



HAL
open science

The Indirect Continuous-GMM Estimation

Rachidi Kotchoni

► **To cite this version:**

| Rachidi Kotchoni. The Indirect Continuous-GMM Estimation. 2013. <hal-00867804>

HAL Id: hal-00867804

<https://hal.science/hal-00867804v1>

Preprint submitted on 30 Sep 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

The Indirect Continuous-GMM Estimation*

Rachidi Kotchoni[†]

Assistant Professor, Université de Cergy-Pontoise.

July 28, 2013

Abstract

A *curse of dimensionality* arises when using the Continuum-GMM procedure to estimate large dimensional models. Two solutions are proposed, both of which convert the high dimensional model into a continuum of reduced information sets. Under certain regularity conditions, each reduced information set can be used to produce a consistent estimator of the parameter of interest. An indirect CGMM estimator is obtained by optimally aggregating all such consistent estimators. The simulation results suggest that the indirect CGMM procedure makes an efficient use of the information content of moment restrictions.

Keywords: Conditional moment restriction, Continuum of moment conditions, Covariance operator, Empirical characteristic function, Generalized method of moments, Indirect estimation.

*The matlab codes used to compute the results of this paper are available as supplementary material.

[†]Correspondence: THEMA, Université de Cergy-Pontoise, 33 boulevard du port, Cergy-Pontoise, 95011 Cedex, France// Email: rachidi.kotchoni@gmail.ca

1 Introduction

In the financial econometrics literature, many models are specified directly in terms of their characteristic function (CF) because their densities are unknown. Typical examples include the stable distributions and discretely sampled continuous time processes. A discrete sample from a square root diffusion is an exception (Singleton, 2001). Its transition density is of the same form as for an autoregressive Gamma model, being an infinite mixture of Gamma densities with Poisson weights (See Gouriéroux and Jasiak, 2005). Unfortunately, infinite mixtures have to be truncated in practice for the sake of feasibility. The challenge raised by the estimation of such models has motivated the use of various instances of the CF based GMM. See e.g. Carrasco and Florens (2000), Jiang and Knight (2002), Knight and Yu (2002), Yu (2004), Taufer et al. (2011) and Li et al. (2012).

In fact, two random variables have the same distribution if and only if their CFs coincide on their whole domain. This suggests that an inference procedure that adequately exploits the information content of the CF has the potential to be as efficient as a likelihood-based approach. The continuum-GMM (CGMM) proposed by Carrasco and Florens (2000) permits an efficient use of the whole continuum of moment conditions obtained by taking the difference between the theoretical CF of an IID random variable and its empirical counterpart. Carrasco, Chernov, Florens and Ghysels (2007) extend the use of the CGMM to Markov and weakly dependent models. However, the scope of the CGMM goes beyond models specified in terms of their CF. To see this, let us consider an economic model summarized by a conditional moment restriction (CMR) of the type $E(g(\theta_0, y_t)|X_t) = 0$, where $y_t \in \mathbb{R}$, $X_t \in \mathbb{R}^d$ and θ_0 is the parameter of interest. This type of CMR is quite prevalent in the macroeconomic and asset pricing literature (e.g., first order conditions of DSGEs or Euler Equations) and it is equivalent to the infinite set of unconditional moment restrictions given by: $E(g(\theta_0, y_t)A(X_t)) = 0$ for all possible functions $A(X_t)$. Dominguez and Lobato (2004) show that using a small number of unconditional moment restrictions selected from the previous infinite set may not warrant the identification of θ_0 . They show that identification and GMM-efficiency are achieved by exploiting the continuum of unconditional moment restrictions given by $E(g(\theta_0, y_t)1(X_t < \tau)) = 0$, $\tau \in \mathbb{R}^d$. Bierens (1982) showed that the CMR $E(g(\theta_0, y_t)|X_t) = 0$ is also equivalent to the continuum of unconditional moment restrictions $E(g(\theta_0, y_t)\exp(i\tau'X_t)) = 0$, $\tau \in \mathbb{R}^d$. The latter continuum shares some similarities with the one used to design the CGMM and it has also been used by Lavergne and Patilea (2008) to derive smoothed minimum distance estimators.

The CGMM builds on the same philosophy as the GMM of Hansen (1982). Both are based on the minimization of a quadratic form associated with some scalar product. The scalar product of the GMM is defined on a finite dimensional vector space while the one used to design the CGMM is defined on an infinite dimensional Hilbert space. An example of scalar product between two functions $h(\tau)$ and $g(\tau)$ in a complex Hilbert space is given by the integral of $h(\tau)\overline{g(\tau)}$ against a continuous measure $\Pi(\tau)d\tau$, where $\overline{g(\tau)}$ is the complex conjugate of $g(\tau)$. The norm of $h(\tau)$ associated with this scalar product is given by the integral of $h(\tau)\overline{h(\tau)}$ against $\Pi(\tau)d\tau$. Hence the multiplicity of the integral is determined by the dimensionality of τ . In Carrasco and Florens (2000), $h(\tau) \equiv h(\tau, \theta_0)$ is the difference between the theoretical CF of a random variable and its empirical counterpart where τ the Fourier transformation index and θ_0 a finite dimensional parameter.

Typically, one chooses $\Pi(\tau)$ to be a multivariate Gaussian measure on \mathbb{R}^d so as to be able to use Gauss-Hermite quadratures. This approach produces satisfactory results when the index τ is either

one-dimensional or two-dimensional (Kotchoni, 2012); however, the complexity of the numerical integration grows exponentially with the dimensionality of τ . For instance, if 10 quadrature points deliver a given precision in numerically evaluating a one dimensional integral, approximately 10^d quadrature points would be required to obtain an equivalent precision for a d -dimensional integral. This “*curse of dimensionality*” is well-known in computational fields. Possible solutions to address this problem include reducing the number of quadrature points or removing quadrature points that have very low weights. Unfortunately, neither of these solutions provides a substantial numerical efficiency gain without jeopardizing the accuracy of the overall estimation procedure.

In an effort to circumvent the aforementioned problem, two approaches are explored in this paper. The first approach consists of converting the multivariate model into a continuum of univariate models. To begin, one draws a vector τ from a continuous distribution defined on a standardized subset Λ of \mathbb{R}^d (e.g. the unit sphere or the unit hypercube). Next, one defines the set of all moment functions along the vector line τ as $h_{\tau,t}(u, \theta_0) \equiv h_t(u\tau, \theta_0)$, $u \in \mathbb{R}$. Under certain regularity conditions, a consistent estimator $\hat{\theta}_{CGMM}(\tau)$ is obtained by minimizing a norm of $h_{\tau,t}(u, \theta_0)$, which is a function of a one dimensional index u for a given τ . A final estimator that does not depend on τ is obtained by integrating $\hat{\theta}_{CGMM}(\tau)$ against a measure $\pi(\tau) d\tau$ on Λ . Unfortunately, some parameters that could be identified from the full information set $\{h_t(\tau, \theta_0), \tau \in \mathbb{R}^d\}$ may no longer be identifiable from the reduced information set $\{h_{\tau,t}(u, \theta_0), u \in \mathbb{R}\}$. This leads us to consider a second approach that relies on a discrete subset of the full information set for the estimation of θ_0 . Here, one draws a set of n indices τ_1, \dots, τ_n independently from a continuous distribution on \mathbb{R}^d , where n is large enough to ensure the identification of θ from $\{h_t(\tau_i, \theta_0), i = 1, \dots, n\}$. A consistent estimator $\hat{\theta}_{GMM}(\tilde{\tau})$ is then obtained by minimizing a norm of $\frac{1}{T} \sum_t \tilde{h}_t(\tilde{\tau}, \theta_0)$, where $\tilde{\tau} = (\tau_1, \dots, \tau_n)$ and $\tilde{h}_t(\tilde{\tau}, \theta_0)$ is the n -vector of relevant moment conditions. A final estimator that does not depend on $\tilde{\tau}$ is obtained by integrating $\hat{\theta}_{GMM}(\tilde{\tau})$ against a measure $\tilde{\Pi}(\tilde{\tau}) d\tilde{\tau}$ on $(\mathbb{R}^d)^n$. In either case, the final estimator consists of the aggregation of estimators obtained from samples of type $\{y_{\tau,t} = \tau'x_t\}_{t=1}^T$ generated from the frequency domain of the distribution of x_t and thus, it has the flavour of one obtained by a resampling method. It is also reminiscent of an indirect estimator because the underlying procedure converts a high dimensional model into a continuum of reduced information sets. As such, we refer to this estimator as the indirect CGMM (henceforth, ICGMM) estimator.

Three major issues are addressed regarding the ICGMM procedure. The first issue is related to the identification of θ_0 from the reduced information sets. The second issue involves the choice of the aggregating measure that warrants minimum variance for the ICGMM estimator. It appears that the optimal weighting scheme is closely related to the inverse of the covariance operators associated with the random element $\hat{\theta}_{CGMM}(\tau)$ and $\hat{\theta}_{GMM}(\tilde{\tau})$. The third issue concerns the implementation of the efficient ICGMM. We propose an implementation strategy that relies on a combination of time domain and frequency domain resampling and we show that the feasible efficient ICGMM estimator converges in probability to its theoretical counterpart.

