estimation of Affine Pricing Models Using the Empirical Characteristic Function," *Journal of Econometrics*, 102, 111-141.

[42] Yu, J. (2004), "Empirical Characteristic Function Estimation and Its Applications," *Econometric Reviews*, 23:2, 93-123.

[43] Zhou, H. (2001), "Finite Sample Properties of EMM, GMM, QMLE, and MLE for a Square-Root Interest Rate Diffusion Model," Mimeo, available at *http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.26.8099.*

Appendix

A Some basic properties of the covariance operator

For more formal proofs of the results mentioned in this appendix, see Carrasco, Florens and Renault (2007). Let K be the covariance operator defined in (5) and (6), and $\widehat{h}_t(\tau,\theta)$ the moment function defined in (1) and (2). Finally, let Φ_β be the subset of $L^2(\pi)$ defined in Assumption 4.

Definition 7 *The range of K denoted $R(K)$ is the set of functions g such that $Kf = g$ for some f in $L^2(\pi)$.*

Proposition 8 *$R(K)$ is a subspace of $L^2(\pi)$.*

Note that the kernel functions $k(s,.)$ and $k(.,r)$ are elements of $L^2(\pi)$ because

$$|k(s,r)|^2 = \left|E\left[h_t(\theta,s)\overline{h_t(\theta,r)}\right]\right|^2 \leq 4,\ \forall\,(s,r)\in\mathbb{R}^{2p} \tag{1}$$

Thus for any $f \in L^2(\pi)$, we have

$$\begin{aligned} |Kf(s)|^2 &= \left|\int k(s,r)f(r)\,\pi(r)\,dr\right|^2 \leq \int |k(s,r)f(r)|^2\,\pi(r)\,dr \\ &\leq 4\int |f(r)|^2\,\pi(r)\,dr < \infty, \end{aligned}$$

implying

$$\|Kf\|^2 = \int |Kf(s)|^2\,\pi(s)\,ds < \infty \Rightarrow Kf \in L^2(\pi).$$

Definition 9 *The null space of K denoted $N(K)$ is the set of functions f in $L^2(\pi)$ such that $Kf = 0$.*

The covariance operator K associated with a moment function based on the CF is such that $N(K) = \{0\}$. See CCFG (2007).

Definition 10 *ϕ is an eigenfunction of K associated with eigenvalue μ if and only if $K\phi = \mu\phi$.*

Proposition 11 *Suppose $\mu_1 \geq \mu_2 \geq .... \geq \mu_j \geq ...$ are the eigenvalues of K. Then the sequence $\{\mu_j\}$ satisfies: (i) $\mu_j > 0$ for all j, (ii) $\mu_1 < \infty$ and $\lim_{j\to\infty}\mu_j = 0$.*

Remark. The covariance operator associated with the CF-based moment function is necessarily compact.

Proposition 12 *Every $f \in L^2(\pi)$ can be decomposed as: $f = \sum_{j=1}^{\infty}\langle f,\phi_j\rangle\phi_j$.*

As a consequence, $Kf = \sum_{j=1}^{\infty}\langle f,\phi_j\rangle K\phi_j = \sum_{j=1}^{\infty}\langle f,\phi_j\rangle\mu_j\phi_j$.

Proposition 13 *If* $0 < \beta_1 \leq \beta_2$, *then* $\Phi_{\beta_2} \subset \Phi_{\beta_1}$.

We recall that Φ_β is the set of functions such that $\left\|K^{-\beta} f\right\| < \infty$. In fact, $f \in R(K^{\beta_2}) \Rightarrow K^{-\beta_2} f$ exist and $\left\|K^{-\beta_2} f\right\|^2 = \sum_{j=1}^{\infty} \mu_j^{-2\beta_2} \left|\left\langle f, \phi_j \right\rangle\right|^2 < \infty$. Thus if $f \in R(K^{\beta_2})$, we have:

$$\left\|K^{-\beta_1} f\right\|^2 = \sum_{j=1}^{\infty} \mu_j^{2(\beta_2-\beta_1)} \mu_j^{-2\beta_2} \left|\left\langle f, \phi_j \right\rangle\right|^2 \leq \mu_1^{2(\beta_2-\beta_1)} \sum_{j=1}^{\infty} \mu_j^{-2\beta_2} \left|\left\langle f, \phi_j \right\rangle\right|^2 < \infty$$

$\Rightarrow K^{-\beta_1} f$ exist $\Rightarrow f \in R(K^{\beta_1})$. This means $R(K) \subset R(K^{1/2})$ so that the function $K^{-1/2} f$ is defined on a wider subset of $L^2(\pi)$ compared to $K^{-1} f$. When $f \in \Phi_1$, $\left\langle K^{-1/2} f, K^{-1/2} f \right\rangle = \left\langle K^{-1} f, f \right\rangle$. But when $f \in \Phi_\beta$ for $1/2 \leq \beta < 1$, the quadratic form $\left\langle K^{-1/2} f, K^{-1/2} f \right\rangle$ is well defined while $\left\langle K^{-1} f, f \right\rangle$ is not.

B Expansion of the MSE and proofs of Theorems 1 and 2

B.1 Preliminary results and proof of Theorem 1

Lemma 14 *Let* $K_\alpha^{-1} = (K^2 + \alpha I)^{-1} K$ *and assume that* $f \in \Phi_\beta$ *for some* $\beta > 1$. *Then as* α *goes to zero and* n *goes to infinity, we have:*

$$\begin{aligned}
\left\|K_{\alpha T}^{-1} - K_\alpha^{-1}\right\| &= O_p\left(\alpha^{-3/2} T^{-1/2}\right), && (2)\\
\left\|\left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right) f\right\| &= O_p\left(\alpha^{-1} T^{-1/2}\right), && (3)\\
\left\|\left(K_\alpha^{-1} - K^{-1}\right) f\right\| &= O\left(\alpha^{\min(1, \frac{\beta-1}{2})}\right), && (4)\\
\left\langle \left(K^{-1} - K_\alpha^{-1}\right) f, f \right\rangle &= O\left(\alpha^{\min\left(1, \frac{2\beta-1}{2}\right)}\right). && (5)
\end{aligned}$$

Proof of Lemma 14. Subsequently, $\phi_j, j = 1, 2..., \infty$ denote the eigenfunctions of the covariance operator K associated respectively with the eigenvalues $\mu_j, j = 1, 2..., \infty$. We first consider (2). By the triangular inequality:

$$\begin{aligned}
&\left\|(K_T^2 + \alpha I)^{-1} K_T - (K^2 + \alpha I)^{-1} K\right\| \\
&\leq \left\|(K_T^2 + \alpha I)^{-1} (K_T - K)\right\| + \left\|(K_T^2 + \alpha I)^{-1} K - (K^2 + \alpha I)^{-1} K\right\| \\
&\leq \underbrace{\left\|(K_T^2 + \alpha I)^{-1}\right\|}_{\leq \alpha^{-1}} \underbrace{\|K_T - K\|}_{= O_p\left(T^{-1/2}\right)} + \left\|\left[(K_T^2 + \alpha I)^{-1} - (K^2 + \alpha I)^{-1}\right] K\right\|,
\end{aligned}$$

where $\|K_T - K\| = O_p\left(T^{-1/2}\right)$ follows from *Proposition 3.3 (i)* of CCFG (2007). We have:

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

This proves (2).

The difference between (2) and (3) is that in (3) we exploit the fact that $f \in \Phi_\beta$ with $\beta > 1$, hence $\left\|K^{-1}f\right\| < \infty$. We can rewrite (3) as

$$\left\|\left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right) f\right\| = \left\|\left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right) K K^{-1} f\right\| \leq \left\|\left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right) K\right\| \left\|K^{-1}f\right\|.$$

We have

$$\begin{aligned}
\left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right) K &= (K_T^2 + \alpha I)^{-1} K_T K - (K^2 + \alpha I)^{-1} K^2 \\
&= (K_T^2 + \alpha I)^{-1} (K_T - K) K \qquad (6) \\
&\quad + \left[(K_T^2 + \alpha I)^{-1} - (K^2 + \alpha I)^{-1}\right] K^2. \qquad (7)
\end{aligned}$$

The term (6) can be bounded in the following manner

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

For the term (7), we use the fact that $A^{-1/2} - B^{-1/2} = A^{-1/2}\left(B^{1/2} - A^{1/2}\right) B^{-1/2}$. It follows that

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

This proves (3).

Now we turn our attention toward equation (4). We can write

$$(K^2 + \alpha I)^{-1} K f - K^{-1} f = \sum_{j=1}^{\infty} \left[\frac{\mu_j}{\alpha + \mu_j^2} - \frac{1}{\mu_j}\right] \langle f, \phi_j \rangle \phi_j = \sum_{j=1}^{\infty} \left(\frac{\mu_j^2}{\alpha + \mu_j^2} - 1\right) \frac{\langle f, \phi_j \rangle}{\mu_j} \phi_j.$$

We now take the norm:

$$\begin{aligned}
(4) &= \left\|(K^2 + \alpha I)^{-1} K f - K^{-1} f\right\| = \left(\sum_{j=1}^{\infty} \left(\frac{\mu_j^2}{\alpha + \mu_j^2} - 1\right)^2 \frac{\left|\langle f, \phi_j \rangle\right|^2}{\mu_j^2}\right)^{1/2} \\
&= \left(\sum_{j=1}^{\infty} \mu_j^{2\beta - 2} \left(\frac{\mu_j^2}{\alpha + \mu_j^2} - 1\right)^2 \frac{\left|\langle f, \phi_j \rangle\right|^2}{\mu_j^{2\beta}}\right)^{1/2} \leq \left(\sum_{j=1}^{\infty} \frac{\left|\langle f, \phi_j \rangle\right|^2}{\mu_j^{2\beta}}\right)^{1/2} \sup_{1 \leq j \leq \infty} \mu_j^{\beta - 1} \frac{\alpha}{\alpha + \mu_j^2}.
\end{aligned}$$

Recall that as K is a compact operator, its largest eigenvalue μ_1 is bounded. We need to find an equivalent to

$$\sup_{0\leq\mu\leq\mu_1}\mu^{\beta-1}\frac{\alpha}{\alpha+\mu_j^2}=\sup_{0\leq\lambda\leq\mu_1^2}\lambda^{\frac{\beta-1}{2}}\left(1-\frac{1}{\alpha/\lambda+1}\right) \tag{8}$$

Case where $1\leq\beta\leq 3$: We apply another change of variables $x=\alpha/\lambda$: $\sup_{x\geq 0}\frac{\alpha^{\beta/2-1/2}}{x^{\beta/2-1/2}}\left(\frac{x}{1+x}\right)$. An equivalent to (8) is $\alpha^{\beta/2-1/2}$ provided that $\frac{1}{x^{\beta/2-1/2}}\left(\frac{x}{1+x}\right)$ is bounded on $\mathbb{R}^+$. Note that $g(x)\equiv\frac{x^{(3-\beta)/2}}{1+x}$ is continuous and therefore bounded on any interval of $(0,+\infty)$. It goes to 0 at $+\infty$ and its limit at 0 also equals 0 for $1\leq\beta<3$. For $\beta=3$, we have: $g(x)\equiv\frac{1}{1+x}$. Then $g(x)$ goes to 1 at $x=0$ and to 0 at $+\infty$.

Case where $\beta>3$: We rewrite the left hand side of (8) as

$$\mu_j^{\beta-1}\frac{\alpha}{\alpha+\mu_j^2}=\alpha\mu_j^{\beta-3}\underbrace{\frac{\mu_j^2}{\alpha+\mu_j^2}}_{\in(0,1)}\leq\alpha\mu_1^{\beta-3}=O(\alpha).$$

To summarize, we have for $f\in\Phi_\beta$: (4)$=O\left(\alpha^{\min(1,\frac{\beta-1}{2})}\right)$.

