Econometric Theory, **33**, 2017, 479–526.
doi:10.1017/S0266466616000025

EFFICIENT ESTIMATION USING THE CHARACTERISTIC FUNCTION

MARINE CARRASCO
Université de Montréal

RACHIDI KOTCHONI
African School of Economics

The method of moments procedure proposed by Carrasco and Florens (2000) permits full exploitation of 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 Donald and Newey (2001), we derive a higher-order expansion of the estimator from which we characterize the finite sample dependence of the AMSE on α. We propose to select the regularization parameter by minimizing an estimate of the AMSE. We show that this procedure delivers a 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.

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. Therefore, an inference procedure based on the empirical CF has the potential to be as efficient as another one that exploits the likelihood function. Paulson, Holcomb, and Leitch (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 (1981a, 1981b, 1981c), Koutrouvelis (1980), Carrasco and Florens (2000),

We thank the editor Yuichi Kitamura and two referees for their insightful comments. 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. Address correspondence to Rachidi Kotchoni, African School of Economics. E-mail: rachidi.kotchoni@africanschoolofeconomics.com.

Jiang and Knight (2002), Chacko and Viceira (2003) and Carrasco et al. (2007) (henceforth, CCFG (2007)). An extensive review of empirical CF-based estimation methods can be found in Yu (2004) paper.

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, 2016). 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. A transition density of the same type arises in the autoregressive Gamma model (see Gourieroux and Jasiak, 2005). Ait-Sahalia and Kimmel (2007) propose closed form approximations for the log-likelihood function of certain 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 $\varphi(\tau, \theta_0)$ of an IID random vector $x_t \in \mathbb{R}^p$ $(t = 1, \ldots, T)$ is nothing but the expectation of $e^{i\tau' x_t}$ with respect to the distribution of x_t, where θ_0 is a finite dimensional 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 is $h_t(\tau, \theta) = e^{i\tau' x_t} - \varphi(\tau, \theta)$. 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 previous 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 $\tau \in (\tau_1, \tau_2, \ldots \tau_N)$. 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. This discretization approach is also advocated in Singleton (2001, Section 5.2.) and Chacko and Viceira (2003). However, the number of points in the grid for τ must not be larger than the sample size of 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 $(\tau_1, \tau_2, \ldots \tau_N)$ is too refined[1]. In fact, the discrete set of moment conditions $\{h_t(\tau_i, \theta)\}_{i=1}^N$ converges to the moment function $\tau \mapsto h_t(\tau, \theta), \tau \in \mathbb{R}$ so that operator methods are needed to handle the estimation procedure in the limit functional space.

Carrasco and Florens (2000) proposed a Continuum GMM (henceforth, CGMM) that efficiently uses the information content of a 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 is a quadratic form with metrics K^{-1} defined in an Hilbert space, 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 K by an estimator K_T obtained from a finite sample. However, the empirical operator K_T is degenerate and not invertible while its theoretical counterpart K is invertible only on a dense subset of the Hilbert space of interest. To circumvent these difficulties, Carrasco and Florens (2000) resorted to a Tikhonov-type regularized inverse of K_T, e.g. $K_{\alpha T} = (K_T^2 + \alpha I)^{-1} K_T$, 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. 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 terms, we find an expression of the form $\alpha_T = cT^{-g(\beta)}$, where $g(\beta)$ inherits some properties from the covariance operator K. To implement the optimal selection of α empirically, we minimize the approximate MSE (AMSE) of the CGMM estimator. This AMSE is not known analytically and must therefore be estimated by simulations. The proposed estimator of α_T, $\widehat{\alpha}_{TM}$, is indexed by the sample size T and the number of Monte Carlo replications M used to estimate the AMSE. Under certain regularity conditions, it is shown that $\widehat{\alpha}_{TM}$ is consistent for α_T as M goes to infinity.

This type of expansion has been used in Newey and Smith (2004) to study the higher order properties of generalized empirical likelihood estimators. Moreover, our method for selecting the regularization parameter is closely related to Donald and Newey (2001), who use an AMSE criterion to select the optimal number of instruments in a linear instrumental variable model.

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 Monte Carlo study based on the CIR term structure model and Section 6 concludes. The proofs are collected in the appendix.

2. OVERVIEW OF THE CGMM

This section presents an overview of the CGMM estimator and a summary of known results. The first subsection presents an example that motivates the use of the CGMM. The second subsection presents the general framework for implementing the CF-based CGMM procedure. The third subsection presents a consistency and asymptotic normality result.

2.1. Interest Rate Model as a Motivating Example

The recent literature on the term structure of interest rates provides an empirically relevant area of implementation for the CGMM estimator.

Indeed, the short-term nominal interest rate plays a key role in the valuation of financial assets. The dynamics of the short-term nominal rate is often specified in continuous-time as:

$$dx_t = \mu(x_t)\,dt + \sigma(x_t)\,dW_t, \tag{1}$$

where x_t denotes an interest rate process, $\mu(x_t)$ is a drift function, $\sigma(x_t)$ is an instantaneous volatility function and W_t is a standard Brownian motion. As noted by Singleton (2001), the transition density of a discrete sample from (1) is known only in the special case of a CIR model, i.e. when $\mu(x_t) = \kappa(\rho - x_t)$ and $\sigma(x_t) = \sigma\sqrt{x_t}$. In this case, the conditional density of x_t given $x_{t-\delta}$ ($\delta > 0$) can be represented as (see Devroye, 1986):

$$f(x_t|x_{t-\delta}) = \sum_{j=0}^{\infty} p_j \frac{x_t^{j+q_0-1} c_0^{j+q_0}}{\Gamma(j+q_0)} \exp(-c_0 x_t), \tag{2}$$

where $c_0 = \frac{2\kappa_0}{\sigma_0^2(1-e^{-\delta\kappa_0})}$, $q_0 = \frac{2\kappa_0\rho_0}{\sigma_0^2}$ and $p_j = \frac{(c_0 e^{-\delta\kappa_0} x_{t-\delta})^j \exp(-c_0 e^{-\delta\kappa_0} x_{t-\delta})}{j!}$.

As apparent from above, the CIR model is indexed by three parameters: κ, ρ and σ. The first parameter (κ) captures the strength of the mean reversion in the interest rate process. The second parameter (ρ) is the long run value around which the interest rate process oscillates according to business cycles. The third parameter (σ) captures the instantaneous volatility of the interest rate process. Imprecise estimations of these parameters may have dramatic implications for the valuation of interest rate sensitive assets, the prediction of the term structure of the market risk premium, and the assessment of interest rate risk. The importance of the short nominal interest rate makes it one of the most frequently modeled financial variables (Gray, 1996).

Contrasting with the infinite mixture of Gamma densities with Poisson weights given above, the conditional CF of the CIR model has a rather simple expression:

$$\varphi(\tau, \theta_0; x_{t-\delta}) \equiv E\left(e^{i\tau x_t}|x_{t-\delta}\right) = (1 - i\tau/c_0)^{-q} \exp\left(\frac{i\tau e^{-\delta\kappa_0} x_{t-\delta}}{1 - i\tau/c_0}\right), \tag{3}$$

where $\theta_0 = (\kappa_0, \rho_0, \sigma_0)'$. Furthermore, the CF is available in closed form for all models of type (1) as well as for more sophisticated affine-jump diffusions (Singleton, 2001; Jiang and Knight, 2002) and stochastic volatility models (Yu, 2004). For these reasons, several dynamic models are more naturally specified by their CFs than by their transition densities.

2.2. 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:

$$h_t(\tau, \theta; \theta_0) = e^{i\tau' x_t} - \varphi(\tau, \theta), \tag{4}$$

where $\varphi(\tau, \theta) = E^\theta\left(e^{i\tau' x_t}\right)$ is the CF of x_t, E^θ is the expectation operator with respect to the data generating process indexed by θ and θ_0 is the true parameter on which $h_t(\tau, \theta; \theta_0)$ depends implicitly via the actual data x_t.

CCFG (2007) extend the scope of the CGMM procedure to Markov and weakly dependent models. The moment function used in CCFG (2007) for the Markov case is:

$$h_t(\tau, \theta; \theta_0) = \left(e^{is' x_t} - \varphi(s, \theta; x_{t-1})\right) e^{ir' x_{t-1}}, \tag{5}$$

where $\varphi(s, \theta; x_{t-1}) = E^\theta(e^{is' x_t} | x_{t-1})$ is the CF of x_t conditional on x_{t-1}, $\tau = (s, r) \in \mathbb{R}^{2p}$ and $h_t(\tau, \theta; \theta_0)$ depends implicitly on θ_0 via the pair (x_t, x_{t-1}). In equation (5), the set of basis functions $\{e^{ir' x_{t-1}}\}$ is being used as instruments. CCFG (2007) show that these instruments are optimal, given the Markovian structure of the model. Note that moment conditions defined by (4) are IID processes whereas equation (5) describes a martingale difference sequence. In this paper, we restrict our attention to IID and Markov cases only.

It is important to stress that the moment function depends on θ_0 via the data because some of our proofs require the partial derivatives of $h_t(\tau, \theta; \theta_0)$ with respect to its third argument to exist. For that purpose, we will introduce a notation θ^0 representing this argument. Thus, $h_t(\tau, \theta; \theta^0)$ is the moment function evaluated at θ when the data are generated by the model with parameter θ^0. For the sake of simplicity, the dependence of $h_t(\tau, \theta; \theta^0)$ on θ^0 will be hidden when $\theta^0 = \theta_0$ (i.e., $h_t(\tau, \theta) \equiv h_t(\tau, \theta; \theta_0)$), where θ_0 is the particular value of θ^0 for the model that generated the actual data. Subsequently, the generic notation $h_t(\tau, \theta)$, $\tau \in \mathbb{R}^d$ denotes a moment function defined by either (4) or (5), where $d = p$ for (4) and $d = 2p$ for (5), and $\theta^0 = \theta_0$ for both cases.

Let π be a probability density 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.:

$$\mathbf{L}^2(\pi) = \left\{ f : \mathbb{R}^d \to \mathbf{C} \middle| \int f(\tau)\overline{f(\tau)}\pi(\tau)d\tau < \infty \right\}, \tag{6}$$

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 $\mathbf{L}^2(\pi) \times \mathbf{L}^2(\pi)$:

$$\langle f, g \rangle = \int f(\tau)\overline{g(\tau)}\pi(\tau)d\tau. \tag{7}$$

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

$$\widehat{\theta} = \arg\min_{\theta} \left\langle K^{-1}\widehat{h}_T(.,\theta), \widehat{h}_T(.,\theta) \right\rangle, \tag{8}$$

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

$$Kf(\tau_1) = \int_{-\infty}^{\infty} k(\tau_1, \tau) f(\tau) \pi(\tau) d\tau, \text{ for any } f \in L^2(\pi), \tag{9}$$

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

$$k(\tau_1, \tau_2) = E\left(h_t(\tau_1, \theta)\overline{h_t(\tau_2, \theta)}\right). \tag{10}$$

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:

$$k_T\left(\tau_1, \tau_2, \widehat{\theta}^1\right) = \frac{1}{T}\sum_{t=1}^{T} h_t\left(\tau_1, \widehat{\theta}^1\right)\overline{h_t\left(\tau_2, \widehat{\theta}^1\right)}. \tag{11}$$

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

$$k_T(\tau_1, \tau_2) = \frac{1}{T}\sum_{t=1}^{T}\left(e^{i\tau_1' x_t} - \widehat{\varphi}_T(\tau_1)\right)\overline{\left(e^{i\tau_2' x_t} - \widehat{\varphi}_T(\tau_2)\right)}, \tag{12}$$

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 (11) or (12) is degenerate and noninvertible. This problem is worsened by the fact that the inverse of K, which K_T estimates, exists only on a dense subset of $L^2(\pi)$. Moreover, when $g \equiv K^{-1}f$ 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 an appropriate rate. The expression for $K_{\alpha T}^{-1}$ uses a Tikhonov regularization, also called ridge regularization. Other forms of regularizations could have been used, see e.g. Carrasco, Florens, and Renault (2007).

The feasible CGMM estimator is given by:

$$\widehat{\theta}_T(\alpha; \theta_0) = \arg\min_{\theta} \widehat{Q}_T(\alpha, \theta; \theta_0), \tag{13}$$

where

$$\widehat{Q}_T(\alpha, \theta; \theta_0) = \left\langle K_{\alpha T}^{-1} \widehat{h}_T(., \theta), \widehat{h}_T(., \theta) \right\rangle, \tag{14}$$

and $\widehat{h}_T(\tau, \theta) \equiv \widehat{h}_T(\tau, \theta; \theta_0)$. An expression of $\widehat{Q}_T(\alpha, \theta; \theta_0)$ in matrix form is given in CCFG (2007, Section 3.3). An alternative expression and an algorithm for the numerical evaluation of this objective function based on Gauss–Hermite quadratures are described in Appendix D.

Note that (13) defined a function $\widehat{\theta}_T(\alpha; .)$ such that:

$$\widehat{\theta}_T\left(\alpha; \theta^0\right) = \arg\min_{\theta} \widehat{Q}_T\left(\alpha, \theta; \theta^0\right) \text{ for all } \theta^0.$$

Therefore, the second argument of $\widehat{\theta}_T(\alpha; .)$ is the third argument of $\widehat{h}_T(\tau, \theta; .)$ and $\widehat{Q}_T(\alpha, \theta; .)$.