We perform two sets of Monte Carlo experiments that are aimed at comparing the performance of the ICGMM estimator to that of feasible benchmarks (e.g., Maximum Likelihood, CGMM or Smoothed Minimum Distance). The first set of experiments is based on a bivariate Gaussian IID model. We use this simple framework as a pretext to introduce a non-technical summary of the implementation steps of the ICGMM estimator. The second simulation study is based on an enriched version of a linear heteroscedastic model used in Cragg (1983). We summarize this model

into a CMR that is turned into a continuum of unconditional moment restrictions for the purpose of applying the ICGMM procedure. The simulation results suggest that the ICGMM procedure compares favorably to the benchmarks and makes an efficient use of the information content of moment restrictions.

The remainder of the paper is organized as follows. Section 2 presents the general framework for implementing the ICGMM and introduces the necessary assumptions. Section 3 discusses the properties of standard CGMM estimators. Section 4 presents the derivation of the optimal aggregating weight for the ICGMM estimator. In Section 5, a feasible version of the efficient ICGMM estimator is presented and its asymptotic optimality established. For sake of clarity, we focus on IID models in Sections 2 to 5. The extension of the ICGMM to CMR, Markov and weakly dependent models is discussed in Section 6. Section 7 presents the Monte Carlo simulations and Section 8 concludes. The proofs of the propositions are gathered in the appendix.

2 The General Framework

This section introduces the ICGMM estimation in IID models along with the assumptions underlying the procedure.

2.1 The ICGMM estimators

Let $x_t \in \mathbb{R}^d$ be an IID random variable and assume that the distribution of x_t is fully characterized by a finite dimensional parameter $\theta_0 \in \mathbb{R}^q$. Let us consider the function $h_t(\tau, \theta)$ given by:

$$h_t(\tau, \theta) = \exp(i\tau'x_t) - \varphi(\tau, \theta), \tau \in \mathbb{R}^d, \quad (1)$$

where $\varphi(\tau, \theta) = E[\exp(i\tau'x_t)]$ and E is the expectation operator. As $E[h_t(\tau, \theta_0)] = 0$ for all $\tau \in \mathbb{R}^d$, the function $h_t(\tau, \theta)$ defines a continuum of valid moment conditions that can be used to estimate θ_0 from observed data.

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

$$\mathbf{L}^2(\Pi) = \{f : \mathbb{R}^d \rightarrow \mathbb{C} \text{ such that } \int f(\tau)\overline{f(\tau)}\Pi(\tau)d\tau < \infty\}.$$

A scalar product $\langle \cdot, \cdot \rangle$ on $\mathbf{L}^2(\Pi) \times \mathbf{L}^2(\Pi)$ is given by:

$$\langle f, g \rangle = \int f(\tau)\overline{g(\tau)}\Pi(\tau)d\tau, \quad (2)$$

where \bar{z} is the complex conjugate of z .

The moment function $h_t(\tau, \theta_0)$ is bounded in modulus and hence, belongs to $\mathbf{L}^2(\Pi)$ for any finite measure $\Pi(\tau)$. Taking advantage of this, Carrasco and Florens (2000) define the objective function of the CGMM by means of the quadratic form associated with the scalar product above. They use:

$$Q_T(\theta) = \left\langle K^{-1/2}\widehat{h}_T(\cdot, \theta), K^{-1/2}\widehat{h}_T(\cdot, \theta) \right\rangle, \quad (3)$$

where $\widehat{h}_T(\tau, \theta) = \frac{1}{T} \sum_{t=1}^T h_t(\tau, \theta)$ and K is the covariance operator associated with the moment function. This operator satisfies:

$$Kf(\tau_1) = \int k(\tau_1, \tau_2) g(\tau_2) \Pi(\tau_2) d\tau_2,$$

for any function $f \in \mathbf{L}^2(\Pi)$, where the kernel $k(\tau_1, \tau_2)$ is given by:

$$k(\tau_1, \tau_2) = E^{\theta_0} \left[h_t(\tau_1, \theta) \overline{h_t(\tau_2, \theta)} \right]. \quad (4)$$

The CGMM estimator is defined as the particular value of θ that minimizes the objective function $Q_T(\theta)$. Note that the CGMM estimation algorithm requires the iterative numerical evaluation of the multiple integrals involved in the expression of $Q_T(\theta)$. Hence, as the dimensionality of x_t grows, a *curse of dimensionality* arises and the CGMM estimator becomes numerically unfeasible.

We explore two possible approaches to circumvent this problem. In the first approach, we consider the normalized set Λ defined by:

$$\Lambda = \{ \tau \in \mathbb{R}^d : \|\tau\|_E = 1 \}, \quad (5)$$

where $\|\cdot\|_E$ is the Euclidian norm on \mathbb{R}^d (One may also choose Λ to be the hypercube $[0, 1]^d$). For any particular $\tau \in \Lambda$, the set of all moment functions along the vector line generated by τ is given by:

$$h_{\tau,t}(u, \theta) \equiv h_t(u\tau, \theta), \quad u \in \mathbb{R}, \quad (6)$$

As a function of u , $h_{\tau,t}(u, \theta)$ is a univariate mapping from \mathbb{R} to \mathbb{C} . Under certain regularity conditions discussed below, a consistent CGMM estimator of θ_0 is given by:

$$\widehat{\theta}_{CGMM}(\tau) = \arg \min_{\theta} \left\{ Q_{\tau,T}(\theta) = \int \widehat{h}_{\tau,T}(u, \theta) \overline{\widehat{h}_{\tau,T}(u, \theta)} \omega(u) du \right\}, \quad (7)$$

where $\omega(\cdot)$ is a continuous and positive weighting function on \mathbb{R} , $\widehat{h}_{\tau,T}(u, \theta) = \frac{1}{T} \sum_{t=1}^T h_{\tau,t}(u, \theta)$ and the dependence of $\widehat{\theta}_{CGMM}(\tau)$ on $\omega(\cdot)$ is hidden for simplicity. In order to make the final estimator independent of the direction τ , we define the ICGMM estimator as:

$$\widehat{\theta}_{ICGMM1} = \int \widehat{\theta}_{CGMM}(\tau) \pi(\tau) d\tau, \quad (8)$$

where $\pi(\cdot)$ is a continuous measure on Λ . An efficient ICGMM estimator is obtained by selecting $\pi(\cdot)$ so as to minimize the variance of $\widehat{\theta}_{ICGMM1}$.

In the second approach to solve the dimensionality problem, we consider estimating θ_0 from the reduced information set $\{h_t(\tau_i, \theta_0), i = 1, \dots, n\}$, where τ_1, \dots, τ_n are independent draws from a distribution $\widetilde{\pi}(\tau)$ on \mathbb{R}^d and $n \geq \dim(\theta_0)$. If θ_0 is identifiable from the full information set $\{h_t(\tau, \theta_0), \tau \in \mathbb{R}^d\}$ and n is large enough, then a consistent estimator of θ_0 is given by:

$$\widehat{\theta}_{GMM}(\widetilde{\tau}) = \arg \min \left(\overline{\frac{1}{T} \sum_t \widetilde{h}_t(\widetilde{\tau}, \theta_0)} \right)' \left(\frac{1}{T} \sum_t \widetilde{h}_t(\widetilde{\tau}, \theta_0) \right), \quad (9)$$

where $\tilde{h}_t(\tilde{\tau}, \theta_0) = (h_t(\tau_1, \theta_0), \dots, h_t(\tau_n, \theta_0))'$ and $\tilde{\tau} = (\tau_1, \dots, \tau_n)$. Note that $\hat{\theta}_{GMM}(\tilde{\tau})$ is the standard first step GMM estimator of Hansen (1982). An ICGMM estimator that does not depend on $\tilde{\tau}$ is obtained by aggregating the estimators $\hat{\theta}_{GMM}(\tilde{\tau}), \tilde{\tau} \in (\mathbb{R}^d)^n$ as follows:

$$\hat{\theta}_{ICGMM2} = \int \hat{\theta}_{GMM}(\tilde{\tau}) \tilde{\Pi}(\tilde{\tau}) d\tilde{\tau}, \quad (10)$$

where $\tilde{\Pi}(\tilde{\tau}) = \tilde{\pi}(\tau_1) \dots \tilde{\pi}(\tau_n)$. An efficient ICGMM estimator is obtained by selecting $\tilde{\pi}(\tau)$ so as to minimize the variance of $\hat{\theta}_{ICGMM2}$.