Finally, we consider (5). We have:

$$\begin{aligned}(5) &= \sum_j\left(\frac{1}{\mu_j}-\frac{\mu_j}{\mu_j^2+\alpha}\right)\langle f,\phi_j\rangle^2=\sum_j\left(1-\frac{\mu_j^2}{\mu_j^2+\alpha}\right)\frac{\langle f,\phi_j\rangle^2}{\mu_j}\\ &= \sum_j\mu_j^{2\beta-1}\left(1-\frac{\mu_j^2}{\mu_j^2+\alpha}\right)\frac{\langle f,\phi_j\rangle^2}{\mu_j^{2\beta}}\leq\sum_j\frac{\langle f,\phi_j\rangle^2}{\mu_j^{2\beta}}\sup_{\mu\leq\mu_1}\mu^{2\beta-1}\frac{\alpha}{\mu^2+\alpha}.\end{aligned}$$

For $\beta\geq 3/2$, we have: $\sup_{\mu\leq\mu_1}\mu^{2\beta-1}\frac{\alpha}{\mu^2+\alpha}\leq\alpha\mu_1^{2\beta-3}=O(\alpha)$. For $\beta<3/2$, we apply the change of variables $x=\alpha/\mu^2$ and obtain $\sup_{x\geq 0}\frac{x}{1+x}\left(\frac{\alpha}{x}\right)^{\frac{2\beta-1}{2}}=O\left(\alpha^{\frac{2\beta-1}{2}}\right)$, as $f(x)=\frac{x}{1+x}x^{-\frac{2\beta-1}{2}}$ is bounded on $\mathbb{R}^+$. Finally: (5)$=O\left(\alpha^{\min(1,\frac{2\beta-1}{2})}\right)$.■

Lemma 15 *Suppose we have a particular function* $f(\theta)\in\Phi_\beta$ *for some* $\beta>1$*, and a sequence of functions* $f_T(\theta)\in\Phi_\beta$ *such that* $\sup_{\theta\in\Theta}\|f_T(\theta)-f(\theta)\|=O_p(T^{-1/2})$*. Then as* α *goes to zero, we have*

$$\sup_{\theta\in\Theta}\left\|K_{\alpha T}^{-1/2}f_T(\theta)-K^{-1/2}f(\theta)\right\|=O_p(\alpha^{-1}T^{-1/2})+O\left(\alpha^{\min(1,\frac{\beta-1}{2})}\right).$$

Proof of Lemma 15.

$$\sup_{\theta\in\Theta}\left\|K_{\alpha T}^{-1}f_T(\theta)-K^{-1}f(\theta)\right\|\leq B_1+B_2,$$

with

$$B_1=\sup_{\theta\in\Theta}\left\|K_{\alpha T}^{-1}f_T(\theta)-K_{\alpha T}^{-1}f(\theta)\right\|\text{ and }B_2=\sup_{\theta\in\Theta}\left\|\left(K_{\alpha T}^{-1}-K^{-1}\right)f(\theta)\right\|.$$

We have

$$\begin{aligned} B_1 &\leq \left\|K_{\alpha T}^{-1}\right\| \sup_{\theta\in\Theta} \|f_T(\theta) - f(\theta)\| \\ &\leq \underbrace{\left\|\left(\alpha_T + K_T^2\right)^{-1/2}\right\|}_{\leq \alpha_T^{-1/2}} \underbrace{\left\|\left(\alpha_T + K_T^2\right)^{-1/2} K_T\right\|}_{\leq 1} \underbrace{\sup_{\theta\in\Theta} \|f_T(\theta) - f(\theta)\|}_{=O_p(T^{-1/2})} \\ &= O_p(\alpha_T^{-1/2} T^{-1/2}). \end{aligned}$$

On the other hand, Lemma 14 implies that:

$$\begin{aligned} B_2 &= \left\|\left(K_{\alpha T}^{-1} - K^{-1}\right) f(\theta)\right\| \\ &\leq \left\|\left(K_{\alpha T}^{-1} - K_{\alpha}^{-1}\right) f(\theta)\right\| + \left\|\left(K_{\alpha}^{-1} - K^{-1}\right) f(\theta)\right\| \\ &= O_p\left(\alpha^{-1} T^{-1/2}\right) + O\left(\alpha^{\min(1, \frac{\beta-1}{2})}\right). \end{aligned}$$

Hence, B_1 is negligible with respect to B_2 and the result follows.■

Lemma 16 *For all nonrandom functions* (u, v), *we have:*

$$E\left[\left\langle u, \widehat{h}_T(.,\theta)\right\rangle \overline{\left\langle v, \widehat{h}_T(.,\theta)\right\rangle}\right] = \frac{1}{T}\langle u, Kv\rangle$$

Proof of Lemma 16. We have:

$$\begin{aligned} E\left[\left\langle u, \widehat{h}_T(.,\theta)\right\rangle \overline{\left\langle v, \widehat{h}_T(.,\theta)\right\rangle}\right] &= E\left[\left(\int u(\tau)\,\overline{\widehat{h}_T(\tau,\theta)}\pi(\tau)\,d\tau\right)\left(\int \overline{v(\tau)}\widehat{h}_T(\tau,\theta)\pi(\tau)\,d\tau\right)\right] \\ &= E\left[\int\int \overline{\widehat{h}_T(\tau_1,\theta)}\widehat{h}_T(\tau_2,\theta) u(\tau_1)\,\overline{v(\tau_2)}\pi(\tau_1)\,\pi(\tau_2)\,d\tau_1 d\tau_2\right] \\ &= \int\int E\left[\overline{\widehat{h}_T(\tau_1,\theta)}\widehat{h}_T(\tau_2,\theta)\right] u(\tau_1)\,\overline{v(\tau_2)}\pi(\tau_1)\,\pi(\tau_2)\,d\tau_1 d\tau_2. \end{aligned}$$

Because the h_ts are uncorrelated, we have:

$$E\left[\widehat{h}_T(\tau_1,\theta)\overline{\widehat{h}_T(\tau_2,\theta)}\right] = \frac{1}{T}E\left[h_t(\tau_1,\theta)\overline{h_t(\tau_2,\theta)}\right] = \frac{1}{T}k(\tau_1,\tau_2),$$

we have

$$\begin{aligned} & E\left[\left\langle u, \widehat{h}_T(.,\theta)\right\rangle \overline{\left\langle v, \widehat{h}_T(.,\theta)\right\rangle}\right] \\ = \frac{1}{T}\int \underbrace{\left(\int \overline{k(\tau_1,\tau_2)v(\tau_2)}\pi(\tau_2)\,d\tau_2\right)}_{\overline{Kv(\tau_1)}} u(\tau_1)\,\pi(\tau_1)\,d\tau_1 &= \frac{1}{T}\langle u, Kv\rangle. \end{aligned}$$

■

Lemma 17 *Let S be a neighborhood of $\widehat{\theta}$, such that $\widetilde{\theta} - \widehat{\theta} = O_p(T^{-1/2})$ for all $\widetilde{\theta} \in S$, where $\widehat{\theta}$ solves:*

$$\left\langle K_{\alpha T}^{-1} \widehat{G}_T(.,\widehat{\theta}), \widehat{h}_T(.,\widehat{\theta}) \right\rangle = 0$$

and $\widehat{G}_T(.,\theta) = \frac{\partial \widehat{h}_T(.,\theta)}{\partial \theta}$. We have:

$$\operatorname{Im} \left\langle K_{\alpha T}^{-1} \widehat{G}_T(.,\widetilde{\theta}), \widehat{h}_T(.,\widetilde{\theta}) \right\rangle = O_p\left(T^{-1}\right) \text{ for all } \widetilde{\theta} \in S.$$

Proof of Lemma 17. Note that S contains θ_0 and:

$$\widetilde{\theta} - \theta_0 = \underbrace{\widetilde{\theta} - \widehat{\theta}}_{O_p\left(T^{-1/2}\right)} + \underbrace{\widehat{\theta} - \theta_0}_{O_p\left(T^{-1/2}\right)} = O_p\left(T^{-1/2}\right).$$

Hence, a first order Taylor expansion of $\widehat{h}_T(.,\widetilde{\theta})$ around θ_0 yields:

$$\widehat{h}_T(.,\widetilde{\theta}) = \widehat{h}_T(.,\theta_0) + \widehat{G}_T(.,\theta_0)\left(\widetilde{\theta} - \theta_0\right) + O_p\left(T^{-1}\right).$$

Likewise, a first order Taylor expansion of $\widehat{G}_T(.,\widetilde{\theta})$ around θ_0 yields:

$$\widehat{G}_T(.,\widetilde{\theta}) = \widehat{G}_T(.,\theta_0) + \sum_{j=1}^{q} \widehat{H}_{j,T}(.,\theta_0)\left(\widetilde{\theta}_j - \theta_{j,0}\right) + O_p\left(T^{-1}\right).$$

Hence, we have:

$$\begin{aligned} \left\langle K_{\alpha T}^{-1} \widehat{G}_T(.,\widetilde{\theta}), \widehat{h}_T(.,\widetilde{\theta}) \right\rangle &= \left\langle K_{\alpha T}^{-1} \widehat{G}_T(.,\theta_0), \widehat{h}_T(.,\theta_0) \right\rangle \\ &\quad + \left\langle K_{\alpha T}^{-1} \widehat{G}_T(.,\theta_0), \widehat{G}_T(.,\theta_0) \right\rangle \left(\widetilde{\theta} - \theta_0\right) + O_p\left(T^{-1}\right). \end{aligned}$$

Note that the term $\left\langle K_{\alpha T}^{-1} \widehat{G}_T(.,\theta_0), \widehat{G}_T(.,\theta_0) \right\rangle \left(\widetilde{\theta} - \theta_0\right)$ is real. At the particular point $\widetilde{\theta} = \widehat{\theta}$ (and for fixed α):

$$0 = \left\langle K_{\alpha T}^{-1} \widehat{G}_T(.,\theta_0), \widehat{h}_T(.,\theta_0) \right\rangle + \left\langle K_{\alpha T}^{-1} \widehat{G}_T(.,\theta_0), \widehat{G}_T(.,\theta_0) \right\rangle \left(\widehat{\theta} - \theta_0\right) + O_p\left(T^{-1}\right).$$

Hence, the imaginary part of $\left\langle K_{\alpha T}^{-1} \widehat{G}_T(.,\theta_0), \widehat{h}_T(.,\theta_0) \right\rangle$ is $O_p\left(T^{-1}\right)$, and so is the imaginary part of $\left\langle K_{\alpha T}^{-1} \widehat{G}_T(.,\widetilde{\theta}), \widehat{h}_T(.,\widetilde{\theta}) \right\rangle$ for all $\widetilde{\theta} \in S$.■

Proof of Theorem 1. The proof follows the same steps as that of Proposition 3.2 in CCFG (2007). However, we now exploit the fact $E\nabla_\theta h_t(\theta) \in \Phi_\beta$ with $\beta \geq 1$. The consistency follows from Lemma 15 provided $\alpha T^{1/2} \to \infty$ and $\alpha \to 0$. For the asymptotic normality to hold, we need to find a

bound for the term B.10 of CCFG (2007). We have:

$$\begin{aligned}
|B.10| &= \left|\left\langle K_{\alpha T}^{-1}\nabla_\theta \hat{h}_T\left(\hat{\theta}_T\right) - K^{-1}E\left(\nabla_\theta \hat{h}_T\left(\theta_0\right)\right), \sqrt{T}\hat{h}_T\left(\theta_0\right)\right\rangle\right| \\
&\leq \left\|K_{\alpha T}^{-1/2}\nabla_\theta \hat{h}_T\left(\hat{\theta}_T\right) - K^{-1/2}E\left(\nabla_\theta \hat{h}_T\left(\theta_0\right)\right)\right\| \underbrace{\left\|\sqrt{T}\hat{h}_T\left(\theta_0\right)\right\|}_{=O_p(1)} \\
&= O_p\left(\alpha^{-1/2}T^{-1/2}\right) + O\left(\alpha^{\min(1,\frac{\beta-1}{2})}\right).
\end{aligned}$$

Hence the asymptotic normality requires the same conditions as the consistency, that is, $\alpha T^{1/2} \to \infty$ and $\alpha \to 0$. The asymptotic efficiency follows from the fact that $K_{\alpha T}^{-1} \to K^{-1}$ under the same conditions.■

B.2 Stochastic expansion of the CGMM estimator: IID case

The objective function is

$$\widehat{\theta} = \arg\min_\theta \left\{Q_{\alpha T}\left(\theta\right) = \left\langle K_{\alpha T}^{-1}\widehat{h}_T(.,\theta), \widehat{h}_T(.,\theta)\right\rangle\right\}.$$

where $\widehat{h}_T(\tau,\theta) = \frac{1}{T}\sum_{t=1}^T \left(e^{i\tau' x_t} - \varphi(\tau,\theta)\right)$. The optimal $\widehat{\theta}$ solves:

$$\frac{\partial Q_{\alpha T}\left(\widehat{\theta}\right)}{\partial\theta} = 2\,\mathrm{Re}\left\langle K_{\alpha T}^{-1}G(.,\widehat{\theta}), \widehat{h}_T(.,\widehat{\theta})\right\rangle = 0 \tag{9}$$

where $G(.,\theta) = -\frac{\partial\varphi(\tau,\theta)}{\partial\theta}$.