2.3. Consistency and Asymptotic Normality

In order to study the properties of the CGMM estimator, the following assumptions are imposed.

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. *(i)* $h_t(\tau, \theta)$ is three times continuously differentiable with respect to θ. Furthermore, *(ii)* the first two derivatives of $h_t(\tau, \theta)$ with respect to its

second argument satisfy:

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial h_t(\tau,\theta)}{\partial\theta_j}-E\left(\frac{\partial h_t(\tau,\theta)}{\partial\theta_j}\right)=O_p\left(T^{-1/2}\right) \text{ and}$$

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 h_t(\tau,\theta)}{\partial\theta_j\partial\theta_k}-E\left(\frac{\partial^2 h_t(\tau,\theta)}{\partial\theta_j\partial\theta_k}\right)=O_p\left(T^{-1/2}\right),$$

for all j and k.

Assumption 4. *(i)* $E^{\theta_0}\left(h_T(\tau,\theta;\theta_0)\right)\in\Phi_\beta$ for all $\theta\in\Theta$ and for some $\beta\geq 1$, where

$$\Phi_\beta=\left\{f\in L^2(\pi) \text{ such that } \left\|K^{-\beta}f\right\|<\infty\right\}. \tag{15}$$

Furthermore, *(ii)* the first two derivatives of $E^{\theta_0}\left(h_T(\tau,\theta;\theta_0)\right)$ with respect to θ belong to Φ_β for all θ in a neighborhood of θ_0 and for the same β as above.

Assumption 5. *(i)* The random variable x_t is stationary Markov and satisfies $x_t=X\left(x_{t-1},\theta_0,\varepsilon_t\right)$ where $X\left(x_{t-1},\theta^0,\varepsilon_t\right)$ 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. Furthermore, *(ii)* the gradient $G(\tau,\theta;\theta^0)=E\left(\frac{\partial h_t(\tau,\theta;\theta^0)}{\partial\theta}\right)$ and covariance operator K are continuously differentiable with respect to in θ^0.

Assumptions 1 and 2 are quite standard and they have been used in Carrasco and Florens (2000). Note that the probability density function π that appears in these assumptions is the one used to build the scalar product $\langle .,.\rangle$ and it has nothing to do with the process that generated the actual data. The first part of Assumption 3 ensures some smoothness properties for $\widehat{\theta}_T(\alpha)\equiv\widehat{\theta}_T(\alpha;\theta_0)$ while the second part is always satisfied for IID models. The largest real number 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 largest eigenvalues. Because $Kf(.)$ involves a d-dimensional integration, β may be affected by both the dimensionality of the index τ and the smoothness of f.

Assumption 5-*(i)* 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 times a standardized representative of the family of distribution. Examples include the exponential family and the stable distribution. An IID model is a special case of Assumption 5 where $X(x_{t-1},\theta_0,\varepsilon_t)$ takes the simpler form $X(\theta_0,\varepsilon_t)$. 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 $X(x_{t-1},\theta_0,\varepsilon_t)$ may not be available in

analytical form. In particular, the relation $x_t = X(x_{t-1}, \theta_0, \varepsilon_t)$ can be the numerical solution of a general equilibrium asset pricing model (e.g., as in Duffie and Singleton, 1993).

By Assumptions 3 and 5-*(i)*, $\frac{\partial \widehat{h}_T(\tau,\theta;\theta^0)}{\partial \theta}$ is twice continuously differentiable with respect to θ^0 while the kernel $k_T(\tau_1, \tau_2, \widehat{\theta}^1)$ given by (11) is three times continuously differentiable with respect to θ^0. Therefore, the differentiability requirement of Assumption 5-*(ii)* already holds for the empirical gradient $G_T(\tau, \theta; \theta^0) = \frac{\partial \widehat{h}_T(\tau,\theta;\theta^0)}{\partial \theta}$ and the empirical covariance operator K_T. Assumption 5-*(ii)* is quite mild as it simply extends this differentiability to the probability limits $G(\tau, \theta; \theta^0)$ and K.

We have the following results for the two-step CGMM estimator.

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

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

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. Note that 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.

3. STOCHASTIC EXPANSION AND AMSE OF THE CGMM ESTIMATOR

Ideally, we would like 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, the MSE matrix is:

$$MSE(\alpha, \theta_0) = E\left[T\left(\widehat{\theta}_T(\alpha) - \theta_0\right)\left(\widehat{\theta}_T(\alpha) - \theta_0\right)'\right], \tag{16}$$

and the trace of $MSE(\alpha, \theta_0)$ is given by $E\left[T\left\|\widehat{\theta}_T(\alpha) - \theta_0\right\|^2\right]$.

Unfortunately, there is no theoretical basis for claiming that the variance of $\widehat{\theta}_T(\alpha)$ is finite for any data generating process and any sample size. Indeed, the

large sample properties of GMM-type estimators like $\widehat{\theta}_T(\alpha)$ are well known but their finite sample properties can be established only in special cases. In particular, the MSE of $\widehat{\theta}_T(\alpha)$ can be infinite in finite samples even though $\widehat{\theta}_T(\alpha)$ is consistent for θ_0. To hedge against situations where this variance is infinite, we consider approximating the MSE of $\widehat{\theta}_T(\alpha)$ by that of the leading terms of its stochastic expansion.

The higher order properties of GMM-type estimators have been studied by Rothenberg (1983, 1984), Koenker et al. (1994), Rilstone, Srivastava, and Ullah (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 that of Nagar (1959) and in particular to that of Donald and Newey (2001), who select the number of instruments to include in a linear instrumental variable model by minimization of an AMSE criterion.

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:

$$\begin{aligned}\left\| K_\alpha^{-1/2}\sqrt{T}\widehat{h}_T(\theta_0)\right\| &\leq \underbrace{\left\| \left(K^2+\alpha I\right)^{-1/4}\right\|}_{\leq \alpha^{-1/4}} \underbrace{\left\| \left(K^2+\alpha I\right)^{-1/4} K^{1/2}\right\|}_{\leq 1} \underbrace{\left\| \sqrt{T}\widehat{h}_T(\theta_0)\right\|}_{=O_p(1)} \\ &= O_p\left(\alpha^{-1/4}\right). \end{aligned} \tag{17}$$

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:

$$\widehat{\theta}_T(\alpha)-\theta_0 = \Delta_1+\Delta_2+\Delta_3+o_p\left(\alpha^{-1}T^{-1}\right)+o_p\left(\alpha^{\min\left(1,\frac{2\beta-1}{2}\right)}T^{-1/2}\right), \tag{18}$$

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 $\Delta_3 = O_p\left(\alpha^{-1}T^{-1}\right)$ are given in Appendix B. The next theorem uses this expansion to establish results on the AMSE matrix of $\widehat{\theta}_T(\alpha)$ and on the optimal convergence rate for the regularization parameter.

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

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

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

where

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

$$TVar = I_{\theta_0}^{-1} + O\left(\alpha^{\min\left(2,\frac{2\beta-1}{2}\right)}\right) + O\left(\alpha^{-1}T^{-1/2}\right),$$

as $T \to \infty$, $\alpha^2 T \to \infty$ *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:*

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

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 et al. (2007). The use of another regularization such as spectral cut-off or Landweber-Fridman would permit improvement in 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:

$$\alpha_T = cT^{-g(\beta)}, \tag{19}$$

for some positive function c that does not depend on T and a positive function $g(\beta)$ that satisfies $\max\left(\frac{1}{6},\frac{1}{2\beta+1}\right) \leq g(\beta) < 1/2$.

From Equations (B.9) and (B.11) in the Appendix, we have:

$$\widehat{\theta}_T(\alpha)-\theta_0 \simeq \Delta_T(\alpha,\theta_0)=\Delta_1+\Delta_2+\Delta_3.$$

Adding up the expressions of Δ_j, $j=1,2,3$ provided in Appendix B yields:

$$\Delta_T(\alpha,\theta_0)=-W_0^{-1}(\theta_0)\left\langle K_{\alpha T}^{-1}G(.,\theta_0),\widehat{h}_T(.,\theta_0)\right\rangle$$
$$+W_0^{-1}(\theta_0)\left[\left\langle K_{\alpha T}^{-1}G(.,\theta_0),G(.,\theta_0)\right\rangle-W_0(\theta_0)\right]W_0^{-1}\Psi_{T,0}\left(\theta^0\right),$$

where

$$G(\tau,\theta_0)=P\lim\frac{1}{T}\sum_{t=1}^{T}\frac{\partial h_t(\tau,\theta_0)}{\partial\theta},$$

$$\Psi_{T,0}(\theta_0)=\mathrm{Re}\left\langle K^{-1}G(.,\theta_0),\widehat{h}_T(.,\theta_0)\right\rangle, \text{ and}$$

$$W_0(\theta_0)=\left\langle K^{-1}G(.,\theta_0),G(.,\theta_0)\right\rangle.$$

By construction, the MSE of Δ_T given by:

$$\Sigma_T(\alpha,\theta_0)=E\left[T\left\|\Delta_T(\alpha,\theta_0)\right\|^2\right] \tag{20}$$

coincides with the AMSE of $\widehat{\theta}_T(\alpha)$ and is always finite. The limit of $\Sigma_T(\alpha,\theta_0)$ as $T\to\infty$ coincides with the asymptotic MSE of $\widehat{\theta}_T(\alpha)$.

4. ESTIMATION OF THE OPTIMAL REGULARIZATION PARAMETER

The minimizer of the AMSE,

$$\alpha_T(\theta_0)=\underset{\alpha\in[0,1]}{\arg\min}\,\Sigma_T(\alpha,\theta_0), \tag{21}$$

corresponds to α_T in Theorem 2. This minimizer is unfeasible because $\Delta_T(\alpha,\theta_0)$, of which $\Sigma_T(\alpha,\theta_0)$ is the MSE, depends on the unknown θ_0. We circumvent this difficulty by using a consistent first step estimator $\widehat{\theta}_T^1$ to simulate $\Sigma_T(\alpha,\theta_0)$.[2]

The expressions of $\Delta_T(\alpha,\theta_0)$ depend on both deterministic and random quantities. The deterministic quantities are the true parameter θ_0, the covariance operator K, the probability limit of the gradient of the moment function $G(\tau,\theta_0)$, and the regularization parameter α. The random quantities are the moment function $\widehat{h}_T(\tau,\theta_0)$ and the empirical covariance operator K_T. Therefore, we write:

$$\Delta_T(\alpha,\theta_0)=\Delta\left(\alpha,K,G(.,\theta_0),K_T(\theta_0),\widehat{h}_T(.,\theta_0)\right). \tag{22}$$

In the IID case, $G(\tau,\theta_0)$ and $K(\theta_0)$ are known in closed form since $G(\tau,\theta_0)=\frac{\partial\varphi(\tau,\theta_0)}{\partial\theta}$ and

$$k(\tau_1,\tau_2)=\varphi(\tau_1-\tau_2,\theta_0)-\varphi(\tau_1,\theta_0)\overline{\varphi(\tau_2,\theta_0)},$$

where $k(\tau_1, \tau_2)$ is the kernel of K.

Let $\widehat{\theta}_T^1$ be a consistent but inefficient estimator of θ. Our algorithm to estimate the AMSE of $\widehat{\theta}_T(\alpha)$ is as follows:

Step 1. Obtain an estimate $\widetilde{K}\left(\widehat{\theta}_T^1\right)$ of $K(\theta_0)$ and an estimate $\widetilde{G}\left(.,\widehat{\theta}_T^1\right)$ of $G(.,\theta_0)$ from a very large sample (e.g., 10000 or 50000 observations) simulated using $\widehat{\theta}_T^1$.[3]

Step 2. For $j = 1, 2, \ldots, M$:

Step 2.1. Draw independent samples $X_T^{(j)}\left(\widehat{\theta}_T^1\right)$ of size T from the data generating process using $\widehat{\theta}_T^1$.

Step 2.2. Use the sample $X_T^{(j)}\left(\widehat{\theta}_T^1\right)$ to compute the moment function $\widehat{h}_T^{(j)}\left(\tau, \widehat{\theta}_T^1\right)$ and the empirical covariance operator $K_T^{(j)}\left(\widehat{\theta}_T^1\right)$.

Step 2.3. Compute $\Delta_T^{(j)}\left(\alpha, \widehat{\theta}_T^1\right) = \Delta\Big(\alpha, \widetilde{K}\left(\widehat{\theta}_T^1\right), \widetilde{G}\left(.,\widehat{\theta}_T^1\right), K_T^{(j)}\left(\widehat{\theta}_T^1\right), \widehat{h}_T^{(j)}\left(.,\widehat{\theta}_T^1\right)\Big)$.

Step 3. Compute the simulated AMSE of $\widehat{\theta}_T(\alpha)$ as:

$$\widehat{\Sigma}_{TM}\left(\alpha, \widehat{\theta}_T^1\right) = \frac{T}{M}\sum_{j=1}^{M}\left\|\Delta_T^{(j)}\left(\alpha, \widehat{\theta}_T^1\right)\right\|^2, \tag{23}$$

where T and M denote the sample size and the number of Monte Carlo replications.[4]

With an estimator of the AMSE in hand, we estimate the optimal regularization parameter as:

$$\widehat{\alpha}_{TM}\left(\widehat{\theta}^1\right) = \underset{\alpha\in[0,1]}{\arg\min}\,\widehat{\Sigma}_{TM}(\alpha, \widehat{\theta}^1). \tag{24}$$

Note that under our notation, $\Sigma_T\left(\alpha, \widehat{\theta}^1\right)$ denotes the probability limit of $\widehat{\Sigma}_{TM}\left(\alpha, \widehat{\theta}^1\right)$ as M goes to infinity[5] and $\alpha_T\left(\widehat{\theta}^1\right)$ denotes the minimizer of $\Sigma_T\left(\alpha, \widehat{\theta}^1\right)$.