Both $\hat{\theta}_{CGMM}(\tau)$ and $\hat{\theta}_{GMM}(\tilde{\tau})$ could be used as first step estimators to build efficient second step estimators based on the reduced information sets. In the first approach, we have:

$$\hat{\theta}_{CGMM}^{(2)}(\tau) = \arg \min_{\theta} \left\{ Q_{\tau, T}^{(2)}(\theta) = \int K_{\tau}^{-1} \hat{h}_{\tau, T}(u, \theta) \overline{\hat{h}_{\tau, T}(u, \theta)} \omega(u) du \right\}, \quad (11)$$

where K_{τ} is the covariance operator associated with $h_{\tau, t}(u, \theta)$. This operator satisfies $K_{\tau} f(u_1) = \int k_{\tau}(u_1, u_2) g(u_2) \omega(u_2) du_2$ for any g that is square integrable with respect to $\omega(\cdot)$, where:

$$k_{\tau}(u_1, u_2) = E \left(h_{\tau, t}(u_1, \theta) \overline{h_{\tau, t}(u_2, \theta)} \right), \quad (12)$$

$\hat{\theta}_{CGMM}^{(2)}(\tau)$ is efficient among the CGMM estimators based on $h_{\tau, t}(u, \theta), u \in \mathbb{R}$ and its asymptotic distribution is independent of $\omega(\cdot)$ (See Carrasco and Florens, 2000). In the second approach, $\hat{\theta}_{GMM}^{(2)}(\tau)$ is the second step GMM estimator given by:

$$\hat{\theta}_{GMM}^{(2)}(\tau) = \arg \min \left(\overline{\left(\frac{1}{T} \sum_t \tilde{h}_t(\tilde{\tau}, \theta_0) \right)'} \tilde{K}^{-1} \left(\frac{1}{T} \sum_t \tilde{h}_t(\tilde{\tau}, \theta_0) \right) \right), \quad (13)$$

where \tilde{K} is the asymptotic covariance matrix of $\tilde{h}_t(\tilde{\tau}, \theta_0)$, that is, $\tilde{K} = E \left(\tilde{h}_t(\tilde{\tau}, \theta_0) \overline{\tilde{h}_t(\tilde{\tau}, \theta_0)'} \right)$.

$\hat{\theta}_{GMM}^{(2)}(\tau)$ is efficient among the GMM estimators based on $\{h_t(\tau_i, \theta_0), i = 1, \dots, n\}$ (See Hansen, 1982). However, we have preferred to design the ICGMM estimators using the first step estimators $\hat{\theta}_{CGMM}(\tau)$ and $\hat{\theta}_{GMM}(\tilde{\tau})$ for reasons that are presented in Section 4.3. The discussions on the choice of the weighting function $\omega(\cdot)$ to use for $\hat{\theta}_{ICGMM1}$ and the number of discretization points n to use for $\hat{\theta}_{ICGMM2}$ are also postponed until that section.

2.2 The Assumptions

With no loss of generality, we study the theoretical properties of the ICGMM estimator by assuming that it is given by Equation (8). Indeed, the alternate ICGMM approach is based on a standard GMM estimator whose properties are well-known in the literature. Hence from here until Section 5, it is assumed that $\hat{\theta} = \hat{\theta}_{ICGMM1}$. The following assumptions are posited.

Assumption 1: The pdf $\omega(u)$ is strictly positive on \mathbb{R} and has finite moments at any order.

Assumption 2: For all $\tau \in \Lambda \setminus \mathbb{N}$, the equation

$$E^{\theta_0} [h_{\tau, t}(u, \theta)] = 0 \text{ for all } u \in \mathbb{R}, \omega - \text{almost everywhere,}$$

has a unique solution θ_0 which is an interior point of a compact set Θ , where \aleph is a null set with respect to $\pi(\cdot)$ and E^{θ_0} denotes the expectation with respect to the distribution of the data at $\theta = \theta_0$.

Assumption 3: $\sqrt{T}\widehat{h}_{\tau,T}(u, \theta_0) \rightarrow N(0, K_\tau)$, where K_τ is the covariance operator associated with the moment function $h_{\tau,t}(\cdot, \theta_0)$.

Assumption 4: For all $\tau \in \Lambda \setminus \aleph$, $h_{\tau,t}(u, \theta)$ is twice continuously differentiable with respect to θ .

Assumption 5: $h_{\tau,t}(u, \theta)$ is at least twice continuously differentiable with respect to τ in $\Lambda \setminus \aleph$.

Assumption 6: At any θ such that $\frac{\partial Q_{\tau,T}}{\partial \theta} = 0$, we have: (i) $\frac{\partial^2 Q_{\tau,T}}{\partial \theta \partial \theta'}$ is positive definite and (ii) $\frac{\partial^2 Q_{\tau,T}}{\partial \theta \partial \tau}$ is of full rank for all $\tau \in \Lambda \setminus \aleph$.

Assumption 7: The measure $\pi(\cdot)$ on Λ satisfies: $\int \pi(\tau) d\tau = 1$.

The first assumption ensures that $0 < E_{\omega(u)} \left(f(u) \overline{f(u)} \right) < \infty$ for all $f \neq 0$.

Assumption 2 must be interpreted in two steps. First, for any given τ in $\Lambda \setminus \aleph$, there might exist a subset A of \mathbb{R} such that the solution to $E^{\theta_0} [h_{\tau,t}(u, \theta)] = 0$ is not unique for all $u \in A$. However, Assumption 2 requires that A be a null set with respect to the measure ω . Second, Assumption 2 requires that \aleph (i.e., the set of all τ such that θ_0 is not identifiable from $h_{\tau,t}(u, \theta)$) be a null set with respect to the continuous measure π on Λ . For instance, if $d = 2$ and $x_t = (x_{1,t}, x_{2,t})'$, choosing $\tau = (\tau_1, 0)$ is equivalent to relying on the marginal distribution of $x_{1,t}$ for the estimation of θ_0 . In this case, the parameters that characterize the dependence between $x_{1,t}$ and $x_{2,t}$ cannot be identified from the distribution of $y_{\tau,t} = \tau_1 x_{1,t}$. However, the set of all $\tau = (\tau_1, \tau_2)$ such that $\tau_1 = 0$ or $\tau_2 = 0$ is a null set with respect to any continuous measure on \mathbb{R}^2 . To assess the strength of Assumption 2, assume that $x_t = (x_{1,t}, x_{2,t})'$ is IID bivariate normal:

$$x_t \sim N \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix} \right), \quad (14)$$

This joint distribution is indexed by five parameters: $\theta = (\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \rho)$. Let us focus on the estimation of ρ by assuming that $(\mu_i, \sigma_i^2) = (0, 1)$, $i = 1, 2$. It can be shown that the MLE of ρ based on the joint distribution (14) solves the fixed-point relation:

$$\widehat{\rho} = \frac{\widehat{\rho}^3 - \widehat{\rho}^2 \frac{1}{T} \sum x_{1,t} x_{2,t} - \frac{1}{T} \sum x_{1,t} x_{2,t}}{1 - \frac{1}{T} \sum x_{1,t}^2 - \frac{1}{T} \sum x_{2,t}^2}. \quad (15)$$

Now, consider the distribution of $y_{\tau,t} = \tau_1 x_{1,t} + \tau_2 x_{2,t}$, with $\tau = (\tau_1, \tau_2)$. We have:

$$y_{\tau,t} = \tau_1 x_{1,t} + \tau_2 x_{2,t} \sim N \left(0, \tau_1^2 + 2\rho \tau_1 \tau_2 + \tau_2^2 \right). \quad (16)$$

The MLE of ρ based on the distribution (16) is:

$$\widehat{\rho}(\tau) = \frac{1}{2\tau_1 \tau_2} \left(\frac{1}{T} \sum y_{\tau,t}^2 - \tau_1^2 - \tau_2^2 \right). \quad (17)$$

For almost all τ , the estimator $\widehat{\rho}(\tau)$ is unbiased for ρ and is consistent. Hence, ρ is identifiable from the reduced information set consisting of the distribution of $y_{\tau,t}$ and the “weights” τ_1 and τ_2 . The reduced information set is strictly included in the joint distribution of $(z_{1,t}, z_{2,t})$, but it is larger than the sole knowledge of the marginal distribution of $y_{\tau,t}$.

Note that all five parameters $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \rho)$ cannot be identified from the reduced information set unless some restrictions are imposed (e.g., by letting σ_1^2 , μ_2 and σ_2^2 be known functions of μ_1 and focusing on the estimation of (μ_1, ρ)). In this respect, Assumption 2 is quite strong. However, this shortcoming is compensated by the generality of the approach to derive the efficient ICGMM estimator presented in Section 4 and the procedure for its implementation presented in Section 5. Indeed, the results derived for $\widehat{\theta}_{ICGMM1}$ in Sections 4 and 5 are easily extended to $\widehat{\theta}_{ICGMM2}$ (i.e., the ICGMM estimator based on $\widehat{\theta}_{GMM}(\tilde{\tau})$) upon adapting the assumptions above. Furthermore, $\widehat{\theta}_{ICGMM2}$ has the advantage of being exempt from the identification issue raised previously for $\widehat{\theta}_{ICGMM1}$. In a more complicated model, it might not be possible to detect all identification issues at glance. In such a case, one may construct bootstrap confidence sets as a means to diagnose the model. Indeed, poor bootstrap confidence set coverage strongly suggest that some of the parameters of interest are (almost) unidentified (See Dufour, 1997).

The remaining assumptions are quite mild. Assumption 3 is satisfied if x_t is IID (See Carrasco and Florens, 2000). The consistency of the CGMM estimator $\widehat{\theta}_{CGMM}(\tau)$ can be shown under a weaker condition than Assumption 4, e.g. that $h_{\tau,t}(u, \theta)$ is once continuously differentiable. However, twice continuous differentiability is required along with Assumptions 5 and 6 to ensure that $\widehat{\theta}_{CGMM}(\tau)$ is a smooth function of τ . Assumption 6 further implies that $\widehat{\theta}_{CGMM}(\tau)$ is unique. Indeed, there is no guaranty that a CGMM objective function computed from an arbitrary small sample will have a unique minimizer. However, for reasonable sample sizes and if the model underlying the moment function is identified, we can expect $\widehat{\theta}_{CGMM}(\tau)$ to be unique “for almost all τ ,” which is enough for the goals pursued in this paper. Finally, the measure $\pi(\cdot)$ to which Assumption 7 refers could be any continuous pdf on Λ .