A third order expansion gives

$$0 = \frac{\partial Q_{\alpha T}\left(\theta_0\right)}{\partial\theta} + \frac{\partial^2 Q_{\alpha T}\left(\theta_0\right)}{\partial\theta\partial\theta'}\left(\widehat{\theta} - \theta_0\right) + \sum_{j=1}^q \left(\widehat{\theta}_j - \theta_{j,0}\right)\frac{\partial^3 Q_{\alpha T}\left(\bar{\theta}\right)}{\partial\theta_j\partial\theta\partial\theta'}\left(\widehat{\theta} - \theta_0\right),$$

where $\bar{\theta}$ lies between $\widehat{\theta}$ and θ_0. The dependence of $\widehat{\theta}$ on α_T is hidden for convenience. Let us define

$$G_j(.,\theta) = -\frac{\partial\varphi(\tau,\theta)}{\partial\theta_j},\ H(.,\theta) = -\frac{\partial^2\varphi(\tau,\theta)}{\partial\theta\partial\theta'},\ H_j(.,\theta) = -\frac{\partial^2\varphi(\tau,\theta)}{\partial\theta\partial\theta_j},\ L_j = -\frac{\partial^3\varphi(\tau,\theta)}{\partial\theta_j\partial\theta\partial\theta'}.$$

and

$$\begin{aligned}
\Psi_T(\theta_0) &= \mathrm{Re}\left\langle K_{\alpha T}^{-1}G(.,\theta_0), \widehat{h}_T(.,\theta_0)\right\rangle, \\
W_T(\theta_0) &= \left\langle K_{\alpha T}^{-1}G(.,\theta_0), G(.,\theta_0)\right\rangle + \mathrm{Re}\left\langle K_{\alpha T}^{-1}H(.,\theta_0), \widehat{h}_T(.,\theta_0)\right\rangle, \\
B_{j,T}(\bar{\theta}) &= 2\,\mathrm{Re}\left\langle K_{\alpha T}^{-1}G(.,\bar{\theta}), H_j(.,\bar{\theta})\right\rangle + \mathrm{Re}\left\langle K_{\alpha T}^{-1}L_j(.,\bar{\theta}), \widehat{h}_T(.,\bar{\theta})\right\rangle \\
&\quad + \mathrm{Re}\left\langle K_{\alpha T}^{-1}H(.,\bar{\theta}), G_j(.,\bar{\theta})\right\rangle.
\end{aligned}$$

Then we can write:

$$0 = \Psi_T(\theta_0) + W_T(\theta_0)\left(\widehat{\theta} - \theta_0\right) + \sum_{j=1}^{q}\left(\widehat{\theta}_j - \theta_{j,0}\right) B_{j,T}(\bar{\theta})\left(\widehat{\theta} - \theta_0\right).$$

Note that the derivatives of the moment functions are deterministic in the IID case. We decompose $\Psi_T(\theta_0)$, $W_T(\theta_0)$ and $B_{j,T}(\bar{\theta})$ as follows:

$$\Psi_T(\theta_0) = \Psi_{T,0}(\theta_0) + \Psi_{T,\alpha}(\theta_0) + \widetilde{\Psi}_{T,\alpha}(\theta_0),$$

where

$$\begin{aligned}
\Psi_{T,0}(\theta_0) &= \operatorname{Re}\left\langle K^{-1}G, \widehat{h}_T\right\rangle = O_p\left(T^{-1/2}\right) \\
\Psi_{T,\alpha}(\theta_0) &= \operatorname{Re}\left\langle \left(K_\alpha^{-1} - K^{-1}\right)G, \widehat{h}_T\right\rangle = O_p\left(\alpha^{\min(1,\frac{\beta-1}{2})}T^{-1/2}\right) \\
\widetilde{\Psi}_{T,\alpha}(\theta_0) &= \operatorname{Re}\left\langle \left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right)G, \widehat{h}_T\right\rangle = O_p\left(\alpha^{-1}T^{-1}\right)
\end{aligned}$$

where the rates of convergence are obtained using the Cauchy-Schwarz inequality and the results of Lemma 14. Similarly, we decompose $W_T(\theta_0)$ into various terms with distinct rates of convergence:

$$W_T(\theta_0) = W_0(\theta_0) + W_\alpha(\theta_0) + \widetilde{W}_\alpha(\theta_0) + W_{T,0}(\theta_0) + \widetilde{W}_{T,\alpha}(\theta_0),$$

where

$$\begin{aligned}
W_0(\theta_0) &= \left\langle K^{-1}G, G\right\rangle = O(1), \\
W_\alpha(\theta_0) &= \left\langle \left(K_\alpha^{-1} - K^{-1}\right)G, G\right\rangle = O\left(\alpha^{\min(1,\frac{2\beta-1}{2})}\right), \\
\widetilde{W}_\alpha(\theta_0) &= \left\langle \left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right)G, G\right\rangle = O_p\left(\alpha^{-1}T^{-1/2}\right), \\
W_{T,0}(\theta_0) &= \operatorname{Re}\left\langle K^{-1}H(.,\theta_0), \widehat{h}_T(.,\theta_0)\right\rangle = O_p\left(T^{-1/2}\right), \\
\widetilde{W}_{T,\alpha}(\theta_0) &= \operatorname{Re}\left\langle \left(K_{\alpha T}^{-1} - K^{-1}\right)H(.,\theta_0), \widehat{h}_T(.,\theta_0)\right\rangle = O_p\left(\alpha^{-1}T^{-1}\right).
\end{aligned}$$

We consider a simpler decomposition for $B_{j,T}(\bar{\theta})$:

$$B_{j,T}(\bar{\theta}) = B_j(\bar{\theta}) + \left(B_{j,T}(\bar{\theta}) - B_j(\bar{\theta})\right)$$

where

$$\begin{aligned}
B_j(\bar{\theta}) &= 2\operatorname{Re}\left\langle K^{-1}G(.,\bar{\theta}), H_j(.,\bar{\theta})\right\rangle + \operatorname{Re}\left\langle K^{-1}H(.,\bar{\theta}), G_j(.,\bar{\theta})\right\rangle = O(1), \\
B_{j,T}(\bar{\theta}) &= B_j(\bar{\theta}) + O\left(\alpha^{\min(1,\frac{\beta-1}{2})}\right) + O_p\left(\alpha^{-1}T^{-1/2}\right).
\end{aligned}$$

By replacing these decompositions into the expansion of the FOC, we can solve for $\widehat{\theta}-\theta_0$ to obtain:

$$\begin{aligned}
\widehat{\theta}-\theta_0 &= -W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0) \\
&\quad -W_0^{-1}(\theta_0)\left[\Psi_{T,\alpha}(\theta_0)+W_\alpha(\theta_0)\left(\widehat{\theta}-\theta_0\right)\right] \\
&\quad -W_0^{-1}(\theta_0)\left[\widetilde{\Psi}_{T,\alpha}(\theta_0)+\widetilde{W}_\alpha(\theta_0)\left(\widehat{\theta}-\theta_0\right)\right] \\
&\quad -W_0^{-1}(\theta_0)W_{T,0}(\theta_0)\left(\widehat{\theta}-\theta_0\right) \\
&\quad -\sum_{j=1}^{q}\left(\widehat{\theta}_j-\theta_{j,0}\right)W_0^{-1}(\theta_0)B_j(\bar{\theta})\left(\widehat{\theta}-\theta_0\right) \\
&\quad -\sum_{j=1}^{q}\left(\widehat{\theta}_j-\theta_{j,0}\right)W_0^{-1}(\theta_0)(B_{j,T}(\bar{\theta})-B_j(\bar{\theta}))\left(\widehat{\theta}-\theta_0\right).
\end{aligned}$$

To complete the expansion, we replace $\widehat{\theta}-\theta_0$ by $-W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0)$ in the higher order terms:

$$\widehat{\theta}-\theta_0=\Delta_1+\Delta_2+\Delta_3+\Delta_4+\Delta_5+\widehat{R},$$

where $\widehat{R}$ is a remainder that goes to zero faster than the following terms:

$$\begin{aligned}
\Delta_1 &= -W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0), \\
\Delta_2 &= -W_0^{-1}(\theta_0)\left[\Psi_{T,\alpha}(\theta_0)-W_\alpha(\theta_0)W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0)\right], \\
\Delta_3 &= -W_0^{-1}(\theta_0)\left[\widetilde{\Psi}_{T,\alpha}(\theta_0)-\widetilde{W}_\alpha(\theta_0)W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0)\right], \\
\Delta_4 &= W_0^{-1}(\theta_0)W_{T,0}(\theta_0)W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0) \\
&\quad -\sum_{j=1}^{q}\left(W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0)\right)_j W_0^{-1}(\theta_0)B_j(\bar{\theta})W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0), \\
\Delta_5 &= -\sum_{j=1}^{q}\left(W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0)\right)_j W_0^{-1}(\theta_0)(B_{j,T}(\bar{\theta})-B_j(\bar{\theta}))W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0).
\end{aligned}$$

To obtain the rates of these terms, we use the fact that $|Af|\leq\|A\|\,|f|$. This yields immediately:

$$\begin{aligned}
\Delta_1 &= O_p\left(T^{-1/2}\right);\ \Delta_2=O_p\left(\alpha^{\min(1,\frac{2\beta-1}{2})}T^{-1/2}\right),\ \Delta_3=O_p\left(\alpha^{-1}T^{-1}\right);\ \Delta_4=O_p\left(T^{-1}\right), \\
\Delta_5 &= O\left(\alpha^{\min(1,\frac{\beta-1}{2})}T^{-1}\right)+O_p\left(\alpha^{-1}T^{-3/2}\right).
\end{aligned}$$

To summarize, we have:

$$\widehat{\theta}-\theta_0=\Delta_1+\Delta_2+\Delta_3+o_p\left(\alpha^{-1}T^{-1}\right)+o_p\left(\alpha^{\min(1,\frac{\beta-1}{2})}T^{-1/2}\right). \tag{10}$$

B.3 Stochastic expansion of the CGMM estimator: Markov case

The objective function here is given by:

$$\widehat{\theta} = \arg\min_{\theta} \left\{ Q_{\alpha T}(\theta) = \left\langle K_{\alpha T}^{-1} \widehat{h}_T(.,\theta), \widehat{h}_T(.,\theta) \right\rangle \right\}.$$

where $\widehat{h}_T(\tau,\theta) = \frac{1}{T}\sum_{t=1}^{T} \left(e^{is'x_{t+1}} - \varphi(s,\theta,x_t) \right) e^{ir'x_t}$ and $\tau = (s,r) \in \mathbb{R}^{2p}$. The optimal $\widehat{\theta}$ solves

$$\frac{\partial Q_{\alpha T}\left(\widehat{\theta}\right)}{\partial \theta} = 2\,\mathrm{Re}\left\langle K_{\alpha T}^{-1} \widehat{G}_T(.,\widehat{\theta}), \widehat{h}_T(.,\widehat{\theta}) \right\rangle = 0 \tag{11}$$

where $\widehat{G}_T(\tau,\theta) = -\frac{1}{T}\sum_{t=1}^{T} \frac{\partial \varphi(s,\theta,x_t)}{\partial \theta} e^{ir'x_t}$.

The third order Taylor expansion of (11) around θ_0 yields:

$$0 = \frac{\partial Q_{\alpha T}(\theta_0)}{\partial \theta} + \frac{\partial^2 Q_{\alpha T}(\theta_0)}{\partial \theta \partial \theta'}(\widehat{\theta} - \theta_0) + \sum_{j=1}^{q} \left(\widehat{\theta}_j - \theta_{j,0} \right) \frac{\partial^3 Q_{\alpha T}\left(\bar{\theta}\right)}{\partial \theta_j \partial \theta \partial \theta'}(\widehat{\theta} - \theta_0),$$

where $\bar{\theta}$ lies between $\widehat{\theta}$ and θ_0.

Let us define:

$$\begin{aligned}
\widehat{H}_T(\tau,\theta) &= -\frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2 \varphi(s,\theta,x_t)}{\partial \theta \partial \theta'} e^{ir'x_t},\ \widehat{G}_{j,T}(\tau,\theta) = -\frac{1}{T}\sum_{t=1}^{T} \frac{\partial \varphi(s,\theta,x_t)}{\partial \theta_j} e^{ir'x_t}, \\
\widehat{H}_{j,T}(\tau,\theta) &= -\frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2 \varphi(s,\theta,x_t)}{\partial \theta_j \partial \theta} e^{ir'x_t},\ \widehat{L}_{j,T}(\tau,\theta) = -\frac{1}{T}\sum_{t=1}^{T} \frac{\partial^3 \varphi(s,\theta,x_t)}{\partial \theta_j \partial \theta \partial \theta'} e^{ir'x_t},
\end{aligned}$$

and

$$\begin{aligned}
\widehat{\Psi}_T(\theta_0) &= \mathrm{Re}\left\langle K_{\alpha T}^{-1} \widehat{G}_T(.,\theta_0), \widehat{h}_T(.,\theta_0) \right\rangle, \\
\widehat{W}_T(\theta_0) &= \left\langle K_{\alpha T}^{-1} \widehat{G}_T(.,\theta_0), \widehat{G}_T(.,_0) \right\rangle + \mathrm{Re}\left\langle K_{\alpha T}^{-1} \widehat{H}_T(.,\theta_0), \widehat{h}_T(.,\theta_0) \right\rangle, \\
\widehat{B}_{j,T}(\bar{\theta}) &= 2\,\mathrm{Re}\left\langle K_{\alpha T}^{-1} \widehat{G}_T(.,\bar{\theta}), \widehat{H}_{j,T}(.,\bar{\theta}) \right\rangle + \mathrm{Re}\left\langle K_{\alpha T}^{-1} \widehat{H}_T(.,\bar{\theta}), \widehat{G}_{j,T}(.,\bar{\theta}) \right\rangle \\
&\quad + \mathrm{Re}\left\langle K_{\alpha T}^{-1} \widehat{L}_{j,T}(.,\bar{\theta}), \widehat{h}_T(.,\bar{\theta}) \right\rangle.
\end{aligned}$$