The approach to estimate the optimal regularization parameter described above is rather fast as it does not require a numerical optimization at each Monte Carlo replication. However, the overall procedure rests on the presumption that $\widetilde{K}$ is a highly accurate approximation of K. This presumption is reasonable given that the simulated sample used to estimate K at Step 1 can be made arbitrarily large. Our subsequent results are established by assuming that $\widetilde{K}$ and $\widetilde{G}$ entails no estimation error, which is equivalent to set $\widetilde{K} = K$ and $\widetilde{G} = G$. We further make the following assumption:

Assumption 6. The regularization parameter α that minimizes $\Sigma_T(\alpha, \theta_0)$ 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 $\max\left(\frac{1}{6}, \frac{1}{2\beta+1}\right) \leq g(\beta) < 1/2$.

Assumption 6 seems reasonable given the finding of Theorem 2 (ii). Such expression for the smoothing parameter is often used as a starting point for optimal bandwidth selection in nonparametric density estimation. Examples in the semiparametric context include Linton (2002) and Jacho-Chavez (2010).

THEOREM 3. *Let $\widehat{\theta}^1$ be a $\sqrt{T}-$consistent estimator of θ_0. Then under Assumptions 1 to 6, we have:*

(i) $\frac{\alpha_T(\widehat{\theta}^1)}{\alpha_T(\theta_0)} - 1$ *converges in probability to zero as T goes to infinity;*

(ii) $\frac{\widehat{\alpha}_{TM}(\theta_0)}{\alpha_T(\theta_0)} - 1$ *converges in probability to zero at rate $M^{-1/2}$ as M goes to infinity and T is fixed;*

(iii) $\widehat{\alpha}_{TM}(\widehat{\theta}^1) - \alpha_T(\theta_0) = O_p(T^{-1/2}) + O_p(M^{-1/2})$ *as M goes to infinity first and T goes to infinity.*

In Theorem 3-*(i)*, 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. Theorem 3-*(i)* guarantees that under our assumptions, the difference $\alpha_T(\widehat{\theta}^1) - \alpha_T(\theta_0)$ goes to zero faster than $\alpha_T(\theta_0)$ itself.

Theorem 3-*(ii)* characterizes the rate of convergence of $\frac{\widehat{\alpha}_{TM}(\theta_0)}{\alpha_T(\theta_0)}$. Indeed, $\widehat{\alpha}_{TM}(\theta_0)$ is the minimum of the AMSE simulated using the true θ_0. In the proof, one first shows that the conditions for the uniform convergence in probability of $\widehat{\Sigma}_{TM}(\alpha, \theta_0) - \Sigma_T(\alpha, \theta_0)$ 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 $\frac{\widehat{\alpha}_{TM}(\theta_0)}{\alpha_T(\theta_0)}$. Note that when T goes to infinity, $\alpha_T(\theta_0)$ goes to zero so that the rate given in Theorem 3-*(ii)* is no longer valid. However, we have that $\widehat{\alpha}_{TM}(\theta_0) - \alpha_T(\theta_0) = O_p\left(M^{-1/2}\right)$ when M goes to infinity first and T goes to infinity second.

Theorem 3-*(iii)* revisits the previous result when θ_0 is replaced by a consistent estimator. It rests on a sequential limit in M and T, which is needed here because $\alpha_T(\theta_0)$ goes to zero as T goes to infinity. Such a 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[6]. Theorem 3-*(iii)* implies that $\widehat{\alpha}_{TM}(\widehat{\theta}^1)$ benefits from an increase in both M and T.

The next theorem establishes that the minimum of the simulated AMSE is consistent for the optimal AMSE, $\Sigma_T(\alpha_T(\theta_0), \theta_0)$.

THEOREM 4. *Let $\widehat{\theta}^1$ be a $\sqrt{T}-$consistent estimator of θ_0. Then under assumptions 1 to 6,*

$$\frac{\widehat{\Sigma}_{TM}\left(\widehat{\alpha}_{TM}(\widehat{\theta}^1), \widehat{\theta}^1\right)}{\Sigma_T(\alpha_T(\theta_0), \theta_0)} - 1 = o_p(1)$$

as M goes to infinity first and $T \leq M$ goes to infinity second.

This result is obtained by exploiting the continuous differentiability of $\widehat{\Sigma}_{TM}(\alpha, \widehat{\theta}^1)$ with respect to α and the consistency of $\widehat{\alpha}_{TM}(\widehat{\theta}^1)$ for $\alpha_T(\theta_0)$. Let $\widehat{\theta}_T(\widehat{\alpha}_{TM}, \theta_0)$ be the CGMM estimator based on the actual sample and the regularization parameter $\widehat{\alpha}_{TM}(\widehat{\theta}^1)$ obtained by solving (24). The last theorem compares $\widehat{\theta}_T(\widehat{\alpha}_{TM}, \theta_0)$ to the ideal CGMM estimator based on $\alpha_T(\theta_0)$.

THEOREM 5. *Let* $\widehat{\alpha}_{TM} \equiv \widehat{\alpha}_{TM}(\widehat{\theta}^1)$ *and* $\widehat{\theta}_T(\alpha, \theta_0) \equiv \widehat{\theta}_T(\alpha)$. *Then under assumptions 1 to 6,*

$$\sqrt{T}\left(\widehat{\theta}_T(\widehat{\alpha}_{TM}) - \widehat{\theta}_T(\alpha_T(\theta_0))\right) = O_p\left(T^{1/2}M^{-1/2}\right),$$

as M *goes to infinity first and* $T \leq M$ *goes to infinity second.*

Theorem 5 implies that the distribution of $\sqrt{T}\left(\widehat{\theta}(\widehat{\alpha}_{TM}) - \theta_0\right)$ is the same as the distribution of $\sqrt{T}\left(\widehat{\theta}(\alpha_T) - \theta_0\right)$ provided $T \leq M$. This ensures that replacing α_T by its estimator $\widehat{\alpha}_{TM}$ does not affect the consistency, asymptotic normality, and efficiency of the final CGMM estimator $\widehat{\theta}(\widehat{\alpha}_{TM})$. The proof of this theorem relies mainly on the fact that $\widehat{\theta}(\alpha)$ is continuously differentiable with respect to α.

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

We pursue two goals in this simulation study. First, we investigate the properties of the feasible AMSE $\widehat{\Sigma}_{TM}\left(\alpha, \widehat{\theta}_T^1\right)$ as the regularization parameter (α), the sample size (T) and the number of replications (M) vary. Second, we compare the CGMM estimator based on the optimally selected regularization parameter with a competing GMM estimator. For this purpose, we consider estimating the parameters of a square-root diffusion. The first subsection below describes the simulation design whilst the second subsection presents the results.

5.1. Simulation Design

A continuous time process r_t is said to follow a square-root (or CIR) diffusion if it obeys the following stochastic differential equation:

$$dr_t = \kappa(\rho - r_t)\,dt + \sigma\sqrt{r_t}\,dW_t \tag{25}$$

where $\kappa > 0$ is the strength of mean reversion in the process, $\rho > 0$ is the long run mean and $\sigma > 0$ is the instantaneous volatility parameter. This model has been widely used in financial econometrics for the term structure of interest rates (Cox, Ingersoll, and Ross, 1985) and also for the volatility of stock prices (Heston, 1993). The stochastic differential equation (25) admits a unique and positive fundamental solution if $\sigma^2 \leq 2\kappa\rho$ (Feller, 1951).

For this simulation exercise, we assume that r_t is an interest rate process observed at regularly spaced time intervals indexed $t = 1, 2, \ldots, T$. The value retained for θ_0 is taken from Singleton (2001):

$$\theta_0 = (\kappa_0, \rho_0, \sigma_0) = (0.4, 6.0, 0.3). \tag{26}$$

Following Zhou (2001), one may consider estimating this CIR model by maximum likelihood by truncating the infinite sum representation of its transition density to 100 terms. The resulting sample log-likelihood function is:

$$\mathcal{L}_T(\theta) \simeq \sum_{t=2}^{T} \log\left(\sum_{j=0}^{100} p_j \frac{r_t^{j+q-1} c^{j+q}}{\Gamma(j+q)} \exp(-cr_t)\right),$$

where $\theta = (\kappa, \rho, \sigma)'$, $c = \frac{2\kappa}{\sigma^2(1-e^{-\kappa})}$, $q = \frac{2\kappa\rho}{\sigma^2}$ and $p_j = \frac{(ce^{-\kappa} r_{t-1})^j \exp(-ce^{-\kappa} r_{t-1})}{j!}$. Unfortunately, our experiments suggest that $\mathcal{L}_T(\theta)$ is often not a concave function of θ in small sample. Therefore, we shall abandon this route and consider CF-based estimators.

Following Singleton (2001), we consider a GMM estimator based on a discrete set of moment conditions given by:

$$h_t(\tau_k, \theta) = \left(e^{i\tau_{k,1} r_t} - \varphi(\tau_{k,2}, \theta; r_{t-1})\right) e^{i\tau_{k,2} r_{t-1}},\ \tau_k = (\tau_{k,1}, \tau_{k,2})' \in \mathbb{R}^2 \text{ and}$$

$$\varphi(\tau_{k,2}, \theta; r_{t-1}) = (1 - i\tau_{k,2}/c)^{-q} \exp\left(\frac{i\tau_{k,2} e^{-\kappa} r_{t-1}}{1 - i\tau_{k,2}/c}\right).$$

Let us define $\widehat{h}_t(\theta) = \left(\widehat{h}_t(\tau_1, \theta), \ldots, \widehat{h}_t(\tau_N, \theta)\right)'$ and $\widehat{h}_T(\theta) = \frac{1}{T}\sum_{i=2}^{T} h_t(\tau_k, \theta)$. The classical GMM estimator of Hansen (1982) is given by:

$$\widehat{\theta}_{GMM} = \arg\min_{\theta} \widehat{h}_T(\theta)' \widehat{S}^{-1} \widehat{h}_T(\theta), \tag{27}$$

where $\widehat{S} = \frac{1}{T}\sum_{i=1}^{T} \widehat{h}_t(\widehat{\theta}^1_{GMM}) \widehat{h}_t(\widehat{\theta}^1_{GMM})'$ is an estimate of the asymptotic covariance matrix of $\widehat{h}_t(\theta)$ and $\widehat{\theta}^1_{GMM}$ is the first step GMM estimator obtained by replacing $\widehat{S}$ by the identity matrix in (27). To compute $\widehat{\theta}_{GMM}$, we use 36 points in $\mathbb{R}^2$ obtained by taking the Cartesian product of six Gauss–Hermite quadrature points. Therefore, $\tau_k \in \Lambda^2$ where:

$$\Lambda = \{-2.3506, -1.3358, -0.4361, 0.4361, 1.3358, 2.3506\}. \tag{28}$$

We compare $\widehat{\theta}_{GMM}$ to the CGMM estimator given by:

$$\widehat{\theta}_{CGMM}(\alpha) = \arg\min_{\theta} \int_{\mathbb{R}^2} \left(K^{-1}_{\alpha T} \widehat{h}_T(\tau, \theta)\right) \overline{\widehat{h}_T(\tau, \theta)} e^{-\tau'\tau} d\tau, \tag{29}$$

where $K^{-1}_{\alpha T}$ is estimated according to (11) and using the first step CGMM estimator given by:

$$\widehat{\theta}^1_{CGMM} = \arg\min_{\theta} \int_{\mathbb{R}^2} \widehat{h}_T(\tau, \theta) \overline{\widehat{h}_T(\tau, \theta)} e^{-\tau'\tau} d\tau. \tag{30}$$

The objective function of the CGMM is evaluated using 36 Gauss–Hermite quadrature points in $\mathbb{R}^2$. The quadrature weight associated with $\tau_k = (\tau_{k,1}, \tau_{k,2})' \in \Lambda^2$ is $\omega(\tau_{k,1})\omega(\tau_{k,2})$, where $\omega(\tau_{k,j})$ is the weight associated with $\tau_{k,j} \in \Lambda$ in a one-dimensional integration (see Appendix D). We have:

$$\omega(\Lambda) = \{0.0045, 0.1571, 0.7246, 0.7246, 0.1571, 0.0045\},$$

where Λ is given by (28).