3 Properties of CGMM Estimators

The CGMM estimator $\widehat{\theta}_{CGMM}(\tau)$ is based on the reduced information set $\{h_{\tau,t}(u, \theta), u \in \mathbb{R}\}$. The objective function that it minimizes does not use the inverse of the covariance operator K_τ as metrics. Hence, it is not the most optimal estimator that can be obtained from this reduced information set. However, its consistency is established by the following proposition.

Proposition 1 *Under Assumptions 1 to 4, $\widehat{\theta}_{CGMM}(\tau)$ is consistent for θ_0 . It is asymptotically normal and we have:*

$$\sqrt{T} \left(\widehat{\theta}_{CGMM}(\tau) - \theta_0 \right) \rightarrow N \left(0, W_\tau^{-1} \langle G_\tau(\cdot, \theta_0), K_\tau G_\tau(\cdot, \theta_0) \rangle W_\tau^{-1} \right),$$

as $T \rightarrow \infty$ and for all $\tau \in \Lambda \setminus \mathcal{N}$, where $G_\tau(\cdot, \theta_0) = P \lim \frac{\partial h_{\tau,t}(u_1, \widehat{\theta}_{CGMM}(\tau))}{\partial \theta}$, $W_\tau = \langle G_\tau(\cdot, \theta_0), G_\tau(\cdot, \theta_0) \rangle$ and K_τ is the covariance operator associated with $h_{\tau,t}(u, \theta)$.

A more general version of this result is stated in Carrasco, Chernov, Florens and Ghysels (2007, Proposition 3.1). Proposition 1 will be used later to prove the consistency and asymptotic normality of the ICGMM estimator. Another property of $\widehat{\theta}_{CGMM}(\tau)$ established below warrants attention.

Proposition 2 *Under Assumptions 1 to 6, $\widehat{\theta}_{CGMM}(\tau)$ is unique and continuously differentiable with respect to τ , for all $\tau \in \Lambda \setminus \mathcal{N}$.*

The result given by Proposition 2 allows us to consider the use of a continuous pdf $\pi(\cdot)$ as weighting functions for the ICGMM estimator. Later on, we will attempt to derive the optimal weighting function $\pi^*(\cdot)$.

If one wishes to compute the second step CGMM $\widehat{\theta}_{CGMM}^{(2)}(\tau)$, an estimate of the covariance operator K_τ is needed. A natural estimator of K_τ is given by the linear empirical operator $K_{\tau,T}$ with kernel:

$$\widehat{k}_\tau(u_1, u_2) = \frac{1}{T} \sum_{t=1}^T h_{\tau,t}(u_1, \widehat{\theta}_{CGMM}(\tau)) \overline{h_{\tau,t}(u_2, \widehat{\theta}_{CGMM}(\tau))}, \quad (18)$$

where $\widehat{\theta}_{CGMM}(\tau)$ is used as first step estimator. In IID models, the first step estimator may be bypassed by considering:

$$\widehat{k}_\tau(u_1, u_2) = \frac{1}{T} \sum_{t=1}^T (e^{iu_1 y_{\tau,t}} - \widehat{\varphi}_{\tau,T}) \overline{(e^{iu_2 y_{\tau,t}} - \widehat{\varphi}_{\tau,T})}, \quad (19)$$

where $\widehat{\varphi}_{\tau,T} = \frac{1}{T} \sum_{t=1}^T e^{iu_1 y_{\tau,t}}$.

The operator K_τ has an infinite and discrete spectrum. By letting $l_{\tau,i}$ be its eigenvalue associated with the eigenfunction $\psi_{\tau,i}$ and assuming that $l_{\tau,i}$ is decreasing in i , we have: (i) $l_{\tau,1} < \infty$, (ii) $l_{\tau,i} > l_{\tau,i+1} > 0$ for all i , and (iii) $\lim_{i \rightarrow \infty} l_{\tau,i} = 0$. By contrast, the empirical operator $K_{\tau,T}$ has a degenerate spectrum. If we let $\widehat{l}_{\tau,i}$ be an eigenvalue of $K_{\tau,T}$ associated with the eigenfunction $\widehat{\psi}_{\tau,i}$, then it is always possible to label $\widehat{l}_{\tau,i}$ and $\widehat{\psi}_{\tau,i}$ so that: (i) $\widehat{l}_{\tau,1} < \infty$, (ii) $\widehat{l}_{\tau,i} > \widehat{l}_{\tau,i+1} \geq 0$ for all i , and (iii) $\widehat{l}_{\tau,i} = 0$ for all $i > T$, where T is the sample size. As a result, $K_{\tau,T}$ is not invertible on $L^2(\omega)$. See Carrasco, Florens and Renault (2007) for more details.

To estimate K_τ^{-1} , the following generalized inverse is used:

$$K_{\tau,T,\alpha}^{-1} = (K_{\tau,T}^2 + \alpha I)^{-1} K_{\tau,T}.$$

With the same notation as above, it can be verified that $\widehat{\psi}_{\tau,i}$ is an eigenfunction of $K_{\tau,T,\alpha}^{-1}$ associated with the eigenvalue $\frac{\widehat{l}_{\tau,i}}{\widehat{l}_{\tau,i}^2 + \alpha}$. Under Assumptions 1 and 2, we have:

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

where K is the covariance operator defined in equation (3). The regularized inverse $K_{\tau,T,\alpha}^{-1}$ has the property that for any function f in the range of $K_{\tau,T}^{1/2}$, the function $K_{\tau,T,\alpha}^{-1/2} f$ converges to $K^{-1/2} f$ as T goes to infinity and α goes to zero. Under Assumptions 1 to 4 and an additional regularity condition on the moment function $h_{\tau,t}(u, \theta)$ (See e.g. Carrasco and Florens, 2000), replacing $K_\tau^{-1/2}$ by $K_{\tau,T,\alpha}^{-1/2}$ in the objective function (11) yields a second step estimator that satisfies:

$$\sqrt{T} \left(\widehat{\theta}_{CGMM}^{(2)}(\tau) - \theta_0 \right) \rightarrow N(0, I_{\tau,\theta_0}^{-1}), \quad (20)$$

as T and $\alpha T^{1/2}$ go to infinity and α goes to zero, where I_{τ,θ_0}^{-1} is the asymptotic variance of the MLE based on the reduced information set.

4 The Efficient ICGMM Estimator

In Equation (8), we have defined the ICGMM estimator as the weighted sum of a continuum of \sqrt{T} -consistent estimators indexed by τ , that is:

$$\widehat{\theta} = \int \widehat{\theta}_{CGMM}(\tau) \pi(\tau) d\tau,$$

where $\pi(\cdot)$ is a continuous measure on Λ that sums to one. We have the following consistency result.

Proposition 3 *The ICGMM estimator $\widehat{\theta}$ is consistent and asymptotically normal for any continuous pdf $\pi(\cdot)$ on Λ . We have:*

$$\sqrt{T} \left(\widehat{\theta} - \theta_0 \right) \rightarrow N(0, \Omega_\pi),$$

as $T \rightarrow \infty$, where:

$$\Omega_\pi = \int \int [W_{\tau_1}^{-1} \langle G_{\tau_1}(\cdot, \theta_0), K_{\tau_1, \tau_2} G_{\tau_2}(\cdot, \theta_0) \rangle W_{\tau_2}^{-1}] \pi(\tau_1) \pi(\tau_2) d\tau_1 d\tau_2,$$

and K_{τ_1, τ_2} is the linear operator with kernel:

$$k_{\tau_1, \tau_2}(u, v) = \lim_{T \rightarrow \infty} Cov \left(\sqrt{T} \widehat{h}_{\tau_1, T}(u, \theta_0), \sqrt{T} \widehat{h}_{\tau_2, T}(v, \theta_0) \right),$$

and $G_\tau(\cdot, \theta_0) = P \lim \frac{\partial h_{\tau, t}(u, \theta_0)}{\partial \theta}$ and $W_\tau = \langle G_\tau(\cdot, \theta_0), G_\tau(\cdot, \theta_0) \rangle$ for any τ .

Below, we derive the optimal weighting function $\pi^*(\cdot)$.

4.1 Approximate Solution to the Exact Problem

We consider selecting the optimal weighting function $\pi(\cdot)$ by minimizing the variance of $\sqrt{T} \lambda' \widehat{\theta}$, for some $\lambda \in \mathbb{R}^q$:

$$Var \left(\sqrt{T} \lambda' \widehat{\theta} \right) = \int \int g_\lambda(\tau_1, \tau_2) \pi(\tau_1) \pi(\tau_2) d\tau_1 d\tau_2, \quad (21)$$

where $g_\lambda(\tau_1, \tau_2) = \lambda' Cov \left(\sqrt{T} \widehat{\theta}(\tau_1), \sqrt{T} \widehat{\theta}(\tau_2) \right) \lambda$. Clearly, the solution of this minimization depends on λ . In practice, λ should be set according to the particular hypothesis one wishes to test on θ_0 . For example, $\lambda' \widehat{\theta}$ is the sum of the coordinates of $\widehat{\theta}$ when $\lambda = (1, \dots, 1)'$, $\lambda' \widehat{\theta}$ selects the first coordinate of $\widehat{\theta}$ when $\lambda = (1, 0, \dots, 0)'$, and so on.