Then the expansion of the FOC becomes:

$$0 = \widehat{\Psi}_T(\theta_0) + \widehat{W}_T(\theta_0)\left(\widehat{\theta} - \theta_0\right) + \sum_{j=1}^{q} \left(\widehat{\theta}_j - \theta_{j,0} \right) \widehat{B}_{j,T}(\bar{\theta}) \left(\widehat{\theta} - \theta_0\right),$$

Unlike in the IID case, the derivatives of the moment function are not deterministic. We thus define:

$$\begin{aligned} G(\tau,\theta) &= \underset{T\to\infty}{p\lim}\widehat{G}_T(\tau,\theta),\ H(\tau,\theta) = \underset{T\to\infty}{p\lim}\widehat{H}_T(\tau,\theta), \\ G_j(\tau,\theta) &= \underset{T\to\infty}{p\lim}\widehat{G}_{j,T}(\tau,\theta),\ H_j(\tau,\theta) = \underset{T\to\infty}{p\lim}\widehat{H}_{j,T}(\tau,\theta). \end{aligned}$$

It follows from Assumption 3 and Markov's inequality that:

$$\begin{aligned} G(\tau,\theta) - \widehat{G}_T(\tau,\theta) &= O_p\left(T^{-1/2}\right),\ H(\tau,\theta) - \widehat{H}_T(\tau,\theta) = O_p\left(T^{-1/2}\right), \\ G_j(\tau,\theta) - \widehat{G}_{j,T}(\tau,\theta) &= O_p\left(T^{-1/2}\right),\ H_j(\tau,\theta) - \widehat{H}_{j,T}(\tau,\theta) = O_p\left(T^{-1/2}\right). \end{aligned}$$

We have the following decomposition for $\widehat{\Psi}_T(\theta_0)$:

$$\widehat{\Psi}_T(\theta_0) = \Psi_{T,0}(\theta_0) + \Psi_{T,\alpha}(\theta_0) + \widetilde{\Psi}_{T,\alpha}(\theta_0) + \widehat{\Psi}_{T,\alpha}(\theta_0) + \widehat{\widetilde{\Psi}}_{T,\alpha}(\theta_0).$$

By using the fact that $\|Af\| \leq \|A\|\,\|f\|$, we obtain the following the rates:

$$\begin{aligned} \Psi_{T,0}(\theta_0) &= \mathrm{Re}\left\langle K^{-1}G, \widehat{h}_T(.,\theta_0)\right\rangle = O_p\left(T^{-1/2}\right), \\ \Psi_{T,\alpha}(\theta_0) &= \mathrm{Re}\left\langle \left(K_\alpha^{-1} - K^{-1}\right)G, \widehat{h}_T(.,\theta_0)\right\rangle = O_p\left(\alpha^{\min(1,\frac{\beta-1}{2})}T^{-1/2}\right), \\ \widetilde{\Psi}_{T,\alpha}(\theta_0) &= \mathrm{Re}\left\langle \left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right)G, \widehat{h}_T(.,\theta_0)\right\rangle = O_p\left(\alpha^{-1}T^{-1}\right), \\ \widehat{\Psi}_{T,\alpha}(\theta_0) &= \mathrm{Re}\left\langle K_\alpha^{-1}\left(\widehat{G}_T - G\right), \widehat{h}_T(.,\theta_0)\right\rangle = O_p\left(\alpha^{-1/2}T^{-1}\right), \\ \widehat{\widetilde{\Psi}}_{T,\alpha}(\theta_0) &= \mathrm{Re}\left\langle \left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right)\left(\widehat{G}_T - G\right), \widehat{h}_T(.,\theta_0)\right\rangle = O_p\left(\alpha^{-3/2}T^{-3/2}\right). \end{aligned}$$

The difference between the above decomposition of $\widehat{\Psi}_T(\theta_0)$ and the one in the IID case only comes from the additional higher order terms $\widehat{\Psi}_{T,\alpha}(\theta_0)$ and $\widehat{\widetilde{\Psi}}_{T,\alpha}(\theta_0)$. Hence we can write $\widehat{\Psi}_T(\theta_0)$ as:

$$\widehat{\Psi}_T(\theta_0) = \Psi_{T,0}(\theta_0) + \Psi_{T,\alpha}(\theta_0) + \widetilde{\Psi}_{T,\alpha}(\theta_0) + R_\Psi,$$

where $R_\Psi = o_p\left(\alpha^{-1}T^{-1}\right) + o_p\left(\alpha^{\min(1,\frac{\beta-1}{2})}T^{-1/2}\right)$.

We have a similar decomposition for $\widehat{W}_T(\theta_0)$:

$$\begin{aligned} \widehat{W}_T(\theta_0) &= W_0(\theta_0) + W_\alpha(\theta_0) + \widetilde{W}_\alpha(\theta_0) + \widehat{W}_\alpha(\theta_0) + \widehat{\widetilde{W}}_\alpha(\theta_0) \\ &\quad + W_1(\theta_0) + W_{1,\alpha}(\theta_0) + \widetilde{W}_{1,\alpha}(\theta_0) + \widehat{W}_{1,\alpha}(\theta_0) + \widehat{\widetilde{W}}_{1,\alpha}(\theta_0), \end{aligned}$$

where

$$\begin{aligned} W_0(\theta_0) &= \left\langle K^{-1}G, G\right\rangle = O(1), \\ W_\alpha(\theta_0) &= \left\langle \left(K_\alpha^{-1} - K^{-1}\right)G, G\right\rangle = O\left(\alpha^{\min(1,\frac{2\beta-1}{2})}\right), \end{aligned}$$

$$\widetilde{W}_\alpha(\theta_0) = \left\langle \left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right) G, G \right\rangle = O_p\left(\alpha^{-1}T^{-1/2}\right),$$
$$\widehat{W}_\alpha(\theta_0) = \left\langle K_\alpha^{-1}\left(\widehat{G}_T - G\right), G \right\rangle = O_p\left(\alpha^{-1/2}T^{-1/2}\right),$$
$$W_1(\theta_0) = \mathrm{Re}\left\langle K^{-1}H, \widehat{h}_T \right\rangle + \left\langle K^{-1}G, \widehat{G}_T - G \right\rangle = O_p\left(T^{-1/2}\right),$$
$$W_{1,\alpha}(\theta_0) = \left\langle \left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right)\left(\widehat{G}_T - G\right), G \right\rangle = O_p\left(\alpha^{-3/2}T^{-1}\right).$$
$$\widehat{\widetilde{W}}_{1,\alpha}(\theta_0) = \mathrm{Re}\left\langle \left(K_\alpha^{-1} - K^{-1}\right) H, \widehat{h}_T \right\rangle + \left\langle \left(K_\alpha^{-1} - K^{-1}\right) G, \widehat{G}_T - G \right\rangle = O\left(\alpha^{\min(1,\frac{\beta-1}{2})}T^{-1/2}\right),$$
$$\widetilde{W}_{1,\alpha}(\theta_0) = \mathrm{Re}\left\langle \left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right) H, \widehat{h}_T \right\rangle + \left\langle \left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right) G, \widehat{G}_T - G \right\rangle = O_p\left(\alpha^{-1}T^{-1}\right),$$
$$\widehat{W}_{1,\alpha}(\theta_0) = \mathrm{Re}\left\langle K_\alpha^{-1}\left(\widehat{H}_T - H\right), \widehat{h}_T \right\rangle + \left\langle K_\alpha^{-1}\left(\widehat{G}_T - G\right), \widehat{G}_T - G \right\rangle = O_p\left(\alpha^{-1/2}T^{-1}\right) \text{ and}$$
$$R_{W,1} = \mathrm{Re}\left\langle \left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right)\left(\widehat{H}_T - H\right), \widehat{h}_T \right\rangle + \left\langle \left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right)\left(\widehat{G}_T - G\right), \widehat{G}_T - G \right\rangle = O_p\left(\alpha^{-3/2}T^{-3/2}\right).$$

For the purpose of finding the optimal α, it is enough to consider the shorter decomposition:

$$\widehat{W}_T(\theta_0) = W_0(\theta_0) + W_\alpha(\theta_0) + \widetilde{W}_\alpha(\theta_0) + \widehat{W}_\alpha(\theta_0) + W_1(\theta_0) + W_{1,\alpha}(\theta_0) + R_{W,}$$

with

$$R_W \equiv \widehat{\widetilde{W}}_{1,\alpha}(\theta_0) + \widetilde{W}_{1,\alpha}(\theta_0) + \widehat{W}_{1,\alpha}(\theta_0) + R_{W,1} = O_p\left(\alpha^{-1}T^{-1}\right) + O\left(\alpha^{\min(1,\frac{\beta-1}{2})}T^{-1/2}\right).$$

Finally, we consider again a simpler decomposition for $B_{j,T}(\bar{\theta})$:

$$B_{j,T}(\bar{\theta}) = B_j(\bar{\theta}) + \left(B_{j,T}(\bar{\theta}) - B_j(\bar{\theta})\right)$$

where

$$\begin{aligned} B_j(\bar{\theta}) &= 2\,\mathrm{Re}\left\langle K^{-1}G(.,\bar{\theta}), H_j(.,\bar{\theta})\right\rangle + \mathrm{Re}\left\langle K^{-1}H(.,\bar{\theta}), G_j(.,\bar{\theta})\right\rangle = O(1) \text{ and} \\ B_{j,T}(\bar{\theta}) &= B_j(\bar{\theta}) + O\left(\alpha^{\min(1,\frac{\beta-1}{2})}\right) + O_p\left(\alpha^{-1}T^{-1/2}\right). \end{aligned}$$

We replace these decompositions into the expansion of the FOC and solve for $\widehat{\theta} - \theta_0$ to obtain:

$$\begin{aligned} \widehat{\theta} - \theta_0 = & -W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0) \\ & -W_0^{-1}(\theta_0)\left[\Psi_{T,\alpha}(\theta_0) + W_\alpha(\theta_0)\left(\widehat{\theta} - \theta_0\right)\right] \\ & -W_0^{-1}(\theta_0)\left[\widetilde{\Psi}_{T,\alpha}(\theta_0) + \widetilde{W}_\alpha(\theta_0)\left(\widehat{\theta} - \theta_0\right)\right] - W_0^{-1}(\theta_0)\widehat{W}_\alpha(\theta_0)\left(\widehat{\theta} - \theta_0\right) \\ & -W_0^{-1}(\theta_0)W_1(\theta_0)\left(\widehat{\theta} - \theta_0\right) - \sum_{j=1}^{q}\left(\widehat{\theta}_j - \theta_{j,0}\right)W_0^{-1}(\theta_0)B_j(\bar{\theta})\left(\widehat{\theta} - \theta_0\right) \\ & -W_0^{-1}(\theta_0)W_{1,\alpha}(\theta_0)\left(\widehat{\theta} - \theta_0\right) \\ & -\sum_{j=1}^{q}\left(\widehat{\theta}_j - \theta_{j,0}\right)W_0^{-1}(\theta_0)(B_{j,T}(\bar{\theta}) - B_j(\bar{\theta}))\left(\widehat{\theta} - \theta_0\right) \\ & -W_0^{-1}(\theta_0)R_W\left(\widehat{\theta} - \theta_0\right) - W_0^{-1}(\theta_0)R_\Psi. \end{aligned}$$

Next, we replace $\widehat{\theta} - \theta_0$ by $-W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0) = O_p\left(T^{-1/2}\right)$ in the higher order terms. This yields:

$$\widehat{\theta} - \theta_0 = \Delta_1 + \Delta_2 + \Delta_3 + \widehat{R}_1 + \widehat{R}_2 + \widehat{R}_3 + \widehat{R}_4,$$

where

$$\begin{aligned}
\Delta_1 &= -W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0) = O_p\left(T^{-1/2}\right), \\
\Delta_2 &= -W_0^{-1}(\theta_0)\left[\Psi_{T,\alpha}(\theta_0) - W_\alpha(\theta_0)W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0)\right] = O_p\left(\alpha^{\min(1,\frac{2\beta-1}{2})}T^{-1/2}\right), \\
\Delta_3 &= -W_0^{-1}(\theta_0)\left[\widetilde{\Psi}_{T,\alpha}(\theta_0) - \widetilde{W}_\alpha(\theta_0)W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0)\right] = O_p\left(\alpha^{-1}T^{-1}\right), \\
\widehat{R}_1 &= W_0^{-1}(\theta_0)\widehat{W}_\alpha(\theta_0)W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0) = O_p\left(\alpha^{-1/2}T^{-1}\right), \\
\widehat{R}_2 &= W_0^{-1}(\theta_0)W_1(\theta_0)W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0) \\
&\quad - \sum_{j=1}^{q}\left(W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0)\right)_j W_0^{-1}(\theta_0)B_j(\bar{\theta})W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0) = O_p\left(T^{-1}\right), \\
\widehat{R}_3 &= W_0^{-1}(\theta_0)W_{1,\alpha}(\theta_0)W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0) = O_p\left(\alpha^{-3/2}T^{-3/2}\right),
\end{aligned}$$

and

$$\begin{aligned}
\widehat{R}_4 &= -W_0^{-1}(\theta_0)R_\Psi + W_0^{-1}(\theta_0)R_W W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0) \\
&\quad - \sum_{j=1}^{q}\left(W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0)\right)_j W_0^{-1}(\theta_0)(B_{j,T}(\bar{\theta}) - B_j(\bar{\theta}))W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0), \\
&= o_p\left(\alpha^{-1}T^{-1}\right) + o_p\left(\alpha^{\min(1,\frac{\beta-1}{2})}T^{-1/2}\right).
\end{aligned}$$

In summary, we have:

$$\widehat{\theta} - \theta_0 = \Delta_1 + \Delta_2 + \Delta_3 + o_p\left(\alpha^{-1}T^{-1}\right) + o_p\left(\alpha^{\min(1,\frac{\beta-1}{2})}T^{-1/2}\right), \tag{12}$$

which is of the same form as in the IID case.