To begin, we use θ_0 to simulate once and for all $S = 500$ samples of size $T_{\max} = 1000$ with in mind that S is the number of Monte Carlo replications and $T_{\max}$ is the maximum sample size that will be considered subsequently.[7] The simulated samples are denoted $X_{T_{\max}}^{(s)}(\theta^0)$, $s = 1, \ldots, S$ and stored as the columns of a matrix of size $(T_{\max}, S)$. Next, we consider all pairs (T, M) such that:

$$T \in [250, 500, 750, 1000] \text{ and } M \in [250, 500, 750, 1000].$$

Let $M_{\max} = 1000$. Considering a given pair (T, M), we define $X_T^{(s)}(\theta^0)$ as the first T elements of $X_{T_{\max}}^{(s)}(\theta^0)$ and use it to compute a two-step GMM estimator $\left(\widehat{\theta}_{GMM}^{(s)}\right)$ and a first step CGMM estimator $\left(\widehat{\theta}_{CGMM}^{(1,s)}\right)$. In turn, $\widehat{\theta}_{CGMM}^{(1,s)}$ is used to simulate a large sample of size $TM_{\max}$, denoted $X_{TM_{\max}}^{(s)}$. This large sample is used to estimate K and $G(.,\theta_0)$ and the estimates are denoted $\widetilde{K}$ and $\widetilde{G}$.[8]

Next, we split $X_{TM_{\max}}^{(s)}$ into $M_{\max}$ nonoverlapping blocks of size T denoted $X_T^{(s,j)}$, $j = 1, \ldots, M_{\max}$. These samples are used to compute an optimal regularization parameter $\widehat{\alpha}_{TM}^{(s)} \equiv \widehat{\alpha}_{TM}\left(\widehat{\theta}_{CGMM}^{(1,s)}\right)$ by minimizing an AMSE function $\widehat{\Sigma}_{TM}\left(\alpha, \widehat{\theta}_{CGMM}^{(1,s)}\right)$ on the following grid:

$$\underline{\alpha} = [\alpha^{(1)}, \alpha^{(2)}, \ldots, \alpha^{(36)}] \text{ with}$$
$$\alpha^{(1)} = 10^{-10} \text{ and } \alpha^{(i)} = \alpha^{(i-1)}\exp(0.5), i = 2, \ldots, 36.$$

The AMSE function $\widehat{\Sigma}_{TM}\left(\alpha, \widehat{\theta}_{CGMM}^{(1,s)}\right)$ is computed using the first M samples $X_T^{(s,j)}$, $j = 1, \ldots, M$ and the first step estimator $\widehat{\theta}_{CGMM}^{(1,s)}$. We have:

$$\widehat{\Sigma}_{TM}\left(\alpha, \widehat{\theta}_{CGMM}^{(1,s)}\right) = \frac{T}{M}\sum_{j=1}^{M}\left\|\Delta_T\left(\alpha, \widetilde{K}, \widetilde{G}, K_T^{(s,j)}, \widehat{h}_T^{(s,j)}\right)\right\|^2,$$

where $K_T^{(s,j)}$ and $\widehat{h}_T^{(s,j)}$ are computed using the sample $X_T^{(s,j)}$ and the first step estimator $\widehat{\theta}_{CGMM}^{(1,s)}$.

Finally, a second step CGMM estimator $\widehat{\theta}_{CGMM}^{(s)}\left(\widehat{\alpha}_{TM}^{(s)}\right)$ is computed based on the optimal regularization parameter $\widehat{\alpha}_{TM}^{(s)}$. To reduce the computational burden, we have computed $\widehat{\theta}_{CGMM}^{(s)}\left(\widehat{\alpha}_{TM}^{(s)}\right)$ only when $M = M_{\max}$. At the end of the

simulation, we have S independent replications of the regularization parameter $\widehat{\alpha}_{TM}^{(s)}, s = 1, \ldots, S$, the second step GMM estimator $\widehat{\theta}_{GMM}^{(s)}, s = 1, \ldots, S$ and the second step CGMM estimator $\widehat{\theta}_{CGMM}^{(s)}\left(\widehat{\alpha}_{T,1000}^{(s)}\right), s = 1, \ldots, S$.

5.2. Simulations Results

Figure 1 shows examples of plots of the AMSE function $\widehat{\Sigma}_{TM}\left(\alpha, \widehat{\theta}_{CGMM}^{(1,s)}\right)$ against $\log\alpha$ for selected pairs (T, M). The convexity of the AMSE function improves as T increases from 250 to 1000, which suggests that the optimal regularization parameter is better identified when T is large. Increasing M seems to improve the convexity of the AMSE function as well.

Table 1 shows summary statistics for the simulated distribution of $\widehat{\alpha}_{TM}$ where "Mean", "Median", and "Std. Dev." are respectively the sample mean, median, and standard deviation of $\left(\widehat{\alpha}_{TM}^{(1)}, \ldots, \widehat{\alpha}_{TM}^{(S)}\right)$. The sample average of $\widehat{\alpha}_{TM}^{(s)}$ decreases monotonically in both M and T. By construction, the most reliable estimates of $\alpha_T(\theta_0)$ are obtained when M is the largest ($M = M_{\max}$). Therefore, our simulations suggest that the speed of convergence of $\widehat{\alpha}_{TM}$ as M increases is rather fast. This is seen by comparing the average variation of $\widehat{\alpha}_{TM}^{(s)}$ for a given T between $M = 250$ and $M = 500$ to the average variation between $M = 750$ and $M = 1000$. The convergence of $\widehat{\alpha}_{TM}$ as M increases seems to be faster when T is larger. This prescribes to always set M as large as possible, in particular when the sample size is small.

The sample median of $\widehat{\alpha}_{TM}^{(s)}$ is smaller than its sample mean everywhere. Therefore, considering that on average $\widehat{\alpha}_{TM}^{(s)}$ is decreasing in M and $\widehat{\alpha}_{T,250}^{(s)}$ is several times larger than $\widehat{\alpha}_{T,1000}^{(s)}$, we can claim that the median of $\widehat{\alpha}_{TM}^{(s)}$ is a more robust estimator of $\alpha_T(\theta_0)$ than is the sample mean. The sample standard deviation of $\widehat{\alpha}_{TM}^{(s)}$ decreases monotonically in M for every T. This standard deviation decreases monotonically in T only when M is large, which is consistent with the

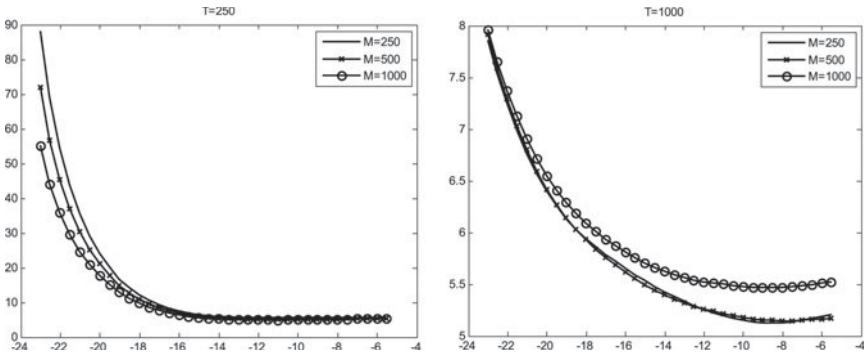

FIGURE 1. Plots of $\widehat{\Sigma}_{TM}\left(\alpha, \widehat{\theta}_{CGMM}^{(1,s)}\right)$ against $\log\alpha$ for selected values of T and M.

TABLE 1. Summary statistics for the simulated distribution of $\widehat{\alpha}_{TM}^{(s)}$ for different T and M

Number of replications: $S = 500$

		M = 250	M = 500	M = 750	M = 1000
T=250	Mean ($\times 10^{-3}$)	0.1658	0.0833	0.0734	0.0720
	Median ($\times 10^{-3}$)	0.0268	0.0442	0.0442	0.0442
	Std. Dev. ($\times 10^{-3}$)	0.5964	0.2417	0.2021	0.1949
T=500	Mean ($\times 10^{-3}$)	0.1534	0.0517	0.0466	0.0359
	Median ($\times 10^{-3}$)	0.0268	0.0268	0.0268	0.0268
	Std. Dev. ($\times 10^{-3}$)	0.6257	0.2529	0.1922	0.0514
T=750	Mean ($\times 10^{-3}$)	0.1061	0.0450	0.0268	0.0260
	Median ($\times 10^{-3}$)	0.0163	0.0163	0.0163	0.0163
	Std. Dev. ($\times 10^{-3}$)	0.5320	0.2525	0.0284	0.0257
T=1000	Mean ($\times 10^{-3}$)	0.0704	0.0485	0.0193	0.0191
	Median ($\times 10^{-3}$)	0.0099	0.0099	0.0163	0.0163
	Std. Dev. ($\times 10^{-3}$)	0.4045	0.3162	0.0224	0.0198

fact the estimator $\widehat{\alpha}_{TM}$ is more precise when M is larger. The simulated distribution of $\widehat{\alpha}_{TM}^{(s)}$ (not reported here) is rather skewed to the right, with a standard

TABLE 2. Comparison of $\widehat{\theta}_{GMM}^{(s)}$ and $\widehat{\theta}_{CGMM}^{(s)}$ $\left(\widehat{\alpha}_{T,1000}^{(s)}\right)$

Number of replications: $S = 500$

		GMM			CGMM		
		$\widehat{\kappa}_{GMM}$	$\widehat{\rho}_{GMM}$	$\widehat{\sigma}_{GMM}$	$\widehat{\kappa}_{CGMM}$	$\widehat{\rho}_{CGMM}$	$\widehat{\sigma}_{CGMM}$
T=250	Bias	0.0017	−0.0060	−0.0352	0.0192	−0.0108	−0.0016
	Std. Dev.	0.0948	0.1609	0.0192	0.0786	0.1187	0.0170
	RMSE	0.0948	0.1610	0.0401	0.0809	0.1192	0.0171
T=500	Bias	0.0099	−0.0194	−0.0296	0.0097	−0.0085	−0.0007
	Std. Dev.	0.0622	0.1093	0.0133	0.0524	0.0845	0.0127
	RMSE	0.0630	0.1110	0.0325	0.0533	0.0850	0.0127
T=750	Bias	0.0159	−0.0297	−0.0245	0.0071	−0.0096	−0.0011
	Std. Dev.	0.0525	0.0827	0.0115	0.0417	0.0673	0.0101
	RMSE	0.0549	0.0879	0.0271	0.0423	0.0680	0.0101
T=1000	Bias	0.0176	−0.0281	−0.0205	0.0060	−0.0083	−0.0008
	Std. Dev.	0.0456	0.0700	0.0106	0.0361	0.0591	0.0086
	RMSE	0.0489	0.0754	0.0231	0.0366	0.0597	0.0087

deviation that is larger than the mean even for large values of T and M. This suggests that our estimator of the regularization parameter may be unstable. However, as the standard deviation of $\widehat{\alpha}_{TM}^{(s)}$ decreases monotonically in M, potential instabilities in $\widehat{\alpha}_{TM}^{(s)}$ can be reduced by increasing M.

Table 2 shows the empirical bias, standard deviation and MSE of the individual coordinates of $\widehat{\theta}_{GMM}^{(s)}$ and $\widehat{\theta}_{CGMM}^{(s)}\left(\widehat{\alpha}_{T,1000}^{(s)}\right)$. These summary statistics are based on the simulated empirical distributions of the estimators and not on analytical formulas. In small samples ($T = 250$ and $T = 500$), the parameters κ and ρ are estimated with less bias by GMM than by CGMM. The opposite is true in large samples. Moreover, $\widehat{\sigma}_{CGMM}$ is less biased than $\widehat{\sigma}_{GMM}$ for all sample sizes. The standard deviation of the CGMM estimator is also always smaller than that of the GMM estimator. As a result, $\widehat{\theta}_{CGMM}^{(s)}\left(\widehat{\alpha}_{T,1000}^{(s)}\right)$ dominates $\widehat{\theta}_{GMM}^{(s)}$ in terms of the MSE for all cases. The largest efficiency gain is achieved by $\widehat{\sigma}_{CGMM}$ whose MSE is always several times smaller than that the MSE of $\widehat{\sigma}_{GMM}$.

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 $\widehat{\theta}_T(\alpha)$ that sheds light on how its finite sample approximate MSE (or AMSE) depends on the regularization parameter. The AMSE is simply the MSE of the leading terms of $\widehat{\theta}_T(\alpha) - \theta_0$ in the stochastic expansion. 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 $\alpha_T = cT^{-g(\beta)}$, where c does not depend on 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 the simulated AMSE. By relying on the leading terms of the expansion of $\widehat{\theta}_T(\alpha) - \theta_0$, our approach to select α remains valid even when the MSE of $\widehat{\theta}_T(\alpha)$ is infinite. The proposed estimator, $\widehat{\alpha}_{TM}$, is shown to be consistent for its theoretical counterpart α_T. The optimal selection of the regularization parameter enables the use of a fully feasible CGMM estimator that is a real alternative to the maximum likelihood estimator.

NOTES

1. These numerical problems are acknowledged by Singleton (2001, p. 131).
2. If analytical expressions were available for all the terms involved in $E\left(\Delta_T'\Delta_T\right)$, one could attempt to estimate such expressions in one shot based on $\widehat{\theta}_T^1$ and the sample. However, $Cov(\Delta_1, \Delta_3)$ cannot be computed explicitly even though Δ_1 and Δ_3 are available in closed form.
3. The simulated sample must be large enough to ensure that the approximation errors of $\widetilde{K}(\theta^0)$ and $\widetilde{G}(.,\theta^0)$ are negligible. Note that Step 1 is not necessary in the IID case.
4. The multiplication by T on the RHS of (23) ensures that $\widehat{\Sigma}_{TM}(\alpha,\theta^0) = O_p(1)$.
5. Indeed, $\widehat{\Sigma}_{TM}(\alpha,\theta)$ converges uniformly in probability to $\Sigma_T(\alpha,\theta)$.