The ideal measure $\pi_\lambda^*(\cdot)$ solves:

$$\pi_\lambda^* = \arg \min_{\pi} \int \int g_\lambda(\tau_1, \tau_2) \pi(\tau_1) \pi(\tau_2) d\tau_1 d\tau_2, \quad (22)$$

subject to $\int \pi(\tau) d\tau = 1$. The Lagrangian for this problem is given by:

$$\mathcal{L}(\pi) = \int \int g_\lambda(\tau_1, \tau_2) \pi(\tau_1) \pi(\tau_2) d\tau_1 d\tau_2 + \mu_\lambda \left(1 - \int \pi(\tau_1) d\tau_1 \right),$$

Let V_λ be the asymptotic covariance operator associated with $\lambda' \widehat{\theta}_{CGMM}(\tau)$. The operator V_λ is linear and its kernel is given by $g_\lambda(\tau_1, \tau_2)$. It is also compact if we have:

$$\int_{\Lambda} \int_{\Lambda} [g_\lambda(\tau_1, \tau_2)]^2 d\tau_1 d\tau_2 < \infty.$$

This condition is met here because Λ is a bounded set while $g_\lambda(\tau_1, \tau_2)$ is finite and continuous at all (τ_1, τ_2) . The compactness of V_λ ensures that it has a discrete spectrum (See Carrasco, Florens and Renault, 2007). If we let $\phi_{\lambda,j}(\tau)$ denote the eigenfunction of V_λ associated with the eigenvalue $\nu_{\lambda,j}$, then we have $\nu_{\lambda,j} \geq 0$ and $\phi_{\lambda,i}(\tau)$ and $\phi_{\lambda,j}(\tau)$ are orthogonal for all $i \neq j$. The first order condition of the minimization problem above is:

$$V_\lambda \pi(\tau) = \mu_\lambda J(\tau),$$

where μ_λ is a Lagrange multiplier and $J(\tau) = 1$ for all τ .

An exact solution to this problem exists only if the constant function $J(\tau)$ lies entirely in the range of V_λ . To our knowledge, it cannot be proved that this is the case. Hence, we target an approximate solution that exploits the projection of $J(\tau)$ onto the range of V_λ . This approximate solution coincides with the exact solution if $J(\tau)$ lies entirely in the range of V_λ . The function $J(\tau)$ is decomposed as:

$$J(\tau) = J_0(\tau) + \widetilde{J}(\tau),$$

where

$$J_0(\tau) = \sum_{j=1}^{\infty} \left(\int \phi_{\lambda,j}(\tau_1) d\tau_1 \right) \phi_{\lambda,j}(\tau),$$

is the projection of $J(\tau)$ onto the range of V_λ and the residual of the projection $\widetilde{J}(\tau) = 1 - J_0(\tau)$ lies in the null space of V_λ . The following proposition characterizes the approximate solution to our optimization problem.

Proposition 4 *The approximate solution of (22) with minimal norm is given by:*

$$\pi_\lambda^*(\tau) = \mu_\lambda^* \sum_{j=1}^{\infty} \frac{1}{\nu_{\lambda,j}} \left(\int \phi_{\lambda,j}(\tau_1) d\tau_1 \right) \phi_{\lambda,j}(\tau), \quad (23)$$

where

$$\mu_\lambda^* = \left[\sum_{j=1}^{\infty} \frac{1}{\nu_{\lambda,j}} \left(\int \phi_{\lambda,j}(\tau) d\tau \right)^2 \right]^{-1}.$$

At the optimum, the variance of $\lambda' \widehat{\theta}_\lambda^*$, is given by:

$$Var \left(\sqrt{T} \lambda' \widehat{\theta}_\lambda^* \right) = \mu_\lambda^*. \quad (24)$$

This approximate solution is the one with minimal norm because $\pi_\lambda^*(\tau) + \widetilde{f}(\tau)$ solves the same first order condition as $\pi_\lambda^*(\cdot)$ for any $\widetilde{f}(\tau)$ that belongs to the null space of V_λ . More compact

expressions than those provided in Proposition (4) are given by:

$$\begin{aligned}\mu_\lambda^* &= \left(\int V_\lambda^{-1} J_0(\tau) d\tau \right)^{-1}, \quad \pi_\lambda^*(\tau) = \mu_\lambda^* V_\lambda^{-1} J_0(\tau) \quad \text{and} \\ \widehat{\theta}_\lambda^* &= \mu_\lambda^* \int [V_\lambda^{-1} J_0(\tau)] \widehat{\theta}_{CGMM}(\tau) d\tau.\end{aligned}$$

We now consider an alternative approach to derive the optimal weights.

4.2 Exact Solution to an Approximate Problem

The presence of the inverse of the eigenvalues of V_λ in the expression of $\pi_\lambda^*(\cdot)$ makes this solution ill-posed. This means that a small perturbation in the eigenvalues of V_λ may create large changes in the behavior of $\pi_\lambda^*(\cdot)$. A well-behaved solution can be found by adding a Ridge penalization to the Lagrangian. This yields:

$$\begin{aligned}\mathcal{L}(\pi) &= \int \int g_\lambda(\tau_1, \tau_2) \pi(\tau_1) \pi(\tau_2) d\tau_1 d\tau_2 + \mu_\lambda \left(1 - \int \pi(\tau_1) d\tau_1 \right) \\ &\quad + \alpha \int \pi(\tau_1)^2 d\tau_1,\end{aligned}\tag{25}$$

The first order optimality condition for this approximate problem is:

$$(V_\lambda + \alpha I) \pi(\tau) = \mu_\lambda J(\tau),$$

where I stands for the identity operator. Interestingly, the null space of the operator $V_\lambda + \alpha I$ reduces to the null function. Hence, an exact solution can be derived as:

$$\pi_{\lambda,\alpha}^*(\tau) = \mu_\lambda (V_\lambda + \alpha I)^{-1} J(\tau).$$

Using the constraint $1 - \int \pi_{\lambda,\alpha}^*(\tau) d\tau = 0$, we find the Lagrange multiplier as:

$$\mu_{\lambda,\alpha}^* = \left(\int (V_\lambda + \alpha I)^{-1} J(\tau) d\tau \right)^{-1}.\tag{26}$$

The exact solution to the regularized problem is given by:

$$\pi_{\lambda,\alpha}^*(\tau) = \mu_{\lambda,\alpha}^* (V_\lambda + \alpha I)^{-1} J(\tau).\tag{27}$$

Finally, the optimal ICGMM estimator based on $\pi_{\lambda,\alpha}^*(\tau)$ is given by:

$$\widehat{\theta}_{\lambda,\alpha}^* \equiv \int \pi_{\lambda,\alpha}^*(\tau) \widehat{\theta}_{CGMM}(\tau) d\tau = \mu_{\lambda,\alpha}^* \int (V_\lambda + \alpha I)^{-1} \widehat{\theta}_{CGMM}(\tau) d\tau,\tag{28}$$

and the variance of $\widehat{\theta}_{\lambda,\alpha}^*$ is given by $\mu_{\lambda,\alpha}^*$. The solution based on the penalized Lagrangian (25) suggests a family of regularized solutions given by:

$$\begin{aligned}\mu_{\lambda,\alpha}^* &= \left(\int V_{\lambda,\alpha}^{-1} J(\tau) d\tau \right)^{-1}, \quad \pi_{\lambda,\alpha}^*(\tau) = \mu_{\lambda,\alpha}^* V_{\lambda,\alpha}^{-1} J(\tau) \quad \text{and} \\ \widehat{\theta}_{\lambda,\alpha}^* &= \mu_{\lambda,\alpha}^* \int V_{\lambda,\alpha}^{-1} \widehat{\theta}_{CGMM}(\tau) d\tau,\end{aligned}$$

where $V_{\lambda,\alpha}^{-1}$ is a regularized inverse of the covariance operator. Subsequently, we advocate the following regularization scheme:

$$V_{\lambda,\alpha}^{-1} = (V_\lambda^2 + \alpha I)^{-1} V_\lambda. \quad (29)$$

This choice leads to a solution $\pi_{\lambda,\alpha}^*$ that lies entirely in the range of V_λ and that converges to the approximate solution of the exact problem as α goes to zero.

4.3 Related Issues: how to choose the weighting measure $\omega(\cdot)$ or the number of discretization n ?

This section discusses the suitable choice of the measure $\omega(\cdot)$ to use for the implementation of $\widehat{\theta}_{ICGMM1}$ and the number of discretization points n to use for the implementation of $\widehat{\theta}_{ICGMM2}$. By the same token, heuristic arguments are put forth to justify our choice to implement $\widehat{\theta}_{ICGMM1}$ and $\widehat{\theta}_{ICGMM2}$ based on the first step estimators $\widehat{\theta}_{CGMM}(\tau)$ and $\widehat{\theta}_{GMM}(\tilde{\tau})$ rather than on the second step estimators $\widehat{\theta}_{CGMM}^{(2)}(\tau)$ and $\widehat{\theta}_{GMM}^{(2)}(\tilde{\tau})$.