B.4 Proof of Theorem 2.

Using the expansions given in (10) and (12), we obtain:

$$\widehat{\theta} - \theta_0 = \Delta_1 + \Delta_2 + \Delta_3 + O_p\left(T^{-1}\right).$$

Lemma 17 ensures that all terms that are slower than $O_p\left(T^{-1}\right)$ in the expansion above are real. Hence the Re symbol may be removed from the expression of Δ_1, Δ_2 and Δ_3.

Asymptotic Variance

The asymptotic variance of $\widehat{\theta}$ is given by

$$\begin{aligned} TVar\left(\Delta_1\right) &= TW_0^{-1}E\left[\Psi_{T,0}(\theta_0)\Psi_{T,0}(\theta_0)'\right]W_0^{-1} \\ &= TW_0^{-1}E\left[\left\langle K^{-1}G,\widehat{h}_T\right\rangle\overline{\left\langle K^{-1}G,\widehat{h}_T\right\rangle}'\right]W_0^{-1} = W_0^{-1}\left\langle K^{-1}G,G\right\rangle W_0^{-1}, \end{aligned}$$

where the last equality follows from Lemma 16. Hence,

$$TVar\left(\Delta_1\right) = W_0^{-1}\left\langle K^{-1}G,G\right\rangle W_0^{-1} = W_0^{-1}.$$

Higher Order Bias

The terms Δ_1 and Δ_2 have zero expectations. Hence, the bias comes from Δ_3:

$$Bias \equiv E\left[\widehat{\theta}-\theta_0\right] = E\left[\Delta_3\right]$$

where $\Delta_3 = -W_0^{-1}\widetilde{\Psi}_{T,\alpha} + W_0^{-1}\widetilde{W}_\alpha W_0^{-1}\Psi_{T,0}$. As W_0^{-1} is a constant matrix, we focus on $\widetilde{\Psi}_{T,\alpha} + \widetilde{W}_\alpha W_0^{-1}\Psi_{T,0}$.

We first consider the term $\widetilde{\Psi}_{T,\alpha}$. By applying Cauchy-Schwarz twice, we obtain:

$$\begin{aligned} \left\|E\left(\widetilde{\Psi}_{T,\alpha}\right)\right\| &= \left\|E\left\langle\left(K_{\alpha T}^{-1}-K_\alpha^{-1}\right)G,\widehat{h}_T\right\rangle\right\| \\ &\leq E\left(\left\|\left(K_{\alpha T}^{-1}-K_\alpha^{-1}\right)G\right\|\left\|\widehat{h}_T\right\|\right) \leq \sqrt{E\left(\left\|\left(K_{\alpha T}^{-1}-K_\alpha^{-1}\right)G\right\|^2\right)E\left(\left\|\widehat{h}_T\right\|^2\right)}. \end{aligned}$$

Using the fact that $h_t\left(\tau,\theta\right)$ is a martingale difference sequence and is bounded, we obtain:

$$\begin{aligned} E\left(\left\|\widehat{h}_T\right\|^2\right) &= E\left(\int \widehat{h}_T\left(\tau,\theta\right)\overline{\widehat{h}}_T\left(\tau,\theta\right)\pi\left(\tau\right)d\tau\right) \\ &= \frac{1}{T}E\left(\int h_t\left(\tau,\theta\right)\overline{h}_t\left(\tau,\theta\right)\pi\left(\tau\right)d\tau\right) = O(T^{-1}). \end{aligned} \tag{13}$$

Next, using (6) and (7), we obtain:

$$\begin{aligned} E\left(\left\|\left(K_{\alpha T}^{-1}-K_\alpha^{-1}\right)G\right\|^2\right) &\leq E\left(\left\|\left(K_{\alpha T}^{-1}-K_\alpha^{-1}\right)K\right\|^2\right)\left\|K^{-1}G\right\|^2 \\ &\leq E\left(\left\|\left(K_T^2+\alpha I\right)^{-1}\left(K_T-K\right)K\right\|^2\right)\left\|K^{-1}G\right\|^2 \qquad (14) \\ &\quad +E\left(\left\|\left(K_T^2+\alpha I\right)^{-1}-\left(K^2+\alpha I\right)^{-1}\right\|^2\right)\|K\|^4\left\|K^{-1}G\right\|^2. \qquad (15) \end{aligned}$$

Hence:

$$\begin{aligned} (14) &= E\left(\left\|\left(K_T^2+\alpha I\right)^{-1}\left(K_T-K\right)K\right\|^2\right)\left\|K^{-1}G\right\|^2 \\ &\leq \alpha^{-2}E\left(\|K_T-K\|^2\right)\|K\|^2\left\|K^{-1}G\right\|^2 = O(\alpha^{-2}T^{-1}), \end{aligned}$$

where $E\left(\|K_T - K\|^2\right) = O(T^{-1})$ follows from Carrasco and Florens (2000, Theorem 4, p. 825). For (15), we use the fact that $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$ to obtain:

$$\begin{aligned}
(15) &= E\left(\left\|(K_T^2 + \alpha I)^{-1} - (K^2 + \alpha I)^{-1}\right\|^2\right)\|K\|^4\left\|K^{-1}G\right\|^2 \\
&\leq E\left(\left\|(K_T^2 + \alpha I)^{-1}\right\|\left\|K_T^2 - K^2\right\|^2\left\|(K^2 + \alpha I)^{-1}\right\|\right)\|K\|^4\left\|K^{-1}G\right\|^2 \\
&\leq \alpha^{-2}E\left(\left\|K_T^2 - K^2\right\|^2\right)\|K\|^4\left\|K^{-1}G\right\|^2 \\
&\leq \alpha^{-2}E\left(\|K_T - K\|^2\|K_T + K\|^2\right)\|K\|^4\left\|K^{-1}G\right\|^2
\end{aligned}$$

By the triangular inequality, $\|K_T + K\| \leq \|K_T\| + \|K\|$. Hence:

$$(15) \leq \alpha^{-2}E\left[\|K_T - K\|^2(\|K_T\| + \|K\|)^2\right]\|K\|^4\left\|K^{-1}G\right\|^2.$$

From (1) in Appendix A, we know that $|k(\tau_1, \tau_2)|^2 \leq 4$. Similarly, $\widehat{k}_T(\tau_1, \tau_2, \widehat{\theta}^1)$ is bounded such that $\left|\widehat{k}_T(\tau_1, \tau_2, \widehat{\theta}^1)\right|^2 \leq 4$. Hence:

$$\begin{aligned}
\|K\| &\leq \sqrt{\int\int |k(\tau_1, \tau_2)|^2 \pi(\tau_1)\pi(\tau_2)\, d\tau_1 d\tau_2} \leq 2 \\
\|K_T\| &\leq \sqrt{\int\int \left|\widehat{k}_T(\tau_1, \tau_2, \widehat{\theta}^1)\right|^2 \pi(\tau_1)\pi(\tau_2)\, d\tau_1 d\tau_2} \leq 2
\end{aligned}$$

Consequently:

$$(15) \leq 16\alpha^{-2}E\left(\|K_T - K\|^2\right)\|K\|^4\left\|K^{-1}G\right\|^2 = O(\alpha^{-2}T^{-1}).$$

Finally,

$$\left\|E\left(\widetilde{\Psi}_{T,\alpha}\right)\right\| = \sqrt{O(\alpha^{-2}T^{-1}) \times O(T^{-1})} = O(\alpha^{-1}T^{-1}). \tag{16}$$

We now consider the term $\widetilde{W}_\alpha W_0^{-1}\Psi_{T,0}$. Again, using Cauchy-Schwarz twice leads to:

$$\left\|E\left(\widetilde{W}_\alpha W_0^{-1}\Psi_{T,0}\right)\right\| \leq E\left(\left\|\widetilde{W}_\alpha\right\|\left\|W_0^{-1}\Psi_{T,0}\right\|\right) \leq \sqrt{E\left(\left\|\widetilde{W}_\alpha\right\|^2\right)E\left(\left\|W_0^{-1}\Psi_{T,0}\right\|^2\right)}.$$

We have:

$$\begin{aligned}
E\left(\left\|W_0^{-1}\Psi_{T,0}\right\|^2\right) &= E\left(\left\|W_0^{-1}\left\langle K^{-1}G, \widehat{h}_T(.,\theta_0)\right\rangle\right\|^2\right) \leq E\left(\left\|W_0^{-1}\right\|^2\left\|(K^{-1}G)\right\|^2\left\|\widehat{h}_T(.,\theta_0)\right\|^2\right) \\
&= \left\|W_0^{-1}\right\|^2\left\|(K^{-1}G)\right\|^2 E\left(\left\|\widehat{h}_T(.,\theta_0)\right\|^2\right) = O(T^{-1}),
\end{aligned}$$

where $E\left(\left\|\widehat{h}_T(.,\theta_0)\right\|^2\right) = O(T^{-1})$ follows from (13). Next:

$$
\begin{aligned}
E\left(\left\|\widetilde{W}_\alpha\right\|^2\right) &= E\left(\left\|\left\langle\left(K_{\alpha T}^{-1}-K_\alpha^{-1}\right)G,G\right\rangle\right\|^2\right) \leq E\left(\left\|\left(K_{\alpha T}^{-1}-K_\alpha^{-1}\right)G\right\|^2\|G\|^2\right) \\
&= \|G\|^2 E\left(\left\|\left(K_{\alpha T}^{-1}-K_\alpha^{-1}\right)G\right\|^2\right) = O(\alpha^{-2}T^{-1}),
\end{aligned}
$$

where the rate follows from (14) and (15). Hence, by the Cauchy-Schwarz inequality,

$$
\begin{aligned}
\left\|E\left(\widetilde{W}_\alpha W_0^{-1}\Psi_{T,0}\right)\right\| &\leq \sqrt{E\left(\left\|\widetilde{W}_\alpha\right\|^2\right)E\left(\left\|W_0^{-1}\Psi_{T,0}\right\|^2\right)} \\
&= \sqrt{O(\alpha^{-2}T^{-1})O(T^{-1})} = O(\alpha^{-1}T^{-1}). \qquad (17)
\end{aligned}
$$

By putting (16) and (17) together, we find $E\left[\Delta_3\right] = O_p\left(\alpha^{-1}T^{-1}\right)$ so that the squared bias satisfies:

$$
TBias.Bias' = O\left(\alpha^{-2}T^{-1}\right).
$$

Higher Order Variance

The dominant terms in the higher order variance are

$$
Cov\left(\Delta_1,\Delta_2\right) + Var\left(\Delta_2\right) + Cov\left(\Delta_1,\Delta_3\right).
$$

We first consider $Cov\left(\Delta_1,\Delta_2\right)$:

$$
Cov\left(\Delta_1,\Delta_2\right) = W_0^{-1}E\left[\Psi_{T,0}\Psi_{T,\alpha}(\theta_0)'\right]W_0^{-1} - W_0^{-1}E\left[\Psi_{T,0}\Psi_{T,0}'\right]W_0^{-1}W_\alpha W_0^{-1}.
$$

From Lemma 16, we have:

$$
E\left[\Psi_{T,0}\Psi_{T,\alpha}'\right] = \frac{1}{T}\left\langle\left(K_\alpha^{-1}-K^{-1}\right)G,G\right\rangle = W_\alpha.
$$

and $E\left[\Psi_{T,0}\Psi_{T,0}'\right] = W_0$. Hence,

$$
Cov\left(\Delta_1,\Delta_2\right) = \frac{1}{T}W_0^{-1}W_\alpha W_0^{-1} - \frac{1}{T}W_0^{-1}W_0W_0^{-1}W_\alpha W_0^{-1} = 0.
$$

Now we consider the term $Cov\left(\Delta_1,\Delta_3\right)$:

$$
Cov\left(\Delta_1,\Delta_3\right) = W_0^{-1}E\left(\Psi_{T,0}\widetilde{\Psi}_{T,\alpha}'\right)W_0^{-1} - W_0^{-1}E\left(\Psi_{T,0}\Psi_{T,0}'W_0^{-1}\widetilde{W}_\alpha\right)W_0^{-1}.
$$