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

7. See DeVroye (1986) and Zhou (2001) for details on how to simulate a CIR process.

8. During our simulations, the matrix approximation of $\widetilde{K}$ based on 36 quadrature points was always invertible. Therefore, there is no need to regularize the inverse of $\widetilde{K}$. This is due to the fact that 36 is small relative to $TM_{\max}$.

REFERENCES

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

Buse, A. (1992) The bias of instrumental variable estimators. *Econometrica* 60(1), 173–180.

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

Carrasco, M. & J.P. Florens (2000) Generalization of GMM to a continuum of moment conditions. *Econometric Theory* 16, 797–834.

Carrasco, M., J.P. Florens, & E. Renault (2007) Linear inverse problems in structural econometrics: Estimation based on spectral decomposition and regularization. In J.J. Heckman & E.E. Leamer (eds.), *Handbook of Econometrics*, vol. 6. pp. 5633–5751.

Chacko, G. & L. Viceira (2003) Spectral GMM estimation of continuous-time processes. *Journal of Econometrics* 116, 259–292.

Cox, J.C., J.E. Ingersoll, & S.A. Ross (1985) A theory of the term structure of interest rates. *Econometrica* 53, 385–407.

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

Donald, S. & W. Newey (2001) Choosing the number of instruments. *Econometrica* 69, 1161–1191.

Duffie, D. & K. Singleton (1993) Simulated moments estimation of Markov models of asset prices. *Econometrica* 61, 929–952.

Feller, W. (1951) Two singular diffusion problems. *Annals of Mathematics* 54, 173–182.

Feuerverger, A. & P. McDunnough (1981a) On efficient inference in symmetry stable laws and processes. In M. Csorgo (ed.), *Statistics and Related Topics*, pp. 109–122. North Holland.

Feuerverger, A. & P. McDunnough (1981b) On some Fourier methods for inference. *Journal of the American Statistical Association* 76, 379–387.

Feuerverger, A. & P. McDunnough (1981c) On the efficiency of empirical characteristic function procedures. *Journal of the Royal Statistical Society. Series B (Methodological)* 43, 20–27.

Feuerverger, A. & R. Mureika (1977) The empirical characteristic function and its applications. *The Annals of Statistics* 5(1), 88–97.

Gouriéroux, C. & J. Jasiak (2005) Autoregressive Gamma Processes. *Journal of Forecasting* 25, 129–152.

Gouriéroux, C. & A. Monfort (1996) *Simulation Based Econometric Methods*. CORE Lectures. Oxford University Press.

Gouriéroux, C., A. Monfort, & E. Renault (1993) Indirect inference. *Journal of Applied Econometrics* 8, S85–S118.

Gray, S.F. (1996) Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics* 42, 27–62.

Hansen, L. (1982) Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029–1054.

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.

Jacho-Chavez, D.T. (2010) Optimal bandwidth choice for estimation of inverse conditional–density–weighted expectations. *Econometric Theory* 26, 94–118.

Jiang, G. & J. Knight (2002) Estimation of continuous time processes via the empirical characteristic function. *Journal of Business and Economic Statistics* 20, 198–212.

Koenker, R., J.A.F. Machado, C.L. Skeels, & A.H.I. Welsh (1994) Momentary lapses: Moment expansions and the robustness of minimum distance estimators. *Econometric Theory* 10, 172–197.

Koutrouvelis, I.A. (1980) Regression-type estimation of the parameters of stable laws. *Journal of the American Statistical Association* 75(372), 918–928.

Linton, O. (2002) Edgeworth approximation for semiparametric instrumental variable and test statistics. *Journal of Econometrics* 106, 325–368.

Liu, Q. & D.A. Pierce (1994) A note on Gauss–Hermite quadrature. *Biometrika* 81(3), 624–629.

Madan, D.B. & E. Senata (1990) The variance gamma model for share market returns. *Journal of Business* 63(4), 511–524.

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.

Newey, W.K. & D. McFadden (1994) Large sample estimation and hypotheses testing. In R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2111–2245

Newey, W.K. & R.J. Smith (2004) Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* 72(1), 219–255.

Nolan, J. P. (2016) *Stable Distributions: Models for Heavy Tailed Data*. Springer Verlag.

Paulson, A.S., W.E. Holcomb, & R.A. Leitch (1975) The estimation of the parameters of the stable laws. *Biometrika* 62(1), 163–170.

Phillips, P.C.B. & H.R. Moon (1999) Linear regression limit theory for nonstationary panel data. *Econometrica* 67(5), 1057–1112.

Rilstone, P., V.K. Srivastava, & A. Ullah (1996) The second-order bias and mean-squared error of nonlinear estimators. *Journal of Econometrics* 75, 369–395.

Rothenberg, T.J. (1983) Asymptotic properties of some estimators in structural models. In S. Karlin, T. Amemiya, & L.A. Goodman (eds.), *Studies in Econometrics, Time Series and Multivariate Statistics*. Academic Press.

Rothenberg, T.J. (1984) Approximating the distributions of econometric estimators and test statistics. In Z. Griliches & M.D. Intriligator (eds.), *Handbook of Econometrics*, vol. 2. North-Holland.

Singleton, K.J. (2001) Estimation of affine pricing models using the empirical characteristic function. *Journal of Econometrics* 102, 111–141.

Yu, J. (2004) Empirical characteristic function estimation and its applications. *Econometric Reviews* 23(2), 93–123.

Zhou, H. (2001) Finite sample properties of EMM, GMM, QMLE, and MLE for a square-root interest rate diffusion model. *Journal of Computational Finance* 2, 89–122.

APPENDIX

Notation: Throughout the appendix, $h_t(\tau,\theta;\theta^0)$ denotes a moment function that depends explicitly on θ and implicitly on θ^0 via the data. A quadratic form $\widehat{Q}_T(\alpha,\theta;\theta^0)$ of $h_t(s,\theta;\theta^0)$ is minimized with respect to θ and the solution is the CGMM estimator of θ^0, denoted $\widehat{\theta}(\alpha,\theta^0)$. The actual data available to the econometrician are thought of as being generated by DGP with a true value θ_0. Thus, θ_0 is a particular value of θ^0. For simplicity, we write $h_t(\tau,\theta)=h_t(s,\theta;\theta_0)$. The asymptotic covariance operator of $h_t(\tau,\theta)$ is denoted K and its sample counterpart K_T. The gradient of $h_t(\tau,\theta)$ is denoted $G_T(\tau,\theta)$, its j^{th} coordinate $G_{j,T}(\tau,\theta)$, and its probability limit $G(\tau,\theta)$. The Hessian of $h_t(\tau,\theta)$ is denoted $H_T(\tau,\theta)$, its j^{th} column $H_{j,T}(\tau,\theta)$, its (k,j) element $H_{k,j,T}(\tau,\theta)$ and its probability limit $H(\tau,\theta)$. The derivative of $H_T(\tau,\theta)$ with respect to θ_j is denoted $L_{j,T}$ and its probability limit L_j. Finally, we recall that $\langle f,g\rangle$ is the scalar product of two functions f and g, $\|f\|=\sqrt{\langle f,f\rangle}$ is the norm of f, $\overline{z}$ is the complex conjugate of z and $|z|=\sqrt{z\overline{z}}$ is the modulus of z.

A. SOME BASIC PROPERTIES OF THE COVARIANCE OPERATOR

For a more formal exposition of the results mentioned in this appendix, see Carrasco, Florens, and Renault (2007). Let K be the covariance operator defined in (9) and (10), $\widehat{h}_t(\tau,\theta)$ the moment function defined in (4) and (5), and Φ_β the subset of $L^2(\pi)$ defined in Assumption 4.

DEFINITION A.1. *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 A.2. *$R(K)$ is a subspace of $L^2(\pi)$.*

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

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

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

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

implying that

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

DEFINITION A.3. *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 A.4. *ϕ is an eigenfunction of K associated with eigenvalue μ if and only if $K\phi = \mu\phi$.*

PROPOSITION A.5. *Suppose $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_j \geq \cdots$ 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 A.6. *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 A.7. *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 $\|K^{-\beta}f\| < \infty$. In fact, $f \in R(K^{\beta_2}) \Rightarrow K^{-\beta_2}f$ exists and $\|K^{-\beta_2}f\|^2 = \sum_{j=1}^{\infty}\mu_j^{-2\beta_2}|\langle f,\phi_j\rangle|^2 < \infty$. Thus if

$f \in R(K^{\beta_2})$, we have:

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

$\Rightarrow K^{-\beta_1} f$ exists $\Rightarrow f \in R(K^{\beta_1})$. 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 notation $\left\langle K^{-1/2} f, K^{-1/2} f\right\rangle$ is well defined while $\left\langle K^{-1} f, f\right\rangle$ is not. By abuse of notation however, we define $\left\langle K^{-1} f, f\right\rangle \equiv \left\langle K^{-1/2} f, K^{-1/2} f\right\rangle$.

B. EXPANSION OF THE MSE AND PROOFS OF THEOREMS 1 AND 2

B.1. Preliminary results and proof of Theorem 1

LEMMA B.1. *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:*

$$\left\|K_{\alpha T}^{-1} - K_\alpha^{-1}\right\| = O_p\left(\alpha^{-3/2} T^{-1/2}\right), \tag{B.1}$$

$$\left\|\left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right) f\right\| = O_p\left(\alpha^{-1} T^{-1/2}\right), \tag{B.2}$$

$$\left\|\left(K_\alpha^{-1} - K^{-1}\right) f\right\| = O\left(\alpha^{\min(1, \frac{\beta-1}{2})}\right), \tag{B.3}$$

$$\left\langle\left(K^{-1} - K_\alpha^{-1}\right) f, f\right\rangle = O\left(\alpha^{\min\left(1, \frac{2\beta-1}{2}\right)}\right). \tag{B.4}$$

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

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

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

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

This proves (B.1).

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

$$\begin{aligned} \left(K_{\alpha T}^{-1} - K_\alpha^{-1} \right) K &= (K_T^2 + \alpha I)^{-1} K_T K - (K^2 + \alpha I)^{-1} K^2 \\ &= (K_T^2 + \alpha I)^{-1} (K_T - K) K \qquad \textbf{(B.5)} \\ &\quad + \left[(K_T^2 + \alpha I)^{-1} - (K^2 + \alpha I)^{-1} \right] K^2. \qquad \textbf{(B.6)} \end{aligned}$$

The term (B.5) can be bounded in the following manner

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

For the term (B.6), 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

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

This proves (B.2).

Now we turn our attention toward equation (B.3). 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:

$$\begin{aligned} \text{(B.3)} &= \left\| (K^2 + \alpha I)^{-1} K f - K^{-1} f \right\| = \left(\sum_{j=1}^{\infty} \left(\frac{\mu_j^2}{\alpha + \mu_j^2} - 1 \right)^2 \frac{\left| \langle f, \phi_j \rangle \right|^2}{\mu_j^2} \right)^{1/2} \\ &= \left(\sum_{j=1}^{\infty} \mu_j^{2\beta - 2} \left(\frac{\mu_j^2}{\alpha + \mu_j^2} - 1 \right)^2 \frac{\left| \langle f, \phi_j \rangle \right|^2}{\mu_j^{2\beta}} \right)^{1/2} \\ &\leq \left(\sum_{j=1}^{\infty} \frac{\left| \langle f, \phi_j \rangle \right|^2}{\mu_j^{2\beta}} \right)^{1/2} \sup_{1 \leq j \leq \infty} \mu_j^{\beta - 1} \frac{\alpha}{\alpha + \mu_j^2}. \end{aligned}$$

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

$$\sup_{0\leq\mu\leq\mu_1}\mu^{\beta-1}\frac{\alpha}{\alpha+\mu_j^2}=\sup_{0\leq\lambda\leq\mu_1^2}\lambda^{\frac{\beta-1}{2}}\left(\frac{\alpha/\lambda}{\alpha/\lambda+1}\right). \tag{B.7}$$

We need to examine two cases:

- If $1\leq\beta\leq 3$: 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 (B.7) 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 $+\infty$.

- If $\beta>3$: We rewrite the left hand side of (B.7) 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 $f\in\Phi_\beta$: (B.3)$=O\left(\alpha^{\min(1,\frac{\beta-1}{2})}\right)$.

Finally, we consider (B.4). We have:

$$\begin{aligned}(\text{B.4})&=\sum_j\left(\frac{1}{\mu_j}-\frac{\mu_j}{\mu_j^2+\alpha}\right)\langle f,\phi_j\rangle^2=\sum_j\left(1-\frac{\mu_j^2}{\mu_j^2+\alpha}\right)\frac{\langle f,\phi_j\rangle^2}{\mu_j}\\&=\sum_j\mu_j^{2\beta-1}\left(1-\frac{\mu_j^2}{\mu_j^2+\alpha}\right)\frac{\langle f,\phi_j\rangle^2}{\mu_j^{2\beta}}\leq\sum_j\frac{\langle f,\phi_j\rangle^2}{\mu_j^{2\beta}}\sup_{\mu\leq\mu_1}\mu^{2\beta-1}\frac{\alpha}{\mu^2+\alpha}.\end{aligned}$$

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: (B.4)$=O\left(\alpha^{\min(1,\frac{2\beta-1}{2})}\right)$. ■

LEMMA B.2. *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 B.2.