We first consider $\widehat{\theta}_{ICGMM1}$ and the choice of $\omega(\cdot)$. Recall that the first step CGMM estimator $\widehat{\theta}_{CGMM}(\tau)$ solves:

$$\widehat{\theta}_{CGMM}(\tau) = \arg \min_{\theta} \left\{ \int \widehat{h}_{\tau,T}(u, \theta) \overline{\widehat{h}_{\tau,T}(u, \theta)} \omega(u) du \right\},$$

where ω is a continuous univariate measure on \mathbb{R} . Thus, let us make the dependence of everything on ω explicit (in this section only). The Efficient ICGMM estimator consists of a summation of $\widehat{\theta}_{CGMM}(\tau) \equiv \widehat{\theta}_{CGMM}(\tau, \omega)$ along the index τ and hence, it also depends on ω . The approximate solution to the exact problem implies an efficient ICGMM estimator of the form:

$$\widehat{\theta}_\lambda^*(\omega) = \mu_\lambda^*(\omega) \int \pi_\lambda^*(\tau, \omega) \widehat{\theta}_{CGMM}(\tau, \omega) d\tau,$$

where $\mu_\lambda^*(\omega)$ and $\pi_\lambda^*(\tau, \omega)$ depend on $\omega(\cdot)$ via the covariance operator $V_\lambda \equiv V_{\lambda,\omega}$. The asymptotic variance of $\widehat{\theta}_\lambda^*(\omega)$ is given by:

$$\mu_\lambda^*(\omega) = \left[\sum_{j=1}^{\infty} \frac{1}{\nu_{\lambda,j}(\omega)} \left(\int \phi_{\lambda,j}(\tau, \omega) d\tau \right)^2 \right]^{-1}, \quad (30)$$

where $\phi_{\lambda,j}(\tau, \omega)$ is the eigenfunction of $V_{\lambda,\omega}$ associated with the eigenvalue $\nu_{\lambda,j}(\omega)$. Thus, minimizing the asymptotic variance of $\widehat{\theta}_\lambda^*(\omega)$ amounts to minimizing the eigenvalues $\nu_{\lambda,j}(\omega)$ while maximizing the squared integrals of the eigenfunctions $(\int \phi_{\lambda,j}(\tau, \omega) d\tau)^2$, $j = 1, \dots, \infty$.

To gain additional insights, let us base our reasoning on the sum of the eigenvalues, $\sum_{j=1}^{\infty} \nu_{\lambda,j}^2(\omega)$,

and the sum of the squared integrals, $\sum_{j=1}^{\infty} \left(\int \phi_{\lambda,j}(\tau_1, \omega) d\tau_1 \right)^2$. First, observe that:

$$\begin{aligned} & \int_{\Lambda} \int_{\Lambda} [g_{\lambda}(\tau_1, \tau_2, \omega)]^2 d\tau_1 d\tau_2 \\ &= \int_{\Lambda} \left(\sum_{j=1}^{\infty} \nu_{\lambda,j}(\omega) \left(\int g_{\lambda}(\tau_1, \tau_2, \omega) \phi_{\lambda,j}(\tau_2, \omega) d\tau_2 \right) \phi_{\lambda,j}(\tau_1, \omega) \right) d\tau_1, \\ &= \int_{\Lambda} \left(\sum_{j=1}^{\infty} \nu_{\lambda,j}^2(\omega) \phi_{\lambda,j}^2(\tau_1, \omega) \right) d\tau_1 = \sum_{j=1}^{\infty} \nu_{\lambda,j}^2(\omega). \end{aligned}$$

Hence, $\sum_{j=1}^{\infty} \nu_{\lambda,j}^2(\omega)$ is made small by selecting the measure $\omega(\cdot)$ so as to minimize the variance of each $\widehat{\theta}_{CGMM}(\tau, \omega)$ for any τ . Second, note that $\sum_{j=1}^{\infty} \left(\int \phi_{\lambda,j}(\tau_1, \omega) d\tau_1 \right)^2$ is the norm of $J_0(\tau)$, that is, the projection of the function $J(\tau) = 1, \tau \in \mathbb{R}^d$ onto the range of $V_{\lambda,\omega}$. This norm is increased if the subspace of L^2 spanned by the eigenfunctions of $V_{\lambda,\omega}$ captures a larger part of $J(\tau)$, or alternatively, if the part of $J(\tau)$ that falls into the null space of $V_{\lambda,\omega}$ is made as small as possible (L^2 is the set of square integrable function $f : \Lambda \rightarrow \mathbb{R}$ with respect to the Lebesgue measure). This is achieved if the functions $\lambda' \widehat{\theta}_{CGMM}(\tau, \omega)$ are scattered along more and more dimensions in L^2 as τ varies in Λ (it helps to think of these dimensions as being identified by the elements of a complete orthonormal sequence in L^2).

The first condition for the optimality of $\omega(\cdot)$ (i.e., minimization of the eigenvalues of $V_{\lambda,\omega}$) concerns the variability of $\lambda' \widehat{\theta}_{CGMM}(\tau, \omega)$ across all possible random samples of size T . The second condition (i.e., maximization of the norm of $J_0(\tau)$) is related to the dimensionality of the subset of L^2 given by $\left\{ \lambda' \widehat{\theta}_{CGMM}(\tau, \omega), \tau \in \Lambda \right\}$. These two conditions are likely to be met by an estimator that is consistent and at the same time quite sensitive to variations in τ . Arguably, $V_{\lambda,\omega}$ has larger eigenvalues compared to the covariance operator associated with $\lambda' \widehat{\theta}_{CGMM}^{(2)}(\tau)$ (since the second step CGMM estimator has smaller variance than $\lambda' \widehat{\theta}_{CGMM}(\tau, \omega)$). However, $\lambda' \widehat{\theta}_{CGMM}^{(2)}(\tau)$ may be less sensitive to variations in τ than $\lambda' \widehat{\theta}_{CGMM}(\tau, \omega)$. This suggests that an efficient ICGMM designed as a linear combination of first step CGMM estimators and based on a reasonable choice of $\omega(\cdot)$ (e.g. the normal density of \mathbb{R}) can be as efficient as one using the second step estimator. These heuristic arguments are supported by the simulation results presented in Section 7 where the ICGMM is compared to maximum likelihood. Furthermore, the computation of $\widehat{\theta}_{CGMM}(\tau, \omega)$ does not require the estimation of the covariance operator associated with the moment conditions and hence, it is obtained in a shorter length of time and with a lesser computational burden than $\widehat{\theta}_{CGMM}^{(2)}(\tau)$.

We now discuss the choice of the number of discretization points when computing $\widehat{\theta}_{ICGMM2}$. Let $\widehat{\theta}_{GMM}(\tilde{\tau}, n)$ and $\widehat{\theta}_{GMM}^{(2)}(\tilde{\tau}, n)$ denote a first step and a second step GMM estimators based on n discretization points. In principle, $n = \dim(\theta_0)$ distinct discretization points should be enough for identification. However, the discretization points are being randomly drawn from a continuum and chances are that two of them be arbitrarily close for a given draw. Furthermore, for some data generating processes, the identifiability of θ_0 from the full information set might be rather weak. For these reasons, caution dictates to set n as large as possible. Let now assume that θ_0 is strongly identifiable from (almost) all reduced information sets consisting of n moment conditions. Our discussion above on the choice of the measure $\omega(\cdot)$ suggests that for a given n , an ICGMM2

estimator based on the first step estimator $\widehat{\theta}_{GMM}(\tilde{\tau}, n)$ is not necessarily more efficient than one based on the second step estimator $\widehat{\theta}_{GMM}^{(2)}(\tilde{\tau}, n)$. Likewise, although $\widehat{\theta}_{GMM}(\tilde{\tau}, n_2)$ is expected to be more efficient than $\widehat{\theta}_{GMM}(\tilde{\tau}, n_1)$ for $n_2 > n_1$, an ICGMM2 procedure based on $\widehat{\theta}_{GMM}(\tilde{\tau}, n_2)$ is not necessarily more efficient than one based on $\widehat{\theta}_{GMM}(\tilde{\tau}, n_1)$. The simulation results suggest that the performance of $\widehat{\theta}_{ICGMM2}$ is not very sensitive to the choice of n (see Section 7). In practice, the ICGMM2 estimator can be fine-tuned by trying different values of n and selecting the one that delivers the smallest bootstrap RMSE.

5 The Feasible Efficient ICGMM

Our goal is to compute the estimator $\widehat{\theta}_{\lambda, \alpha}^* = \int \widehat{\theta}_{CGMM}(\tau) \pi_{\lambda, \alpha}^*(\tau) d\tau$. Unfortunately, the expression of $\pi_{\lambda, \alpha}^*(\tau)$ is not stated in closed form. To circumvent this difficulty, we consider approximating $\widehat{\theta}_{\lambda, \alpha}^*$ by the Monte Carlo average:

$$\widehat{\theta}_{\lambda, \alpha}^* \simeq \frac{1}{S} \sum_{i=1}^S \pi_{\lambda, \alpha}^*(\tau_i) \widehat{\theta}(\tau_i), \quad (31)$$

where $\tau_i, i = 1, \dots, S$ are independent draws from the uniform distribution on Λ . The formula (31) can be implemented upon knowing how to estimate $\pi_{\lambda}^*(\cdot)$. The remainder of this section is devoted to the derivation of an estimator $\widehat{\pi}_{\lambda}^*(\cdot)$ of the optimal weighting function $\pi_{\lambda}^*(\cdot)$.