We first consider $E\left[\Psi_{T,0}\widetilde{\Psi}'_{T,\alpha}\right]$. By the Cauchy-Schwarz inequality:

$$\begin{aligned}\left\|E\left(\Psi_{T,0}\widetilde{\Psi}'_{T,\alpha}\right)\right\| &\leq \sqrt{E\left(\left\|\Psi_{T,0}\right\|^2\right)E\left(\left\|\widetilde{\Psi}'_{T,\alpha}\right\|^2\right)} \\ &= \sqrt{E\left(\left\|\left\langle K^{-1}G,\widehat{h}_T\right\rangle\right\|^2\right)E\left(\left\|\left\langle\left(K_{\alpha T}^{-1}-K_{\alpha}^{-1}\right)G,\widehat{h}_T\right\rangle\right\|^2\right)}\end{aligned}$$

Hence we have:

$$E\left(\left\|\left\langle K^{-1}G,\widehat{h}_T\right\rangle\right\|^2\right)\leq\left\|K^{-1}G\right\|^2E\left(\left\|\widehat{h}_T\right\|^2\right)=O(T^{-1})$$

Also,

$$\begin{aligned}E\left(\left\|\left\langle\left(K_{\alpha T}^{-1}-K_{\alpha}^{-1}\right)G,\widehat{h}_T\right\rangle\right\|^2\right) &= E\left(\left\|\left(K_{\alpha T}^{-1}-K_{\alpha}^{-1}\right)G\right\|^2\left\|\widehat{h}_T\right\|^2\right) \\ &\leq \sqrt{E\left(\left\|\left(K_{\alpha T}^{-1}-K_{\alpha}^{-1}\right)G\right\|^4\right)E\left(\left\|\widehat{h}_T\right\|^4\right)}\end{aligned}$$

We first consider $\left\|\widehat{h}_T\right\|^4$:

$$\begin{aligned}\left\|\widehat{h}_T\right\|^4 &= \left(\int\widehat{h}_T\overline{\widehat{h}}_T\pi\left(\tau\right)d\tau\right)^2=\frac{1}{T^4}\left(\int\sum_{t=1}^{T}h_t\overline{h}_t\pi\left(\tau\right)d\tau+\int\sum_{t\neq s}^{T}h_t\overline{h}_s\pi\left(\tau\right)d\tau\right)^2 \\ &= \frac{1}{T^4}\left(\sum_{t=1}^{T}\int h_t\overline{h}_t\pi\left(\tau\right)d\tau\right)^2+\frac{1}{T^4}\left(\sum_{t\neq s}^{T}\int h_t\overline{h}_s\pi\left(\tau\right)d\tau\right)^2 \\ &\quad +2\left(\frac{1}{T^2}\sum_{t=1}^{T}\int h_t\overline{h}_t\pi\left(\tau\right)d\tau\right)\left(\frac{1}{T^2}\sum_{t\neq s}^{T}\int h_t\overline{h}_s\pi\left(\tau\right)d\tau\right)\end{aligned}$$

Consider the first squared term of $\left\|\widehat{h}_T\right\|^4$:

$$\begin{aligned}E\left[\frac{1}{T^4}\left(\sum_{t=1}^{T}\int h_t\overline{h}_t\pi\left(\tau\right)d\tau\right)^2\right] &= \frac{T}{T^4}E\left[\left(\int h_t\overline{h}_t\pi\left(\tau\right)d\tau\right)^2\right]+\frac{T(T-1)}{T^4}\left[E\left(\int h_t\overline{h}_t\pi\left(\tau\right)d\tau\right)\right]^2 \\ &= O(T^{-2})\end{aligned}$$

The second squared term leads to:

$$\begin{aligned}
& E\left[\frac{1}{T^4}\left(\sum_{t\neq s}^{T}\int h_t\overline{h}_s\pi\left(\tau\right)d\tau\right)^2\right] \\
= \; & E\left[\frac{1}{T^4}\sum_{t\neq s}^{T}\left(\int h_t\overline{h}_s\pi\left(\tau\right)d\tau\right)^2\right]+E\left[\frac{1}{T^4}\sum_{t\neq s,l\neq j,(t,s)\neq(l,j)}^{T}\left(\int h_t\overline{h}_s\pi\left(\tau\right)d\tau\right)\left(\int h_l\overline{h}_j\pi\left(\tau\right)d\tau\right)\right] \\
= \; & \frac{T(T-1)}{T^4}E\left[\left(\int h_t\overline{h}_s\pi\left(\tau\right)d\tau\right)^2\right]=O(T^{-2}),\text{ for } t\neq s.
\end{aligned}$$

As the h_ts are uncorrelated, the cross-term is equal to zero:

$$\left[\left(\frac{1}{T^2}\sum_{t=1}^{T}\int h_t\overline{h}_t\pi\left(\tau\right)d\tau\right)\left(\frac{1}{T^2}\sum_{t\neq s}^{T}\int h_t\overline{h}_s\pi\left(\tau\right)d\tau\right)\right]=0$$

In total, we obtain: $E\left(\left\|\widehat{h}_T\right\|^4\right)=O(T^{-2})$.

We now consider $E\left(\left\|\left(K_{\alpha T}^{-1}-K_{\alpha}^{-1}\right)G\right\|^4\right)$. Using the same decomposition as in (14) and (15) leads to:

$$\begin{aligned}
E\left(\left\|\left(K_{\alpha T}^{-1}-K_{\alpha}^{-1}\right)G\right\|^4\right) &\leq E\left(\left\|\left(K_{\alpha T}^{-1}-K_{\alpha}^{-1}\right)K\right\|^4\right)\left\|K^{-1}G\right\|^4 \\
&\leq E\left(\left\|\left(K_T^2+\alpha I\right)^{-1}\left(K_T-K\right)K\right\|^4\right)\left\|K^{-1}G\right\|^4 \qquad (18) \\
&\quad +E\left(\left\|\left(K_T^2+\alpha I\right)^{-1}-\left(K^2+\alpha I\right)^{-1}\right\|^4\right)\|K\|^8\left\|K^{-1}G\right\|^4, \qquad (19)
\end{aligned}$$

Hence:

$$(18)=E\left(\left\|\left(K_T^2+\alpha I\right)^{-1}\left(K_T-K\right)K\right\|^4\right)\left\|K^{-1}G\right\|^4\leq\alpha^{-4}E\left(\|K_T-K\|^4\right)\|K\|^4\left\|K^{-1}G\right\|^4$$

For (19), we use $A^{-1}-B^{-1}=A^{-1}\left(B-A\right)B^{-1}$ to obtain:

$$\begin{aligned}
(19) &= E\left(\left\|\left(K_T^2+\alpha I\right)^{-1}-\left(K^2+\alpha I\right)^{-1}\right\|^4\right)\|K\|^8\left\|K^{-1}G\right\|^4 \\
&\leq E\left(\left\|\left(K_T^2+\alpha I\right)^{-1}\right\|^2\left\|K_T^2-K^2\right\|^4\left\|\left(K^2+\alpha I\right)^{-1}\right\|^2\right)\|K\|^8\left\|K^{-1}G\right\|^4 \\
&\leq \alpha^{-4}E\left(\left\|K_T^2-K^2\right\|^4\right)\|K\|^8\left\|K^{-1}G\right\|^4 \\
&\leq \alpha^{-4}E\left(\|K_T-K\|^4\|K_T+K\|^4\right)\|K\|^8\left\|K^{-1}G\right\|^4
\end{aligned}$$

By the triangular inequality:

$$\begin{aligned}(19) &\leq \alpha^{-4}E\left(\|K_T-K\|^4\left(\|K_T\|+\|K\|\right)^4\right)\|K\|^8\left\|K^{-1}G\right\|^4 \\ &\leq 256\alpha^{-4}E\left(\|K_T-K\|^4\right)\|K\|^8\left\|K^{-1}G\right\|^4,\end{aligned}$$

due to $\|K_T\| \leq 2$ and $\|K\| \leq 2$.

The rates of (18) and (19) depend on the rate of $E\left(\|K_T-K\|^4\right)$.

$$\begin{aligned}\|K_T-K\|^2 &\leq \int\int\left|\frac{1}{T}\sum_{t=1}^{T}\chi_t(\tau_1,\tau_2)\right|^2\pi(\tau_1)\pi(\tau_2)\,d\tau_1 d\tau_2 \\ &= \frac{1}{T^2}\sum_{t=1}^{T}\int\int|\chi_t(\tau_1,\tau_2)|^2\pi(\tau_1)\pi(\tau_2)\,d\tau_1 d\tau_2 \qquad (20) \\ &\quad +\frac{1}{T^2}\sum_{t\neq l}^{T}\int\int\chi_t(\tau_1,\tau_2)\overline{\chi_l(\tau_1,\tau_2)}\pi(\tau_1)\pi(\tau_2)\,d\tau_1 d\tau_2 \qquad (21)\end{aligned}$$

where $\chi_t(\tau_1,\tau_2) = k_t(\tau_1,\tau_2,\widehat{\theta}^1) - k(\tau_1,\tau_2)$. Hence

$$E\left(\|K_T-K\|^4\right) \leq E\left([(20)]^2\right) + 2E\left([(20)]\,[(21)]\right) + E\left([(21)]^2\right)$$

Because $E\left([(20)]\,[(21)]\right) \leq \sqrt{E\left([(20)]^2\right)\left(E\,[(21)]^2\right)}$, we only need to check the rates of the squared terms. We have:

$$\begin{aligned}&E\left([(20)]^2\right) = \tfrac{T}{T^4}E\left[\left(\int\int|\chi_t(\tau_1,\tau_2)|^2\pi(\tau_1)\pi(\tau_2)\,d\tau_1 d\tau_2\right)^2\right] \\ &+\tfrac{T(T-1)}{T^4}E\left[\left(\int\int|\chi_t(\tau_1,\tau_2)|^2\pi(\tau_1)\pi(\tau_2)\,d\tau_1 d\tau_2\right)\left(\int\int|\chi_l(\tau_1,\tau_2)|^2\pi(\tau_1)\pi(\tau_2)\,d\tau_1 d\tau_2\right)\right], \text{ for } l\neq t.\end{aligned}$$

Hence $E\left([(20)]^2\right) = O(T^{-2})$. Next:

$$\begin{aligned}&E\left([(21)]^2\right) \\ =\ &\frac{1}{T^4}\sum_{t\neq l}^{T}E\left[\left(\int\int\chi_t(\tau_1,\tau_2)\overline{\chi_l(\tau_1,\tau_2)}\pi(\tau_1)\pi(\tau_2)\,d\tau_1 d\tau_2\right)^2\right] \\ &+\frac{1}{T^4}\sum_{t\neq l,n\neq j,(t,l)\neq(n,j)}^{T}E\left[\left(\int\int\chi_t\overline{\chi_l}\pi(\tau_1)\pi(\tau_2)\,d\tau_1 d\tau_2\right)\left(\int\int\chi_n\overline{\chi_j}\pi(\tau_1)\pi(\tau_2)\,d\tau_1 d\tau_2\right)\right]\end{aligned}$$

with $\chi_t \equiv \chi_t(\tau_1,\tau_2)$. Due to the m.d.s property, the last term has expectation zero. Hence:

$$E\left([(21)]^2\right) = \frac{T(T-1)}{T^4}E\left[\left(\int\int\chi_t(\tau_1,\tau_2)\overline{\chi_l(\tau_1,\tau_2)}\pi(\tau_1)\pi(\tau_2)\,d\tau_1 d\tau_2\right)^2\right] = O(T^{-2}).$$

By putting these together, we obtain $E\left(\|K_T - K\|^4\right) = O(T^{-2})$ so that:

$$\begin{aligned} E\left(\left\|\left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right) G\right\|^4\right) &\leq (18)+(19) = O\left(\alpha^{-4}T^{-2}\right) \text{ and} \\ E\left(\left\|\left\langle\left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right) G, \widehat{h}_T\right\rangle\right\|^2\right) &\leq \sqrt{E\left(\left\|\left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right) G\right\|^4\right) E\left(\left\|\widehat{h}_T\right\|^4\right)} \\ &= \sqrt{O\left(\alpha^{-4}T^{-2}\right) \times O\left(T^{-2}\right)} = O\left(\alpha^{-2}T^{-2}\right) \end{aligned}$$

In total:

$$\begin{aligned} \left\|E\left(\Psi_{T,0}\widetilde{\Psi}'_{T,\alpha}\right)\right\| &\leq \sqrt{E\left(\left\|\left\langle K^{-1}G, \widehat{h}_T\right\rangle\right\|^2\right) E\left(\left\|\left\langle\left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right) G, \widehat{h}_T\right\rangle\right\|^2\right)} \\ &= \sqrt{O(T^{-1}) \times O\left(\alpha^{-2}T^{-2}\right)} = O\left(\alpha^{-1}T^{-3/2}\right) \end{aligned}$$

We now check the rate of the second term of $Cov\left(\Delta_1, \Delta_3\right)$:

$$\left\|E\left(\Psi_{T,0}\Psi'_{T,0}W_0^{-1}\widetilde{W}_\alpha\right)\right\| \leq \sqrt{E\left(\left\|\Psi_{T,0}\Psi'_{T,0}\right\|^2\right) E\left(\left\|W_0^{-1}\widetilde{W}_\alpha\right\|^2\right)}$$