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

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 B.1 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 B.3. *For all nonrandom functions* (u, v), *we have:*

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

Proof of Lemma B.3. We have:

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

Because the h_ts are uncorrelated, we have:

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

and

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

■

LEMMA B.4. *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:

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

Proof of Lemma B.4. Note that S contains θ_0 and

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

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

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

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

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

Hence, we have:

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

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

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

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 those of Propositions 3.2 and 4.1 in CCFG (2007). However, we now exploit the fact $E\left(\frac{\partial\hat{h}_T(.,\theta)}{\partial\theta}\right)=E\left(\frac{\partial h_t(.,\theta)}{\partial\theta}\right)\in\Phi_\beta$ with $\beta\geq 1$. The consistency follows from Lemma B.2 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:

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

Hence the asymptotic normality requires the same conditions as the consistency, that is, $\alpha T^{1/2} \to \infty$ and $\alpha \to 0$. The proof of the asymptotic efficiency is identical to that of CCFG. ■

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

The objective function is

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

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

$$
\frac{\partial \widehat{Q}_T\left(\alpha,\widehat{\theta}\right)}{\partial \theta} = 2\,\mathrm{Re}\left\langle K_{\alpha T}^{-1} G\left(.,\widehat{\theta}\right), \widehat{h}_T\left(.,\widehat{\theta}\right) \right\rangle = 0, \tag{B.8}
$$

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

A third order expansion gives

$$
0 = \frac{\partial \widehat{Q}_T(\alpha,\theta_0)}{\partial \theta} + \frac{\partial^2 \widehat{Q}_T(\alpha,\theta_0)}{\partial \theta \partial \theta'}\left(\widehat{\theta}-\theta_0\right) + \sum_{j=1}^{q}\left(\widehat{\theta}_j - \theta_{j,0}\right)\frac{\partial^3 \widehat{Q}_T\left(\alpha,\overline{\theta}\right)}{\partial \theta_j \partial \theta \partial \theta'}\left(\widehat{\theta}-\theta_0\right),
$$

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 $\Psi_T(\theta_0)$ as follows:

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

where

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

where the rates of convergence are obtained using the Cauchy-Schwarz inequality and the results of Lemma B.1. Similarly, we decompose $W_T(\theta_0)$ into 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

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

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

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

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

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

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:

$$\begin{aligned}\Delta_1 &= -W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0),\\ \Delta_2 &= -W_0^{-1}(\theta_0)\left[\Psi_{T,\alpha}(\theta_0)-W_\alpha(\theta_0)W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0)\right],\\ \Delta_3 &= -W_0^{-1}(\theta_0)\left[\widetilde{\Psi}_{T,\alpha}(\theta_0)-\widetilde{W}_\alpha(\theta_0)W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0)\right],\end{aligned}$$

$$\begin{aligned}\Delta_4 = &W_0^{-1}(\theta_0)W_{T,0}(\theta_0)W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0)\\ &-\sum_{j=1}^{q}\left(W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0)\right)_j W_0^{-1}(\theta_0)B_j(\bar{\theta})W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0),\end{aligned}$$

$$\Delta_5=-\sum_{j=1}^{q}\left(W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0)\right)_j W_0^{-1}(\theta_0)(B_{j,T}(\bar{\theta})-B_j(\bar{\theta}))W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0).$$

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

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

To summarize, we have:

$$\widehat{\theta}-\theta_0=\Delta_1+\Delta_2+\Delta_3+o_p\left(\alpha^{-1}T^{-1}\right)+o_p\left(\alpha^{\min(1,\frac{2\beta-1}{2})}T^{-1/2}\right). \qquad \textbf{(B.9)}$$

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

The objective function here is given by:

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

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

$$\frac{\partial \widehat{Q}_T(\alpha,\widehat{\theta})}{\partial \theta} = 2\,\mathrm{Re}\left\langle K_{\alpha T}^{-1} \widehat{G}_T(.,\widehat{\theta}), \widehat{h}_T(.,\widehat{\theta}) \right\rangle = 0 \qquad \textbf{(B.10)}$$

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 (B.10) around θ_0 yields:

$$0 = \frac{\partial \widehat{Q}_T(\alpha,\theta_0)}{\partial \theta} + \frac{\partial^2 \widehat{Q}_T(\alpha,\theta_0)}{\partial \theta \partial \theta'}(\widehat{\theta}-\theta_0) + \sum_{j=1}^{q} (\widehat{\theta}_j - \theta_{j,0}) \frac{\partial^3 \widehat{Q}_T(\alpha,\overline{\theta})}{\partial \theta_j \partial \theta \partial \theta'}(\widehat{\theta}-\theta_0),$$

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

Let us define:

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

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

and

$$\widehat{\Psi}_T(\theta_0) = \mathrm{Re}\left\langle K_{\alpha T}^{-1} \widehat{G}_T(.,\theta_0), \widehat{h}_T(.,\theta_0) \right\rangle,$$

$$\widehat{W}_T(\theta_0) = \left\langle K_{\alpha T}^{-1} \widehat{G}_T(.,\theta_0), \widehat{G}_T(.,_0) \right\rangle + \mathrm{Re}\left\langle K_{\alpha T}^{-1} \widehat{H}_T(.,\theta_0), \widehat{h}_T(.,\theta_0) \right\rangle,$$

$$\widehat{B}_{j,T}(\bar{\theta}) = 2\,\mathrm{Re}\left\langle K_{\alpha T}^{-1} \widehat{G}_T(.,\bar{\theta}), \widehat{H}_{j,T}(.,\bar{\theta}) \right\rangle + \mathrm{Re}\left\langle K_{\alpha T}^{-1} \widehat{H}_T(.,\bar{\theta}), \widehat{G}_{j,T}(.,\bar{\theta}) \right\rangle + \mathrm{Re}\left\langle K_{\alpha T}^{-1} \widehat{L}_{j,T}(.,\bar{\theta}), \widehat{h}_T(.,\bar{\theta}) \right\rangle.$$

Then, the expansion of the FOC becomes:

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

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

$$G(\tau,\theta) = \operatorname*{plim}_{T\to\infty} \widehat{G}_T(\tau,\theta),\ H(\tau,\theta) = \operatorname*{plim}_{T\to\infty} \widehat{H}_T(\tau,\theta),$$

$$G_j(\tau,\theta) = \operatorname*{plim}_{T\to\infty} \widehat{G}_{j,T}(\tau,\theta),\ H_j(\tau,\theta) = \operatorname*{plim}_{T\to\infty} \widehat{H}_{j,T}(\tau,\theta).$$

It follows from Assumption 3 that:

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

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 Lemma B.1, we obtain the following rates:

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

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 $\widehat{W}_T(\theta_0)$:

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

where

$$W_0(\theta_0)=\left\langle K^{-1}G,G\right\rangle=O(1),$$
$$W_\alpha(\theta_0)=\left\langle\left(K_\alpha^{-1}-K^{-1}\right)G,G\right\rangle=O\left(\alpha^{\min(1,\frac{2\beta-1}{2})}\right),$$
$$\widetilde{W}_\alpha(\theta_0)=\left\langle\left(K_{\alpha T}^{-1}-K_\alpha^{-1}\right)G,G\right\rangle=O_p\left(\alpha^{-1}T^{-1/2}\right),$$
$$\widehat{W}_\alpha(\theta_0)=\left\langle K_\alpha^{-1}\left(\widehat{G}_T-G\right),G\right\rangle=O_p\left(\alpha^{-1/2}T^{-1/2}\right),$$
$$W_1(\theta_0)=\mathrm{Re}\left\langle K^{-1}H,\widehat{h}_T\right\rangle+\left\langle K^{-1}G,\widehat{G}_T-G\right\rangle=O_p\left(T^{-1/2}\right),$$
$$W_{1,\alpha}(\theta_0)=\left\langle\left(K_{\alpha T}^{-1}-K_\alpha^{-1}\right)\left(\widehat{G}_T-G\right),G\right\rangle=O_p\left(\alpha^{-3/2}T^{-1}\right).$$

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

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

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

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

$$B_j(\bar{\theta}) = 2\,\mathrm{Re}\left\langle K^{-1}G(.,\bar{\theta}), H_j(.,\bar{\theta})\right\rangle + \mathrm{Re}\left\langle K^{-1}H(.,\bar{\theta}), G_j(.,\bar{\theta})\right\rangle = O_p(1) \text{ and}$$

$$B_{j,T}(\bar{\theta}) = B_j(\bar{\theta}) + O_p\left(\alpha^{\min(1,\frac{\beta-1}{2})}\right) + O_p\left(\alpha^{-1}T^{-1/2}\right).$$

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

$$\Delta_1 = -W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0) = O_p\left(T^{-1/2}\right),$$

$$\Delta_2 = -W_0^{-1}(\theta_0)\left[\Psi_{T,\alpha}(\theta_0) - W_\alpha(\theta_0)W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0)\right] = O_p\left(\alpha^{\min(1,\frac{2\beta-1}{2})}T^{-1/2}\right),$$

$$\Delta_3 = -W_0^{-1}(\theta_0)\left[\widetilde{\Psi}_{T,\alpha}(\theta_0) - \widetilde{W}_\alpha(\theta_0)W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0)\right] = O_p\left(\alpha^{-1}T^{-1}\right),$$

$$\widehat{R}_1 = W_0^{-1}(\theta_0)\widehat{W}_\alpha(\theta_0)W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0) = O_p\left(\alpha^{-1/2}T^{-1}\right),$$

$$\widehat{R}_2 = W_0^{-1}(\theta_0)W_1(\theta_0)W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0)$$
$$-\sum_{j=1}^{q}\left(W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0)\right)_j W_0^{-1}(\theta_0)B_j(\bar{\theta})W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0) = O_p\left(T^{-1}\right),$$

$$\widehat{R}_3 = W_0^{-1}(\theta_0)W_{1,\alpha}(\theta_0)W_0^{-1}(\theta_0)\Psi_{T,0}(\theta_0) = O_p\left(\alpha^{-3/2}T^{-3/2}\right),$$

and

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

In summary, we have:

$$\widehat{\theta}-\theta_0 = \Delta_1+\Delta_2+\Delta_3+o_p\left(\alpha^{-1}T^{-1}\right)+o_p\left(\alpha^{\min(1,\frac{2\beta-1}{2})}T^{-1/2}\right), \tag{B.11}$$

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

Proof of Theorem 2. Using the expansions given in (B.9) and (B.11), we obtain:

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

Lemma B.4 ensures that all terms that are slower than $O_p\left(T^{-1}\right)$ in the expansion of $\widehat{\theta}-\theta_0$ are real. Hence, the Re symbol are removed from the expression of Δ_1, Δ_2, and Δ_3 subsequently.

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 $h_t(\tau,\theta)$ is a martingale difference sequence and is bounded, we obtain:

$$\begin{aligned}E\left(\left\|\widehat{h}_T\right\|^2\right) &= E\left(\int \widehat{h}_T(\tau,\theta)\,\overline{\widehat{h}}_T(\tau,\theta)\,\pi(\tau)\,d\tau\right) \\ &= \frac{1}{T}E\left(\int h_t(\tau,\theta)\,\overline{h}_t(\tau,\theta)\,\pi(\tau)\,d\tau\right) = O\left(T^{-1}\right).\end{aligned} \tag{B.12}$$

Next, using (B.5) and (B.6), we obtain:

$$\begin{aligned}E\left(\left\|\left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right)G\right\|^2\right) &\leq E\left(\left\|\left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right)K\right\|^2\right)\left\|K^{-1}G\right\|^2 \\ &\leq E\left(\left\|(K_T^2 + \alpha I)^{-1}(K_T - K)K\right\|^2\right)\left\|K^{-1}G\right\|^2 \qquad \text{(B.13)} \\ &\quad + E\left(\left\|(K_T^2 + \alpha I)^{-1} - (K^2 + \alpha I)^{-1}\right\|^2\right)\|K\|^4 \\ &\quad \times \left\|K^{-1}G\right\|^2. \qquad \text{(B.14)}\end{aligned}$$

Hence:

$$\begin{aligned}\text{(B.13)} &= E\left(\left\|(K_T^2 + \alpha I)^{-1}(K_T - K)K\right\|^2\right)\left\|K^{-1}G\right\|^2 \\ &\leq \alpha^{-2}E\left(\|K_T - K\|^2\right)\|K\|^2\left\|K^{-1}G\right\|^2 = O\left(\alpha^{-2}T^{-1}\right),\end{aligned}$$

where $E\left(\|K_T - K\|^2\right) = O\left(T^{-1}\right)$ follows from Carrasco and Florens (2000, Theorem 4, p. 825). For (B.14), we use the fact that $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$ to obtain:

$$\begin{aligned}\text{(B.14)} &= E\left(\left\|(K_T^2 + \alpha I)^{-1} - (K^2 + \alpha I)^{-1}\right\|^2\right)\|K\|^4\left\|K^{-1}G\right\|^2 \\ &\leq E\left(\left\|(K_T^2 + \alpha I)^{-1}\right\|\left\|K_T^2 - K^2\right\|^2\left\|(K^2 + \alpha I)^{-1}\right\|\right)\|K\|^4\left\|K^{-1}G\right\|^2 \\ &\leq \alpha^{-2}E\left(\left\|K_T^2 - K^2\right\|^2\right)\|K\|^4\left\|K^{-1}G\right\|^2 \\ &\leq \alpha^{-2}E\left(\|K_T - K\|^2\|K_T + K\|^2\right)\|K\|^4\left\|K^{-1}G\right\|^2.\end{aligned}$$