Let $\tau_i, i = 1, \dots, S$ be S draws from the multivariate uniform distribution on Λ and assume that we can draw samples from the data generating process of x_t . Further let $\left\{x_t^{(l)}\right\}_{t=1}^T, l = 1, \dots, L$ be L independent samples of size T simulated from the distribution of interest. For each sample indexed by l and each possible τ , we compute the univariate samples:

$$\left\{y_{\tau, t}^{(l)}\right\} = \left\{\tau' x_t^{(l)}\right\}, l = 1, \dots, L. \quad (32)$$

Finally, let $\widehat{\theta}(\tau, l)$ be the first step CGMM estimator based on the sample $\left\{y_{\tau, t}^{(l)}\right\}$. Note that the CGMM estimator $\widehat{\theta}(\tau_i, l), l = 1, \dots, L$ are IID copies of $\widehat{\theta}(\tau_i)$.

An estimator of the asymptotic covariance operator associated with $\widehat{\theta}_{CGMM}(\tau)$ is given by the following $(S \times S)$ matrix:

$$\widehat{V}_{\lambda} = \frac{T}{SL} \left(\widehat{\Theta}_{\lambda} - \overline{\Theta}_{\lambda}\right)' \left(\widehat{\Theta}_{\lambda} - \overline{\Theta}_{\lambda}\right), \quad (33)$$

where $\widehat{\Theta}_{\lambda}$ is the matrix with (l, i) element given by $\lambda' \widehat{\theta}(\tau_i, l)$ and $\overline{\Theta}_{\lambda}$ contains the means of the columns of $\widehat{\Theta}_{\lambda}$. The (l, i) element of \widehat{V}_{λ} is:

$$\widehat{g}_{\lambda}(\tau_i, \tau_j) = \frac{T}{SL} \sum_{l=1}^L \lambda' \left(\widehat{\theta}(\tau_i, l) - \overline{\theta}(\tau_i, l)\right) \left(\widehat{\theta}(\tau_j, l) - \overline{\theta}(\tau_j, l)\right)' \lambda, \quad (34)$$

where $\overline{\theta}(\tau_i, l) = \frac{1}{L} \sum_{l=1}^L \widehat{\theta}(\tau_i, l)$. Given that θ_0 is unknown, we may use $\widehat{\theta}_S = \frac{1}{S} \sum_{i=1}^S \widehat{\theta}(\tau_i)$ as a proxy to simulate the independent samples $\left\{x_t^{(l)}\right\}_{t=1}^T, l = 1, \dots, L$.

Instead of using Monte Carlo simulations to obtain the samples $\left\{x_t^{(l)}\right\}_{t=1}^T$, $l = 1, \dots, L$, one may also resort to a standard time domain resampling to generate bootstrap copies $\widehat{\theta}(\tau_i, l)$, $l = 1, \dots, L$ of $\widehat{\theta}(\tau_i)$. We have the following convergence result for \widehat{V}_λ .

Proposition 5 *Let $\underline{f} = (f(\tau_1), \dots, f(\tau_S))'$ where τ_1, \dots, τ_S are S draws from the multivariate uniform distribution on $\bar{\Lambda}$ and f is continuous. Then as L and S go to infinity, we have:*

$$\left(\widehat{V}_\lambda \underline{f}\right)_i - V_\lambda f(\tau_i) = O_p(L^{-1/2}) + O_p(S^{-1/2}),$$

for all τ_i , where $\left(\widehat{V}_\lambda \underline{f}\right)_i$ is the i^{th} element of the vector $\widehat{V}_\lambda \underline{f}$.

We estimate the optimal weighting function $\pi_{\lambda, \alpha}^*$ by:

$$\widehat{\pi}_{\lambda, \alpha}^* = \left(\widehat{\pi}_{\lambda, \alpha}^*(\tau_1), \dots, \widehat{\pi}_{\lambda, \alpha}^*(\tau_S)\right)' = S \left(\iota' \widehat{V}_{\lambda, \alpha}^{-1} \iota\right)^{-1} \iota' \widehat{V}_{\lambda, \alpha}^{-1}, \quad (35)$$

where ι is a vector of ones and $\widehat{V}_{\lambda, \alpha}^{-1}$ is a regularized inverse inspired by (29):

$$\widehat{V}_{\lambda, \alpha}^{-1} = \left(\widehat{V}_\lambda^2 + \alpha I\right)^{-1} \widehat{V}_\lambda, \quad \alpha \in (0, 1). \quad (36)$$

The following convergence result holds for $\widehat{V}_{\lambda, \alpha}^{-1}$.

Proposition 6 *Let $\underline{f} = (f(\tau_1), \dots, f(\tau_S))'$ where τ_1, \dots, τ_S are S draws from the multivariate uniform distribution on $\bar{\Lambda}$ and f is continuous. Then, under Assumptions 1 to 7, we have:*

$$\left\|\widehat{V}_{\lambda, \alpha}^{-1} \underline{f} - V_{\lambda, \alpha}^{-1} \underline{f}\right\| = O_p(\alpha^{-3/2} L^{-1/2}) + O_p(\alpha^{-3/2} S^{-1/2}), \quad (37)$$

where $V_{\lambda, \alpha}^{-1} = (V_{\lambda, \alpha}^2 + \alpha I) V_\lambda$.

Hence, $\widehat{V}_{\lambda, \alpha}^{-1} \underline{f}$ is consistent for $V_{\lambda, \alpha}^{-1} \underline{f}$ provided that $\alpha^3 L \rightarrow \infty$ and $\alpha^3 S \rightarrow \infty$. In particular, we have:

$$\begin{aligned} \left\|\widehat{V}_{\lambda, \alpha}^{-1} \iota - V_{\lambda, \alpha}^{-1} J\right\| &= \left\|\widehat{V}_{\lambda, \alpha}^{-1} \iota - V_{\lambda, \alpha}^{-1} J_0\right\|, \\ &= O_p(\alpha^{-3/2} L^{-1/2}) + O_p(\alpha^{-3/2} S^{-1/2}), \end{aligned}$$

where ι is a vector of ones, $J(\tau) = 1$ for all τ and J_0 is the projection of J onto the range of V_λ . Now, we need to find the rate at which $V_{\lambda, \alpha}^{-1} J_0$ converges to $V_\lambda^{-1} J_0$, where:

$$V_\lambda^{-1} J_0(\tau) = \sum_{j=1}^{\infty} \frac{1}{\nu_{\lambda, j}} \left(\int \phi_{\lambda, j}(\tau_1) d\tau_1\right) \phi_{\lambda, j}(\tau).$$

Note that the expression above is proportional to $\pi_\lambda^*(\cdot)$ (See Proposition 4).

If J_0 belongs to the range of V_λ^3 , then we could write:

$$\begin{aligned} (V_{\lambda, \alpha}^{-1} - V_\lambda^{-1}) f &\equiv ((V_\lambda^2 + \alpha I)^{-1} V_\lambda - V_\lambda^{-1}) f, \\ &= -\alpha [(V_\lambda^2 + \alpha I)^{-1} V_\lambda^2] [V_\lambda^{-3} f_0]. \end{aligned}$$

This would imply that $\|V_\lambda^{-3} J_0\| = O(1)$ and $\|(V_\lambda^2 + \alpha I)^{-1} V_\lambda^2\| = O(1)$, and thus $(V_{\lambda, \alpha}^{-1} - V_\lambda^{-1}) J_0 = O(\alpha)$ as $\alpha \rightarrow 0$. Unfortunately, the requirement that J_0 belongs to the range of V_λ^3 might be too restrictive. The following proposition allows for more generality.

Proposition 7 Let $J(\tau) = 1$ for all τ and J_0 be the projection of J onto the range of V_λ . Then, under Assumptions 1 to 7, we have:

$$\|(V_{\lambda,\alpha}^{-1} - V_\lambda^{-1}) J_0\| = O\left(\alpha^{\min(1, \frac{\epsilon-1}{2})}\right). \quad (38)$$

where ϵ the maximum real number such that J_0 belongs to the range of V_λ^ϵ .