We first consider $E\left(\left\|\Psi_{T,0}\Psi'_{T,0}\right\|^2\right)$. By the Cauchy-Schwarz inequality:

$$E\left(\left\|\Psi_{T,0}\Psi'_{T,0}\right\|^2\right) = E\left(\left\|\left\langle K^{-1}G, \widehat{h}_T\right\rangle\left\langle K^{-1}G, \widehat{h}_T\right\rangle'\right\|^2\right) \leq \left\|K^{-1}G\right\|^4 E\left(\left\|\widehat{h}_T\right\|^4\right) = O(T^{-2})$$

For the second term, we have:

$$\begin{aligned} E\left(\left\|W_0^{-1}\widetilde{W}_\alpha\right\|^2\right) &= \|W_0\|^{-2} E\left(\left\|\left\langle\left(K_\alpha^{-1} - K^{-1}\right) G, G\right\rangle\right\|^2\right) \\ &\leq \|W_0\|^{-2} \|G\|^2 E\left(\left\|\left(K_\alpha^{-1} - K^{-1}\right) G\right\|^2\right) = O(\alpha^{-2}T^{-1}), \end{aligned}$$

according to (14)-(15). Hence,

$$\left\|E\left(\Psi_{T,0}\Psi'_{T,0}W_0^{-1}\widetilde{W}_\alpha\right)\right\| \leq \sqrt{O(T^{-2}) \times O(\alpha^{-2}T^{-1})} = O(\alpha^{-1}T^{-3/2})$$

It now remains to find the rate of $Var\left(\Delta_2\right)$. We recall that $\Delta_2 = -W_0^{-1}\Psi_{T,\alpha} + W_0^{-1}W_\alpha W_0^{-1}\Psi_{T,0}$. We have

$$\begin{aligned} Var\left(\Delta_2\right) &= W_0^{-1}E\left[\Psi_{T,\alpha}\Psi'_{T,\alpha}\right]W_0^{-1} - W_0^{-1}E\left[\Psi_{T,\alpha}\Psi'_{T,0}\right]W_0^{-1}W_\alpha W_0^{-1} \\ &\quad -W_0^{-1}W_\alpha W_0^{-1}E\left[\Psi_{T,0}\Psi'_{T,\alpha}\right]W_0^{-1} + W_0^{-1}W_\alpha W_0^{-1}E\left[\Psi_{T,0}\Psi'_{T,0}\right]W_0^{-1}W_\alpha W_0^{-1}. \end{aligned}$$

Replacing $E\left[\Psi_{T,0}\Psi'_{T,\alpha}\right] = \frac{1}{T}W_\alpha$ and $E\left[\Psi_{T,0}\Psi'_{T,0}\right] = \frac{1}{T}W_0$, we see immediately that the last two terms

cancel out so that

$$Var\left(\Delta_2\right) = W_0^{-1} E\left[\Psi_{T,\alpha}\Psi_{T,\alpha}'\right] W_0^{-1} - W_0^{-1} W_\alpha W_0^{-1} W_\alpha W_0^{-1}.$$

For the first term of $Var\left(\Delta_2\right)$, we use Lemma 16 to obtain:

$$\begin{aligned} E\left[\Psi_{T,\alpha}\Psi_{T,\alpha}'\right] &= E\left[\left\langle \left(K_\alpha^{-1} - K^{-1}\right) G, \widehat{h}_T \right\rangle \left\langle \left(K_\alpha^{-1} - K^{-1}\right) G, \widehat{h}_T \right\rangle\right] \\ &= \frac{1}{T}\left\langle \left(K_\alpha^{-1} - K^{-1}\right) G, \left(K_\alpha^{-1} - K^{-1}\right) KG \right\rangle = \sum_j \left(\frac{\mu_j}{\mu_j^2 + \alpha} - \frac{1}{\mu_j}\right)^2 \mu_j \left\langle G, \phi_j \right\rangle^2 \\ &= \sum_j \left(\frac{\mu_j}{\mu_j^2 + \alpha} - \frac{1}{\mu_j}\right)^2 \mu_j^{2\beta+1} \frac{\left\langle G, \phi_j \right\rangle^2}{\lambda_j^{2\beta}} \leq \sum_j \frac{\left\langle G, \phi_j \right\rangle^2}{\mu_j^{2\beta}} \sup_{\mu \leq \mu_1} \left(\frac{\mu}{\mu^2 + \alpha} - \frac{1}{\mu}\right)^2 \mu^{2\beta+1}. \end{aligned}$$

We focus on the square-root of $\left(\frac{\mu}{\mu^2+\alpha} - \frac{1}{\mu}\right)^2 \mu^{2\beta+1}$, namely:

$$\sup_{\mu \leq \mu_1} \left(\frac{1}{\mu} - \frac{\mu}{\mu^2 + \alpha}\right) \mu^{(2\beta+1)/2} = \sup_{\mu \leq \mu_1} \left(1 - \frac{\mu^2}{\mu^2 + \alpha}\right) \mu^{\beta - 1/2}.$$

Case where $\beta \geq 5/2$

$$\sup_{\mu \leq \mu_1} \left(1 - \frac{\mu^2}{\mu^2 + \alpha}\right) \mu^{\beta-1/2} = \alpha \sup_{\mu \leq \mu_1} \frac{\mu^{\beta-1/2}}{\mu^2 + \alpha} \leq \alpha \sup_{\mu \leq \mu_1} \mu^{\beta - 5/2} \leq \alpha \mu_1^{\beta - 5/2}.$$

Case where $\beta < 5/2$

We apply the change of variable $x = \alpha/\mu^2$ and obtain

$$\sup_{\mu \leq \mu_1} \left(1 - \frac{\mu^2}{\mu^2 + \alpha}\right) \mu^{\beta-1/2} = \sup_{x \geq 0} \left(1 - \frac{1}{1+x}\right) \left(\frac{\alpha}{x}\right)^{\frac{\beta-1/2}{2}} = \alpha^{\frac{2\beta-1}{4}} \sup_{x \geq 0} \frac{x}{1+x} x^{-\frac{2\beta-1}{4}}.$$

The function $f\left(x\right) = \frac{x}{1+x} x^{-\frac{2\beta-1}{4}}$ is continuous and hence bounded for x away from 0 and infinity. When x goes to infinity, $f\left(x\right)$ goes to zero because $2\beta - 1 > 0$. When x goes to zero, $f\left(x\right) = \frac{x^{\frac{5-2\beta}{4}}}{1+x}$ goes to zero because $5 - 2\beta > 0$. Hence, $f\left(x\right)$ is bounded on $\mathbb{R}^+$. In conclusion, the rate of convergence of $E\left(\Psi_{T,\alpha}\Psi_{T,\alpha}'\right)$ is given by: $\alpha^{\min\left(2, \frac{2\beta-1}{2}\right)} T^{-1}$. Note that this rate is an equivalent, not a big O.

For the second term of $Var\left(\Delta_2\right)$, we use the fact that $W_\alpha = O\left(\alpha^{\min(1, \frac{2\beta-1}{2})}\right)$ according to Equation (5) in Lemma 14:

$$\begin{aligned} \frac{1}{T} W_0^{-1} W_\alpha W_0^{-1} W_\alpha W_0^{-1} &= \frac{1}{T} \times O\left(1\right) \times O\left(\alpha^{\min(1, \frac{2\beta-1}{2})}\right) \times O\left(1\right) \times O\left(\alpha^{\min(1, \frac{2\beta-1}{2})}\right) \times O\left(1\right) \\ &= O\left(\alpha^{\min(2, 2\beta-1)} T^{-1}\right). \end{aligned}$$

Optimal Rate for α

Note that the bias term $TBias * Bias' = O\left(\alpha^{-2}T^{-1}\right)$ goes to zero faster than the covariance term $TCov\left(\Delta_1, \Delta_3\right) = O\left(\alpha^{-1}T^{-1/2}\right)$. Hence the optimal α is the one that achieves the best trade-off between $TVar\left(\Delta_2\right) \sim \alpha^{\min(2,\beta-\frac{1}{2})}$ which is increasing in α and $TCov\left(\Delta_1, \Delta_3\right)$ which is decreasing in α. We have

$$\alpha^{\min(2,\beta-\frac{1}{2})} = \alpha^{-1}T^{-1/2} \Rightarrow \alpha^* = T^{-\max(\frac{1}{6}, \frac{1}{2\beta+1})}.$$

Note that this rate satisfies $\alpha^{-1}T^{-1/2} = o(1)$.

C Consistency of $\widehat{\alpha}_{TM}\left(\widehat{\theta}^1\right)$

We first prove the following lemma.

Lemma 18 *:Under Assumptions 1 to 5, $\widehat{\theta}_T(\alpha;\theta_0)$ is once continuously differentiable with respect to α and twice continuously differentiable with respect to θ_0 and $\alpha_T(\theta_0)$ is a continuous in θ_0.*

Proof of Lemma 18: The objective function $\widehat{Q}_T\left(\alpha, \theta\right)$ involves the following operator:

$$K_{\alpha T}^{-1}\widehat{h}_T(.,\theta) = \sum_{j=1}^{T} \frac{\widehat{\mu}_j}{\alpha + \widehat{\mu}_j^2} \left\langle \widehat{h}_T(.,\theta), \widehat{\phi}_j \right\rangle \widehat{\phi}_j$$

where $\widehat{\phi}_j$ is the eigenfunction of K_T associated with the eigenvalue $\widehat{\mu}_j$. By assumption 3, the moment function $\widehat{h}_T(.,\theta)$ is three times continuously differentiable with respect to θ, the argument with respect to which we minimize the objective function of the CGMM. By assumption 5, $x_t = x\left(x_{t-1}, \theta_0, \varepsilon_t\right)$ where r is three times continuously differentiable with respect to θ_0 (the true unknown parameter) and ε_t is an IID white noise whose distribution does not depend on θ_0. Thus as an exponential function of x_t, the moment function is also three times continuously differentiable with respect to θ_0. Thus Assumptions 3 and 5 imply that the objective function of the CGMM is three times continuously differentiable with respect to θ and θ_0. Now we turn our attention toward the differentiability with respect to α. It is easy to check that

$$\frac{\partial^3 K_{\alpha T}^{-1}\widehat{h}_T(.,\theta)}{\partial \alpha^3} = \widetilde{K}_{\alpha T}\widehat{h}_T(.,\theta)$$

where $\widetilde{K}_{\alpha T} \equiv -\left(K_T^2 + \alpha_T I\right)^{-2} K_T$ which is well defined on $L^2\left(\pi\right)$ for α_T fixed. When α_T goes to zero, we have to be more careful. We check that $\left|\left\langle \widetilde{K}_{\alpha T}\widehat{h}_T(.,\theta), \widehat{h}_T(.,\theta)\right\rangle\right|$ is bounded. We have

$$\begin{aligned}
\left|\left\langle \widetilde{K}_{\alpha T}\widehat{h}_T(.,\theta), \widehat{h}_T(.,\theta)\right\rangle\right| &\leq \left\|\widetilde{K}_{\alpha T}\widehat{h}_T(.,\theta)\right\| \left\|\widehat{h}_T(.,\theta)\right\| \leq \left\|\left(K_T^2 + \alpha_T I\right)^{-2} K_T\right\| \left\|\widehat{h}_T(.,\theta)\right\|^2 \\
&= \underbrace{\left\|\left(K_T^2 + \alpha_T I\right)^{-3/2}\right\|}_{\leq \alpha_T^{-3/2}} \underbrace{\left\|\left(K_T^2 + \alpha_T I\right)^{-1/2} K_T\right\|}_{\leq 1} \underbrace{\left\|\widehat{h}_T(.,\theta)\right\|^2}_{=O_p(T^{-1})} \\
&= O_p\left(\alpha_T^{-3/2}T^{-1}\right) = o_p\left(1\right),
\end{aligned}$$

where the last equality follows from Theorem 2(ii). This shows that $\widehat{Q}_T(\alpha,\theta)$ is once continuously differentiable with respect to α and three times continuously differentiable with respect to θ. By the implicit function theorem, $\widehat{\theta}_T(\alpha;\theta_0) = \arg\min_{\theta}\widehat{Q}_T(\alpha,\theta)$ is once continuously differentiable with respect to α and twice continuously differentiable w.r.t. θ_0. The MSE $\widehat{\theta}_T(\alpha;\theta_0)$ is an expectation of a quadratic function in $\widehat{\theta}_T(\alpha;\theta_0)$. Hence $\Sigma_T(\alpha;\theta_0)$ is also once continuously differentiable w.r.t. α and twice continuously differentiable w.r.t. θ_0. Finally, the Maximum theorem implies that $\alpha_T(\theta_0) = \arg\min_{\alpha\in[0,1]}\Sigma_T(\alpha;\theta_0,\nu)$ is continuous w.r.t. θ_0.■

Proof of Theorem 3: Using Assumption 6, we see that $\frac{\alpha_T(\widehat{\theta}^1)}{\alpha_T(\theta_0)} = \frac{c(\widehat{\theta}^1)}{c(\theta_0)}$. Moreover by Lemma 18, $\alpha_T(\theta)$ and hence $c(\theta)$ are continuous functions of θ. Since $\widehat{\theta}^1$ is a consistent estimator of θ_0, the continuous mapping theorem implies that $\frac{c(\widehat{\theta}^1)}{c(\theta_0)} \xrightarrow{P} 1$ as $T\to\infty$.■

Proof of Theorem 4: Here, we consider the expression of the MSE given by (15) but the same proof can be easily adapted to the expressions given by (16) and (17). Consider $\widehat{\Sigma}_{TM}(\alpha,\theta_0,\nu) = \frac{T}{(1-\nu)M}\sum_{j=1}^{M}\xi_{j,T}(\alpha)\,1\left(\xi_{jT}(\alpha) < \widehat{n}_{\nu,TM}\right)$ where $\xi_{j,T}(\alpha)\equiv\xi_{j,T}(\alpha,\theta_0)$ is IID (across j) and continuous in α. We have:

$$\begin{aligned}&\widehat{\Sigma}_{TM}(\alpha,\theta_0,\nu) - \Sigma_T(\alpha,\theta_0,\nu)\\ =\;& \frac{T}{(1-\nu)M}\sum_{j=1}^{M}\left[\xi_{j,T}(\alpha)\,1\left(\xi_{jT}(\alpha) < \widehat{n}_{\nu,TM}\right) - E\left(\xi_{j,T}(\alpha)\,1\left(\xi_{jT}(\alpha) < n_{\nu,T}\right)\right)\right],\end{aligned}$$

where $n_{\nu,T} = \lim_{M\to\infty} n_{\nu,TM}$.