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

$$(B.14) \leq \alpha^{-2} E\left[\|K_T - K\|^2 (\|K_T\| + \|K\|)^2\right] \|K\|^4 \left\|K^{-1}G\right\|^2.$$

From (A.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:

$$\|K\| \leq \sqrt{\int\int |k(\tau_1, \tau_2)|^2 \pi(\tau_1) \pi(\tau_2) d\tau_1 d\tau_2} \leq 2$$

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

Consequently:

$$(B.14) \leq 16\alpha^{-2} E\left(\|K_T - K\|^2\right) \|K\|^4 \left\|K^{-1}G\right\|^2 = O\left(\alpha^{-2}T^{-1}\right).$$

Finally,

$$\left\|E\left(\widetilde{\Psi}_{T,\alpha}\right)\right\| = \sqrt{O\left(\alpha^{-2}T^{-1}\right) \times O\left(T^{-1}\right)} = O\left(\alpha^{-1}T^{-1}\right). \quad \textbf{(B.15)}$$

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

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

We have:

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

where $E\left(\|\widehat{h}_T(.,\theta_0)\|^2\right) = O\left(T^{-1}\right)$ follows from (B.12). Next:

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

where the rate follows from (B.13) and (B.14). Hence, by the Cauchy-Schwarz inequality,

$$\begin{aligned} \left\|E\left(\widetilde{W}_\alpha W_0^{-1}\Psi_{T,0}\right)\right\| &\leq \sqrt{E\left(\|\widetilde{W}_\alpha\|^2\right) E\left(\left\|W_0^{-1}\Psi_{T,0}\right\|^2\right)} \\ &= \sqrt{O\left(\alpha^{-2}T^{-1}\right) O\left(T^{-1}\right)} = O\left(\alpha^{-1}T^{-1}\right). \quad \textbf{(B.16)} \end{aligned}$$

By putting (B.15) and (B.16) together, we find $E\left[\Delta_3\right] = O\left(\alpha^{-1}T^{-1}\right)$ so that the squared bias satisfies:

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

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 B.3. 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 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 B.3, 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

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

Also,

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

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

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

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

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

The second squared term leads to:

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

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

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

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

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 (B.13) and (B.14) leads to

$$E\left(\left\|\left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right) G\right\|^4\right) \leq E\left(\left\|\left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right) K\right\|^4\right) \left\|K^{-1}G\right\|^4$$

$$\leq E\left(\left\|\left(K_T^2 + \alpha I\right)^{-1}(K_T - K) K\right\|^4\right) \left\|K^{-1}G\right\|^4 \quad \textbf{(B.17)}$$

$$+ E\left(\left\|\left(K_T^2 + \alpha I\right)^{-1} - \left(K^2 + \alpha I\right)^{-1}\right\|^4\right) \|K\|^8$$

$$\times \left\|K^{-1}G\right\|^4. \quad \textbf{(B.18)}$$

Hence,

$$\text{(B.17)} = E\left(\left\|\left(K_T^2 + \alpha I\right)^{-1}(K_T - K) K\right\|^4\right) \left\|K^{-1}G\right\|^4$$

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

For (B.18), we use $A^{-1} - B^{-1} = A^{-1}(B - A) B^{-1}$ to obtain

$$\text{(B.18)} = E\left(\left\|\left(K_T^2 + \alpha I\right)^{-1} - \left(K^2 + \alpha I\right)^{-1}\right\|^4\right) \|K\|^8 \left\|K^{-1}G\right\|^4$$

$$\leq E\left(\left\|\left(K_T^2 + \alpha I\right)^{-1}\right\|^2 \left\|K_T^2 - K^2\right\|^4 \left\|\left(K^2 + \alpha I\right)^{-1}\right\|^2\right) \|K\|^8 \left\|K^{-1}G\right\|^4$$

$$\leq \alpha^{-4} E\left(\left\|K_T^2 - K^2\right\|^4\right) \|K\|^8 \left\|K^{-1}G\right\|^4$$

$$\leq \alpha^{-4} E\left(\|K_T - K\|^4 \|K_T + K\|^4\right) \|K\|^8 \left\|K^{-1}G\right\|^4.$$

By the triangular inequality,

$$\text{(B.18)} \leq \alpha^{-4} E\left(\|K_T - K\|^4 (\|K_T\| + \|K\|)^4\right) \|K\|^8 \left\|K^{-1}G\right\|^4$$

$$\leq 256\alpha^{-4} E\left(\|K_T - K\|^4\right) \|K\|^8 \left\|K^{-1}G\right\|^4,$$

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

The rates of (B.17) and (B.18) depend on the rate of $E\left(\|K_T - K\|^4\right)$.

$$\|K_T - K\|^2 \leq \int\int \left|\frac{1}{T}\sum_{t=1}^{T} \chi_t(\tau_1, \tau_2)\right|^2 \pi(\tau_1)\pi(\tau_2) d\tau_1 d\tau_2$$

$$= \frac{1}{T^2}\sum_{t=1}^{T} \int\int |\chi_t(\tau_1, \tau_2)|^2 \pi(\tau_1)\pi(\tau_2) d\tau_1 d\tau_2 \quad \textbf{(B.19)}$$

$$+ \frac{1}{T^2}\sum_{t \neq l}^{T} \int\int \chi_t(\tau_1, \tau_2)\overline{\chi_l(\tau_1, \tau_2)}\pi(\tau_1)\pi(\tau_2) d\tau_1 d\tau_2, \quad \textbf{(B.20)}$$

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

$$E\left(\|K_T - K\|^4\right) \leq E\left([(\text{B.19})]^2\right) + 2E\left([(\text{B.19})][(\text{B.20})]\right) + E\left([(\text{B.20})]^2\right).$$

Because $E\left([(\text{B.19})][(\text{B.20})]\right) \leq \sqrt{E\left([(\text{B.19})]^2\right)\left(E[(\text{B.20})]^2\right)}$, we only need to check the rates of the squared terms. We have

$$\begin{aligned} E\left([(\text{B.19})]^2\right) &= \frac{T}{T^4} E\left[\left(\int\int |\chi_t(\tau_1,\tau_2)|^2 \pi(\tau_1)\pi(\tau_2)d\tau_1 d\tau_2\right)^2\right] \\ &\quad + \frac{T(T-1)}{T^4} E\left[\left(\int\int |\chi_t(\tau_1,\tau_2)|^2 \pi(\tau_1)\pi(\tau_2)d\tau_1 d\tau_2\right)\right. \\ &\quad \left.\times \left(\int\int |\chi_l(\tau_1,\tau_2)|^2 \pi(\tau_1)\pi(\tau_2)d\tau_1 d\tau_2\right)\right], \text{ for } l \neq t. \end{aligned}$$

Hence $E\left([(\text{B.19})]^2\right) = O(T^{-2})$. Next:

$$\begin{aligned} &E\left([(\text{B.20})]^2\right) \\ &= \frac{1}{T^4}\sum_{t\neq l}^{T} E\left[\left(\int\int \chi_t(\tau_1,\tau_2)\overline{\chi_l(\tau_1,\tau_2)}\pi(\tau_1)\pi(\tau_2)d\tau_1 d\tau_2\right)^2\right] \\ &\quad + \frac{1}{T^4}\sum_{t\neq l, n\neq j, (t,l)\neq(n,j)}^{T} E\left[\left(\int\int \chi_t\overline{\chi_l}\pi(\tau_1)\pi(\tau_2)d\tau_1 d\tau_2\right)\right. \\ &\quad \left.\times\left(\int\int \chi_n\overline{\chi_j}\pi(\tau_1)\pi(\tau_2)d\tau_1 d\tau_2\right)\right] \end{aligned}$$

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

$$\begin{aligned} E\left([(\text{B.20})]^2\right) &= \frac{T(T-1)}{T^4} E\left[\left(\int\int \chi_t(\tau_1,\tau_2)\overline{\chi_l(\tau_1,\tau_2)}\pi(\tau_1)\pi(\tau_2)d\tau_1 d\tau_2\right)^2\right] \\ &= O\left(T^{-2}\right). \end{aligned}$$

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

$$E\left(\left\|\left(K_{\alpha T}^{-1} - K_\alpha^{-1}\right)G\right\|^4\right) \leq (\text{B.17})+(\text{B.18}) = O\left(\alpha^{-4}T^{-2}\right) \text{ and}$$

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

In total:

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

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,

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

For the second term, we have:

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

according to (B.13)-(B.14). Hence,

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

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

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

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

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

For the first term of $Var(\Delta_2)$, we use Lemma B.3 to obtain

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

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 $f(x)=\frac{x}{1+x}x^{-\frac{2\beta-1}{4}}$ is continuous and hence bounded for x away from 0 and infinity. When x goes to infinity, $f(x)$ goes to zero because $2\beta-1>0$. When x goes to zero, $f(x)=\frac{x^{\frac{5-2\beta}{4}}}{1+x}$ goes to zero because $5-2\beta>0$. Hence, $f(x)$ 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 rate, not a big O rate.

For the second term of $Var(\Delta_2)$, we use the fact that $W_\alpha=O\left(\alpha^{\min(1,\frac{2\beta-1}{2})}\right)$ according to Equation (B.4) in Lemma B.1:

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

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 C.1. *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.*

Proof of Lemma C.1. The objective function $\widehat{Q}_T\left(\alpha,\theta;\theta^0\right)$ is:

$$\widehat{Q}_T\left(\alpha,\theta;\theta^0\right) = \sum_{j=1}^{T} \frac{\widehat{\mu}_j}{\alpha+\widehat{\mu}_j^2}\left(\left\langle \widehat{h}_T(.,\theta;\theta^0), \widehat{\phi}_j\right\rangle\right)^2,$$

where $\widehat{\phi}_j$ is the eigenfunction of K_T associated with the eigenvalue $\widehat{\mu}_j$. By assumption 3, the moment function $\widehat{h}_T(.,\theta;\theta^0)$ 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, $x_t = X\left(x_{t-1},\theta^0,\varepsilon_t\right)$ is three times continuously differentiable with respect to θ_0. As an exponential function of x_t, $h_t(\tau,\theta;\theta_0)$ is three times continuously differentiable with respect to θ^0. The spectrum of the empirical covariance operator K_T with kernel given by (11) is also three times continuously differentiable with respect to θ^0. Therefore, $\widehat{Q}_T\left(\alpha,\theta;\theta^0\right)$ is three times continuously differentiable with respect to θ and θ^0. We now turn our attention to the differentiability with respect to α. It is easy to check that:

$$\frac{\partial \widehat{Q}_T\left(\alpha,\theta;\theta^0\right)}{\partial\alpha} = \left\langle \widetilde{K}_{\alpha T}\widehat{h}_T(.,\theta;\theta^0), \widehat{h}_T(.,\theta;\theta^0)\right\rangle,$$

where $\widetilde{K}_{\alpha T} \equiv -\left(K_T^2+\alpha I\right)^{-2}K_T$ which is well defined on $L^2(\pi)$ for α fixed. When α goes to zero however, we have to be more careful. We check that $\left|\left\langle \widetilde{K}_{\alpha T}\widehat{h}_T(.,\theta;\theta^0), \widehat{h}_T(.,\theta;\theta^0)\right\rangle\right|$ is bounded:

$$\begin{aligned}&\left|\left\langle \widetilde{K}_{\alpha T}\widehat{h}_T(.,\theta;\theta^0), \widehat{h}_T(.,\theta;\theta^0)\right\rangle\right| \\ &\quad\le \left\|\widetilde{K}_{\alpha T}\widehat{h}_T(.,\theta;\theta^0)\right\|\left\|\widehat{h}_T(.,\theta;\theta^0)\right\| \le \left\|\left(K_T^2+\alpha_T I\right)^{-2}K_T\right\|\left\|\widehat{h}_T(.,\theta;\theta^0)\right\|^2\end{aligned}$$

$$= \underbrace{\left\| \left(K_T^2 + \alpha I\right)^{-3/2} \right\|}_{\leq \alpha^{-3/2}} \underbrace{\left\| \left(K_T^2 + \alpha I\right)^{-1/2} K_T \right\|}_{\leq 1} \underbrace{\left\| \widehat{h}_T(.,\theta;\theta^0) \right\|^2}_{= O_p(T^{-1})}$$
$$= O_p\left(\alpha^{-3/2} T^{-1}\right) = o_p(1),$$

where the last equality follows from Theorem 2(ii). This shows that $\widehat{Q}_T(\alpha,\theta;\theta^0)$ is once continuously differentiable with respect to α. By the implicit function theorem, $\widehat{\theta}_T(\alpha;\theta^0) = \arg\min_\theta \widehat{Q}_T(\alpha,\theta;\theta^0)$ is once continuously differentiable with respect to α and twice continuously differentiable with respect to θ^0. ■

LEMMA C.2. *Under Assumptions 1 to 5, $\widehat{\Sigma}_{TM}(\alpha;\theta^0)$ and $\Sigma_T(\alpha;\theta^0)$ are once continuously differentiable with respect to α on $[\underline{\alpha},1]$ whenever $\underline{\alpha} > 0$, and twice continuously differentiable with respect to θ^0. Furthermore, $\alpha_T(\theta^0) = \arg\min_{\alpha\in[0,1]} \Sigma_T(\alpha;\theta_0)$ is continuous in θ^0.*