If we can show that there exists a function $b_T > 0$ independent of α such that

$$\left\|\frac{\partial\xi_{j,T}(\alpha)}{\partial\alpha}\right\| < b_T, \tag{22}$$

and $E(b_T) < \infty$, then, by Lemma 2.4 of Newey and McFadden (1994), we would have

$$\sup_{\alpha\in[0,1]}\left|\widehat{\Sigma}_{TM}(\alpha,\theta_0,\nu) - \Sigma_T(\alpha,\theta_0,\nu)\right| = O_p\left(M^{-1/2}\right).$$

and it would follow from Theorem 2.1 of Newey and McFadden (1994) that $\widehat{\alpha}_{TM}(\theta_0) - \alpha_T(\theta_0) = O_p\left(M^{-1/2}\right)$. This would imply that $\frac{\widehat{\alpha}_{TM}(\theta_0)}{\alpha_T(\theta_0)} - 1 = O_p\left(M^{-1/2}\right)$, given that $\alpha_T(\theta_0)$ is bounded away from zero when T is fixed.

Let ξ_T be any of the $\xi_{j,T}, j = 1,...,M$. To prove inequality (22), we first compute:

$$\frac{\partial\xi_T(\alpha)}{\partial\alpha} = 2\frac{\partial\widehat{\theta}_T(\alpha,\theta_0)}{\partial\alpha}'\left(\widehat{\theta}_T(\alpha,\theta_0) - \theta_0\right)$$

where by the implicit function theorem:

$$\frac{\partial\widehat{\theta}_T(\alpha,\theta_0)}{\partial\alpha}=-\left(\frac{\partial^2\widehat{Q}_T(\alpha,\theta)}{\partial\theta\partial\theta'}\right)^{-1}\frac{\partial^2\widehat{Q}_T(\alpha,\theta)}{\partial\theta\partial\alpha}.$$

The expressions involved are:

$$\begin{aligned}\frac{\partial^2\widehat{Q}_T(\alpha,\theta)}{\partial\theta\partial\theta'} &= \left\langle K_{\alpha T}^{-1}\widehat{G}_T(.,\theta_0),\widehat{G}_T(.,\theta_0)\right\rangle+\left\langle K_{\alpha T}^{-1}\widehat{H}_T(.,\theta_0),\widehat{h}_T(.,\theta_0)\right\rangle,\\ \frac{\partial^2\widehat{Q}_T(\alpha,\theta)}{\partial\theta\partial\alpha} &= \left\langle K_{\alpha T}^{*}\widehat{G}_T(.,\theta_0),\widehat{h}_T(.,\theta_0)\right\rangle+\left\langle K_{\alpha T}^{*}\widehat{h}_T(.,\theta_0),\widehat{G}_T(.,\theta_0)\right\rangle\end{aligned}$$

and $K_{\alpha T}^{*}\equiv-\left(K_T^2+\alpha I\right)^{-2}K_T$. Next recall that for fixed T, $\alpha_T(\theta_0)$ is bounded away from zero so that there exists a sequence $\underline{\alpha}_T$ such that $\alpha_T(\theta_0)\geq\underline{\alpha}_T$ for all T.

Hence, the minimization problem for the selection of α may be re-written as $\alpha(\theta_0)=\underset{\alpha\in[\underline{\alpha}_T,1]}{\arg\min}\Sigma_T(\alpha;\theta_0,\nu)$, so that $\Sigma_T(\alpha;\theta_0,\nu)$ is a bounded function of α on the choice set. Because $\frac{\partial^2\widehat{Q}_T(\alpha,\theta;\theta_0)}{\partial\theta\partial\theta'}$ and $\frac{\partial^2\widehat{Q}_T(\alpha,\theta;\theta_0)}{\partial\theta\partial\alpha}$ are continuous with respect to α, $\frac{\partial\xi_T(\alpha)}{\partial\alpha}$ is also continuous with respect to α. Hence, we indeed have $\|b_T(\overline{\alpha}_T)\|=\left\|\frac{\partial\xi_T(\overline{\alpha}_T)}{\partial\alpha}\right\|<\infty$ where $\overline{\alpha}_T=\underset{\alpha\in[\underline{\alpha}_T,1]}{\arg\sup}\left\|\frac{\partial\xi_T(\alpha)}{\partial\alpha}\right\|$.■

Proof of Theorem 5: We first make the following decomposition

$$\frac{\widehat{\alpha}_{TM}(\widehat{\theta}^1)}{\alpha_T(\theta_0)}-1=\left(\frac{\alpha_T(\widehat{\theta}^1)}{\alpha_T(\theta_0)}-1\right)+\left(\frac{\alpha_T(\widehat{\theta}^1)}{\alpha_T(\theta_0)}-1\right)\left(\frac{\widehat{\alpha}_{TM}(\widehat{\theta}^1)}{\alpha_T(\widehat{\theta}^1)}-1\right)+\left(\frac{\widehat{\alpha}_{TM}(\widehat{\theta}^1)}{\alpha_T(\widehat{\theta}^1)}-1\right).$$

By first letting M go to infinity, we obtain the result of Theorem 4 which has been proved for fix T: $\frac{\widehat{\alpha}_{TM}(\theta_0)}{\alpha_T(\theta_0)}-1=O_p(M^{-1/2})$. Next, we let T go to infinity in order to obtain the result of Theorem 3: $\frac{\alpha_T(\widehat{\theta}^1)}{\alpha_T(\theta_0)}-1=O_p(T^{-1/2})$. The product of $\left(\frac{\alpha_T(\widehat{\theta}^1)}{\alpha_T(\theta_0)}-1\right)$ and $\left(\frac{\widehat{\alpha}_{TM}(\widehat{\theta}^1)}{\alpha_T(\widehat{\theta}^1)}-1\right)$ is negligible with respect to either of the other terms. Thus, it follows that $\frac{\widehat{\alpha}_{TM}(\widehat{\theta}^1)}{\alpha_T(\theta_0)}-1=O_p(T^{-1/2})+O_p(M^{-1/2})$.■

Proof of Theorem 6: The mean value Theorem yields:

$$\widehat{\theta}(\widehat{\alpha}_{TM})-\widehat{\theta}(\alpha_T)=\frac{\partial\widehat{\theta}(\overline{\alpha})}{\partial\alpha}(\widehat{\alpha}_{TM}-\alpha_T),$$

where $\overline{\alpha}$ lies between $\widehat{\alpha}_{TM}$ and α_T and α_T is bounded away from zero, i.e., $\exists\ \underline{\alpha}_T>0:\underline{\alpha}_T\leq\alpha_T\leq1,\ \forall\ T$. From the proof of Theorem 4, we know that $\widehat{\theta}(\alpha)$ is continuously differentiable with respect to α. This implies that:

$$\left\|\frac{\partial\widehat{\theta}(\overline{\alpha})}{\partial\alpha}\right\|<\sup_{\alpha\in[\underline{\alpha}_T,1]}\left\|\frac{\partial\widehat{\theta}(\alpha)}{\partial\alpha}\right\|=O_p(1).$$

Consequently, the rate of $\widehat{\theta}(\widehat{\alpha}_{TM}) - \widehat{\theta}(\alpha_T)$ is determined by the rate at which $\widehat{\alpha}_{TM} - \alpha_T$ converges to zero. We have:

$$\widehat{\alpha}_{TM} - \alpha_T = \alpha_T \left(\frac{\widehat{\alpha}_{TM}}{\alpha_T} - 1 \right) = c(\theta_0) T^{-g(\beta)} \left(\frac{\widehat{\alpha}_{TM}}{\alpha_T} - 1 \right) = O_p(T^{-g(\beta)-1/2}),$$

provided that $M \geq T$. Hence:

$$\sqrt{T} \left(\widehat{\theta}(\widehat{\alpha}_{TM}) - \widehat{\theta}(\alpha_T) \right) = \frac{\partial \widehat{\theta}(\overline{\alpha})}{\partial \alpha} \sqrt{T} (\widehat{\alpha}_{TM} - \alpha_T) = O_p(T^{-g(\beta)}) = o_p(1),$$

which shows that $\sqrt{T} \left(\widehat{\theta}(\widehat{\alpha}_{TM}) - \theta_0 \right)$ and $\sqrt{T} \left(\widehat{\theta}(\alpha_T) - \theta_0 \right)$ have the same asymptotic distribution.

D Numerical algorithms: Computing the objective function of the CGMM

The moment function $h_t(\theta, \tau) \in L^2(\pi)$ for any finite measure π. Hence, we can take $\pi(\tau)$ to be the standard normal density up to a multiplicative constant: $\pi(\tau) = \exp\{-\tau'\tau\}$. We have:

$$K_T \hat{h}_T(\theta, \tau) = \int_{R^d} \widehat{k}_T(s, \tau) \widehat{h}_T(\theta, s) \exp\{-s's\} ds.$$

This integral can be well approximated numerically by using the Gauss-Hermite quadrature. This amounts to find m points $s_1, s_2, ...s_m$ and weighs $\omega_1, \omega_2, ...\omega_m$ such that:

$$\int_{R^d} P(s) \exp\{-s's\} dx = \sum_{k=1}^{m} \omega_k P(s_k)$$

for any polynomial function $P(.)$ of order smaller than or equal to $2m-1$. See for example Liu and Pierce (1994).

If f is differentiable at any order (for example an analytic function), it can be shown that for any positive ε arbitrarily small, there exist m such that:

$$\left| \int_{R^d} f(s) \exp\{-s's\} dx - \sum_{k=1}^{m} \omega_k f(s_k) \right| < \varepsilon.$$

The choice of the quadrature point does not depend on the function f. The quadrature points and weights are determined by solving:

$$\int s^l \exp\{-s^2\} ds = \sum_{k=1}^{n} \omega_k s_k^l \text{ for all } l = 1, ..., 2n-1$$

Applying that method to evaluate the above integral, we get

$$K_T\widehat{h}_T(\theta,\tau) \approx \sum_{k=1}^{m} \omega_k \widehat{k}_T(s_k,\tau)\,\widehat{h}_T(\theta,s_k).$$

Let $\widehat{h}_T(\theta)$ denote the vector $\left(\widehat{h}_T(\theta,s_k),\widehat{h}_T(\theta,s_k),...,\widehat{h}_T(\theta,s_k)\right)'$ and $\widehat{W}_T$ denote the matrix with elements: $W_{jk} = \omega_k \widehat{k}_T(s_k,s_j)$. Thus we can simply write:

$$K_T\widehat{h}_T(\theta) \approx \widehat{W}_T\widehat{h}_T(\theta).$$

For any given level of precision, the matrix $\widehat{W}_T$ can be looked at as the best finite dimensional reduction of the operator K_T. From the spectral decomposition of $K_{\alpha T}^{-1}$, it is easy to deduce the approximation:

$$K_{\alpha T}^{-1}\widehat{h}_T(\theta) \approx \left(\widehat{W}_T^2 + \alpha I\right)^{-1}\widehat{W}_T\widehat{h}_T(\theta) \equiv \widetilde{h}_T(\theta).$$

Finally, the objective function of the CGMM is computed as:

$$\left\langle K_{\alpha T}^{-1}\widehat{h}_T(\theta,.),\widehat{h}_T(\theta,.)\right\rangle = \int \left|K_{\alpha T}^{-1/2}\widehat{h}_T(\theta,\tau)\right|^2 \exp\left\{-\tau'\tau\right\} d\tau \approx \sum_{k=1}^{m} \omega_k \left|\widetilde{h}_T(\theta,s_k)\right|^2$$

where $\widetilde{h}_T(\theta,s_k)$ is the k^{th} component of $\widetilde{h}_T(\theta)$.