Proof of Lemma C.2. Recall that $\widehat{\Sigma}_{TM}(\alpha,\theta^0) = \frac{T}{M}\sum_{j=1}^M \left\| \Delta_T^{(j)}(\alpha,\theta^0) \right\|^2$, where the generic expression of $\Delta_T^{(j)}(\alpha,\theta^0)$ is of the form $\Delta_T(\alpha,\theta^0)$:

$$\begin{aligned}\Delta_T\left(\alpha,\theta^0\right) = &-W_0^{-1}(\theta^0)\left\langle K_{\alpha T}^{-1} G(.,\theta^0;\theta^0), \widehat{h}_T(.,\theta^0;\theta^0)\right\rangle \\ &+W_0^{-1}(\theta^0)\left[\left\langle K_{\alpha T}^{-1} G(.,\theta^0;\theta^0), G(.,\theta^0;\theta^0)\right\rangle - W_0(\theta^0)\right] \\ &\times W_0^{-1}(\theta^0)\Psi_{T,0}(\theta^0),\end{aligned}$$

with $\Psi_{T,0}(\theta^0) = \mathrm{Re}\left\langle K^{-1}G(.,\theta^0;\theta^0), \widehat{h}_T(.,\theta^0;\theta^0)\right\rangle$ and $W_0(\theta^0) = \left\langle K^{-1}G(.,\theta^0;\theta^0), G(.,\theta^0;\theta^0)\right\rangle$. As a smooth function of $\widehat{h}_T(.,\theta^0;\theta^0)$, $G(.,\theta^0;\theta^0)$, K_T and K, $\Delta_T\left(\alpha,\theta^0\right)$ is twice continuously differentiable with respect to θ^0 by Assumptions 3 and 5. Concerning the differentiability with respect to α, we note that $\frac{\partial K_{\alpha T}^{-1} f}{\partial \alpha} = \widetilde{K}_{\alpha T} f$, where $\widetilde{K}_{\alpha T} \equiv -\left(K_T^2 + \alpha I\right)^{-2} K_T$. It was shown in the proof of Lemma C.1 that $\left|\left\langle \widetilde{K}_{\alpha T}\widehat{h}_T(.,\theta^0), \widehat{h}_T(.,\theta^0)\right\rangle\right|$ is bounded. We now examine the term $\left|\left\langle \widetilde{K}_{\alpha T} G(.,\theta^0), G(.,\theta^0)\right\rangle\right|$:

$$\begin{aligned}\left|\left\langle \widetilde{K}_{\alpha T} G\left(.,\theta^0\right), G\left(.,\theta^0\right)\right\rangle\right| &\leq \left\| \widetilde{K}_{\alpha T} K^2 K^{-1} G\left(.,\theta^0\right)\right\| \left\| K^{-1} G\left(.,\theta^0\right)\right\| \\ &\leq \left\| \left(K_T^2 + \alpha I\right)^{-2} K_T K^2 \right\| \left\| K^{-1} G\left(.,\theta^0\right)\right\|^2 \\ &\leq \underbrace{\left\| \left(K_T^2 + \alpha I\right)^{-3/2} K^2 \right\|}_{= O_p(\alpha^{-1})} \underbrace{\left\| \left(K_T^2 + \alpha I\right)^{-1/2} K_T \right\|}_{\leq 1} \\ &\quad \times \left\| K^{-1} G(.,\theta^0)\right\|^2.\end{aligned}$$

Thus, $\left|\left\langle \widetilde{K}_{\alpha T} G(.,\theta^0), G(.,\theta^0)\right\rangle\right| = O_p\left(\underline{\alpha}^{-1}\right)$ if $\alpha \in [\underline{\alpha}, 1]$. This shows that $\Delta_T\left(\alpha,\theta^0\right)$ is once continuously differentiable with respect to α on $[\underline{\alpha}, 1]$ as long as $\underline{\alpha}$ is bounded away from zero. $\widehat{\Sigma}_{TM}\left(\alpha,\theta^0\right)$ and $\Sigma_T\left(\alpha;\theta^0\right)$ inherit the differentiability properties of $\Delta_T\left(\alpha,\theta^0\right)$ of which they are averages quadratic functions. Note that for a fixed T, $\Sigma_T(\alpha,\theta_0)$ increases without bound as $\alpha \to 0$. Therefore, $\alpha\left(\theta_0\right) = \underset{\alpha\in[0,1]}{\arg\min}\,\Sigma_T\left(\alpha;\theta_0\right)$ is bounded away from zero. Consequently, we define a sequence $\underline{\alpha}_T$ such that $\alpha_T\left(\theta_0\right) \geq \underline{\alpha}_T$ for all T. Finally, the Maximum theorem implies that $\alpha_T\left(\theta_0\right) = \underset{\alpha\in[\underline{\alpha},1]}{\arg\min}\,\Sigma_T\left(\alpha;\theta_0\right)$ is continuous in θ_0 for $\underline{\alpha} > 0$. ■

Proof of Theorem 3. (i) Using Assumption 6, we have $\frac{\alpha_T(\widehat{\theta}^1)}{\alpha_T(\theta_0)} = \frac{c(\widehat{\theta}^1)}{c(\theta_0)}$. Moreover by Assumption 6, $c\left(\theta\right)$ is a continuous function 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 $T \to \infty$. Hence the first result.

(ii) By construction, $\widehat{\Sigma}_{TM}(\alpha,\theta_0) - \Sigma_T(\alpha,\theta_0) = O_p\left(M^{-1/2}\right)$ for all α and T. However, a uniform convergence statement of type

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

is problematic given that $\widehat{\Sigma}_{TM}(\alpha,\theta_0)$ and $\Sigma_T(\alpha,\theta_0)$ are not bounded away from zero as $\alpha \to 0$. As noted in the proof of Lemma C.2 however, the minimizer of $\Sigma_T(\alpha,\theta_0)$, $\alpha_T\left(\theta_0\right)$, is bounded away from zero for any fixed T. Consequently, there exists a sequence $\underline{\alpha}_T$ such that $\alpha_T\left(\theta_0\right) \geq \underline{\alpha}_T$ for all T, implying that $\alpha\left(\theta_0\right) = \underset{\alpha\in[\underline{\alpha}_T,1]}{\arg\min}\,\Sigma_T\left(\alpha;\theta_0\right)$. The boundedness of $\Sigma_T(\alpha;\theta_0)$ on the choice set $\alpha \in \left[\underline{\alpha}_T, 1\right]$ and its continuity imply that $\sup_{\alpha\in[\underline{\alpha}_T,1]} \left\| \frac{\partial \widehat{\Sigma}_{TM}(\alpha,\theta_0)}{\partial\alpha} \right\|$ is finite, which is the requirement of Lemma 2.4 of Newey and McFadden (1994) for uniform convergence in probability. It thus follows from Theorem 2.1 of Newey and McFadden (1994) that $\widehat{\alpha}_{TM}\left(\theta_0\right) - \alpha_T\left(\theta_0\right) = O_p\left(M^{-1/2}\right)$ as $M \to \infty$. As T remains fixed, $\alpha_T\left(\theta_0\right)$ is bounded away from zero and hence, $\frac{\widehat{\alpha}_{TM}(\theta_0)}{\alpha_T(\theta_0)} - 1 = O_p\left(M^{-1/2}\right)$ as $M \to \infty$. Hence the second result.

(iii) We consider the following decomposition:

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

By first letting M go to infinity, we use the result of Theorem 3-*(ii)*: $\widehat{\alpha}_{TM}\left(\theta_0\right) - \alpha_T\left(\theta_0\right) = O_p\left(M^{-1/2}\right)$. As $\widehat{\theta}^1$ does not depend on M, we can apply Theorem 3-*(ii)* to $\widehat{\alpha}_{TM}(\widehat{\theta}^1) - \alpha_T(\widehat{\theta}^1)$ for fixed T. Next, we let T go to infinity in order to use the result of Theorem 3-*(i)*: $\alpha_T(\widehat{\theta}^1) - \alpha_T(\theta_0) = O_p\left(T^{-1/2}\right)$. The result follows. ■

Proof of Theorem 4. By the mean value theorem, we have:

$$\widehat{\Sigma}_{TM}\left(\widehat{\alpha}_{TM}\left(\widehat{\theta}^1\right),\widehat{\theta}^1\right) = \widehat{\Sigma}_{TM}\left(\alpha_T\left(\widehat{\theta}^1\right),\widehat{\theta}^1\right) + \frac{\partial \widehat{\Sigma}_{TM}\left(\overline{\alpha}_T,\widehat{\theta}^1\right)}{\partial\alpha} \times \left(\widehat{\alpha}_{TM}\left(\widehat{\theta}^1\right) - \alpha_T\left(\widehat{\theta}^1\right)\right).$$

Thus, we have:

$$\frac{\widehat{\Sigma}_{TM}\left(\widehat{\alpha}_{TM}\left(\widehat{\theta}^1\right),\widehat{\theta}^1\right)}{\Sigma_T\left(\alpha_T\left(\theta^0\right),\theta^0\right)}-1=\left(\frac{\widehat{\Sigma}_{TM}\left(\alpha_T\left(\widehat{\theta}^1\right),\widehat{\theta}^1\right)}{\Sigma_T\left(\alpha_T\left(\theta^0\right),\theta^0\right)}-1\right)$$
$$+\frac{1}{\Sigma_T\left(\alpha_T\left(\theta^0\right),\theta^0\right)}\frac{\partial\widehat{\Sigma}_{TM}\left(\overline{\alpha}_T,\widehat{\theta}^1\right)}{\partial\alpha}$$
$$\times\left(\widehat{\alpha}_{TM}\left(\widehat{\theta}^1\right)-\alpha_T\left(\widehat{\theta}^1\right)\right).$$

Considering the first term, we note that $\frac{\widehat{\Sigma}_{TM}(\alpha_T(\widehat{\theta}^1),\widehat{\theta}^1)}{\Sigma_T(\alpha_T(\theta^0),\theta^0)}$ converges in probability to $\frac{\Sigma_T(\alpha_T(\widehat{\theta}^1),\widehat{\theta}^1)}{\Sigma_T(\alpha_T(\theta^0),\theta^0)}$ as M goes to infinity first. Also, $\Sigma_T\left(\alpha,\theta^0\right)$ is continuously differentiable with respect to its two arguments and $\alpha_T\left(\theta^0\right)$ is continuous in θ^0. By the continuous mapping theorem, $\frac{\Sigma_T(\alpha_T(\widehat{\theta}^1),\widehat{\theta}^1)}{\Sigma_T(\alpha_T(\theta^0),\theta^0)}\to 1$ as T goes to infinity second. As M goes to infinity first, $\widehat{\alpha}_{TM}\left(\widehat{\theta}^1\right)-\alpha_T\left(\widehat{\theta}^1\right)$ converges to zero at a rate $M^{-1/2}$ (by Theorem 3-*(ii)*). By Theorem 2, $\Sigma_T\left(\alpha_T\left(\theta^0\right),\theta^0\right)$ converges to zero at a rate which is polynomial in α so that the ratio $\frac{\partial\widehat{\Sigma}_{TM}(\overline{\alpha}_T,\widehat{\theta}^1)}{\partial\alpha}/\,\Sigma_T\left(\alpha_T\left(\theta^0\right),\theta^0\right)$ diverges at the rate $1/\alpha_T\left(\theta^0\right)$, hence $T^{g(\beta)}$ where $g(\beta)<1/2$. The rate of the second term is given by $T^{g(\beta)}M^{-1/2}$. The result follows. ■

Proof of Theorem 5. From Lemma C.1, we know that $\widehat{\theta}(\alpha)\equiv\widehat{\theta}_T(\alpha,\theta_0)$ is continuously differentiable with respect to α. Using the notation $\alpha_T=\alpha_T(\theta_0)$, 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., $\exists\ \underline{\alpha}_T>0:\ \underline{\alpha}_T\le\alpha_T\le 1,\ \forall\ T$. Furthermore, the continuous differentiability 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. ■

D. NUMERICAL ALGORITHMS: COMPUTING THE OBJECTIVE FUNCTION OF THE CGMM

The moment function $h_t(\tau,\theta)\equiv h_t(\tau,\theta;\theta_0)\in L^2(\pi)$ 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:

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

This integral can be well approximated numerically by using the Gauss–Hermite quadrature. This amounts to find m points $s_1, s_2, \ldots s_m$ and weighs $\omega_1, \omega_2, \ldots \omega_m$ 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 exists 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 points does not depend on the function f. The quadrature points and weights are determined by solving:

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

Applying that method to evaluate the above integral, we get

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

Let $\widehat{h}_T(\theta)$ denote the vector $\left(\widehat{h}_T(s_1,\theta), \widehat{h}_T(s_2,\theta), \ldots, \widehat{h}_T(s_m,\theta)\right)'$ 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:

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

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:

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

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

$$\widehat{Q}_T(\alpha,\theta;\theta_0) = \left\langle K_{\alpha T}^{-1} \widehat{h}_T(.,\theta), \widehat{h}_T(.,\theta) \right\rangle \simeq \sum_{k=1}^{m} \omega_k \widetilde{h}_T(s_k,\theta)\overline{\widehat{h}_T(s_k,\theta)},$$

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