The lower bound of ϵ is equal to one because J_0 belongs to the range of V_λ by construction. Hence $\widehat{V}_{\lambda,\alpha}^{-1} \iota$ is consistent for $V_\lambda^{-1} J_0$ provided that ϵ is strictly larger than 1 and α is selected wisely. Also note that $(V_{\lambda,\alpha}^{-1} - V_\lambda^{-1}) f_0 = O(\alpha)$ as $\alpha \rightarrow 0$ if $\epsilon \geq 3$. By putting Proposition 6 and 7 together, we have:

$$\begin{aligned} \left\| \widehat{V}_{\lambda,\alpha}^{-1} \iota - V_\lambda^{-1} J_0 \right\| &\leq \left\| \widehat{V}_{\lambda,\alpha}^{-1} \iota - V_{\lambda,\alpha}^{-1} J_0 \right\| + \|(V_{\lambda,\alpha}^{-1} - V_\lambda^{-1}) J_0\| \\ &= O_p\left(\alpha^{-3/2} L^{-1/2}\right) + O_p\left(\alpha^{-3/2} S^{-1/2}\right) + O\left(\alpha^{\min(1, \frac{\epsilon-1}{2})}\right). \end{aligned}$$

We compute the feasible optimal ICGMM estimator as:

$$\begin{aligned} \widehat{\theta}_{\lambda,\alpha}^* &= \frac{1}{S} \sum_{i=1}^S \widehat{\pi}_{\lambda,\alpha}^*(\tau_i) \widehat{\theta}(\tau_i), \\ &= \left(\frac{1}{S} \iota' \widehat{V}_{\lambda,\alpha}^{-1} \iota \right)^{-1} \left(\frac{1}{S} \iota' \widehat{V}_{\lambda,\alpha}^{-1} \underline{\widehat{\theta}} \right). \end{aligned} \quad (39)$$

where $\widehat{\theta}(\tau_i), i = 1, \dots, S$ are computed from the actual data and $\underline{\widehat{\theta}} = \left(\widehat{\theta}(\tau_1), \dots, \widehat{\theta}(\tau_S) \right)'$. The next proposition compares $\widehat{\theta}_{\lambda,\alpha}^*$ with the optimal ICGMM estimator.

Proposition 8 Let $\underline{\widehat{\theta}} = \left(\widehat{\theta}(\tau_1), \dots, \widehat{\theta}(\tau_S) \right)'$ where τ_1, \dots, τ_S are S draws from the multivariate uniform distribution on Λ . Under Assumptions 1 to 7, we have:

$$\begin{aligned} &\frac{1}{S} \sum_{i=1}^S \widehat{\pi}_{\lambda,\alpha}^*(\tau_i) \widehat{\theta}(\tau_i) - \int \pi_\lambda^*(\tau) \widehat{\theta}_{CGMM}(\tau) d\tau \\ &= O_p\left(\alpha^{-3/2} L^{-1/2}\right) + O_p\left(\alpha^{-3/2} S^{-1/2}\right) + O\left(\alpha^{\min(1, \frac{\epsilon-1}{2})}\right), \end{aligned}$$

as L and S go to infinity.

Proposition 8 establishes that the feasible optimal ICGMM estimator converges in probability to the unfeasible one. This implies that both estimators are asymptotically equivalent.

Using the l^{th} sample, we can compute the following estimator:

$$\widehat{\theta}_{\lambda,\alpha}^{*(l)} = \frac{1}{S} \sum_{i=1}^S \widehat{\pi}_{\lambda,\alpha}^*(\tau_i) \widehat{\theta}(\tau_i, l), \quad l = 1, \dots, L. \quad (40)$$

The distribution of $\widehat{\theta}_{\lambda,\alpha}^{*(l)}$, $l = 1, \dots, L$ inferred by Monte Carlo can be used to perform inferences about θ_0 . In particular, it can be used to select the regularization parameter α , e.g., by minimizing the following approximate mean square error:

$$\widehat{\alpha}^* = \arg \min_{\alpha \in (0,1)} \frac{T}{L} \sum_{k=1}^L \left(\widehat{\theta}_{\lambda,\alpha}^{*(l)} - \widehat{\theta}_S \right)' \left(\widehat{\theta}_{\lambda,\alpha}^{*(l)} - \widehat{\theta}_S \right), \quad (41)$$

where $\widehat{\theta}_S = \frac{1}{S} \sum_{i=1}^S \widehat{\theta}(\tau_i)$.

6 Extensions of the ICGMM to CMR, Markov and Weakly Dependent Models

Let us consider the following CMR model:

$$E[g(\theta_0, y_t) | X_t] = 0, \quad t = 1, \dots, T, \quad (42)$$

where $y_t \in \mathbb{R}$, $X_t \in \mathbb{R}^d$ and $g(\theta_0, y_t) \in \mathbb{R}$. For simplicity, assume that the observations are uncorrelated across t . This CMR is equivalent to $E[g(\theta_0, y_t)A(X_t)] = 0$ for all possible functions $A(X_t)$. Thus, let $A(\tau, X_t) = \exp(i\tau'X_t)$ so that:

$$h_t(\tau, \theta) = g(\theta, y_t) \exp(i\tau'X_t), \quad \tau \in \mathbb{R}^d. \quad (43)$$

The parameter θ_0 can be efficiently estimated by CGMM from the continuum of unconditional moment restrictions (43). In particular, one may resort to the ICGMM when d is large. Note that the design (43) has also been used in Lavergne and Patilea (2008) and Antoine and Lavergne (2011).

Let us now consider a Markov process $x_t \in \mathbb{R}^d$ for which the conditional CF is known. For this case, Carrasco, Chernov, Florens and Ghysels (2007) recommend designing the moment function as:

$$h_t(\tau, \theta) = (\exp(is_0'x_t) - \varphi_t(s_0, \theta)) \exp\left(\sum_{k=1}^p is_k'x_{t-k}\right), \quad (44)$$

where p is the order of the Markov process, $\varphi_t(\tau_0, \theta) = E^\theta(\exp(is_0'x_t) | \{x_{t-k}\}_{k=1}^p)$, $s \in \mathbb{R}^d$ is the conditional CF of x_t and $\tau = (s_0, \dots, s_p) \in \mathbb{R}^{d(p+1)}$.

Finally, let us assume that $x_t \in \mathbb{R}^d$ is a weakly dependent process for which the joint CF of an arbitrary number of consecutive observations is known in closed form. For this case, Carrasco, Chernov, Florens and Ghysels (2007) propose writing $h_t(\tau, \theta)$ as:

$$h_t(\tau, \theta) = e^{i\tau'Y_t} - \varphi_t(\tau, \theta), \quad (45)$$

where $Y_t = (x_t', x_{t-1}', \dots, x_{t-p}')'$, $\tau \in \mathbb{R}^{d(p+1)}$ and $\varphi_t(\tau, \theta) = E^\theta(e^{i\tau'Y_t})$ is the CF of Y_t . Here, p must be large enough to ensure identification.

In either of the above cases, the objective function of the CGMM is of the same form as in Equation (3). However, the definitions of the scalar product $\langle \cdot, \cdot \rangle$ need to be adapted to each

situation (i.e., $\langle \cdot, \cdot \rangle$) must be defined on $\mathbf{L}^2(\Pi) \times \mathbf{L}^2(\Pi)$, where $\Pi(\tau)$ is a probability measure on $\mathbb{R}^{\dim(\tau)}$. Also, note that the general expression of the kernel function $k(\tau_1, \tau_2)$ is given by:

$$k(\tau_1, \tau_2) = E \left(h_t(\tau_1, \theta) \overline{h_t(\tau_2, \theta)} \right) + \sum_{j=1}^{\infty} E \left(h_t(\tau_1, \theta) \left(\overline{h_{t-j}(\tau_2, \theta)} + \overline{h_{t+j}(\tau_2, \theta)} \right) \right). \quad (46)$$

This expression reduces to $k(\tau_1, \tau_2) = E \left(h_t(\tau_1, \theta) \overline{h_t(\tau_2, \theta)} \right)$ when $h_t(\tau_1, \theta)$ is given by either (42) or (44). Under weak dependence, $k(\tau_1, \tau_2)$ may be estimated as in Newey and West (1987) or Andrews and Monahan (1992) using the Bartlett kernel:

$$\widehat{k}_T(\tau_1, \tau_2) = \frac{1}{T} \sum_{t=1}^T h_t(\tau_1, \widehat{\theta}^1) \overline{h_t(\tau_2, \widehat{\theta}^1)} + \sum_{j=1}^{J_T} \left(1 - \frac{j-1}{J_T} \right) \sum_{t=1}^T h_t(\tau_1, \widehat{\theta}^1) \left(\overline{h_{t-j}(\tau_2, \widehat{\theta}^1)} + \overline{h_{t+j}(\tau_2, \widehat{\theta}^1)} \right),$$

where $\widehat{\theta}^1$ is a consistent first step estimator of θ_0 and J_T is a bandwidth that is increasing in T . See Carrasco, Florens, Chernov and Ghysels (2007).

The approach presented in Section 2.1 to tackle the curse of dimensionality that arises when implementing the CGMM requires turning the full information set $\{h_t(\tau, \theta), \tau \in \mathbb{R}^d\}$ into the reduced information set $\{h_{\tau,t}(u, \theta), u \in \mathbb{R}\}$ for a given $\tau \in \Lambda$. However, it is not always possible to identify θ_0 from such a reduced information set. This is particularly true in CMR, Markov and weakly dependent models. In CMR models for instance, trying to identify θ_0 from the previous reduced information set amounts to trying to estimate θ_0 based on only one instrument. Interestingly, the second approach which consists of discretizing the full information set into $\{h_t(\tau_i, \theta), i = 1, \dots, n\}$ does not suffer from identification problems if a large enough n is chosen.

Finally, note that the method described in Section 5 for the implementation of the feasible efficient ICGMM applies here with no change as it does not depend on the data generating process.

7 Monte Carlo Simulations

This section presents two simulation studies. The first study is based on a bivariate Gaussian IID model and is aimed at presenting a non-technical description of the implementation of the ICGMM estimator. The second set of experiments is based on a CMR model that we turn into a continuum of unconditional moment conditions.

7.1 Simulation study 1

We base this simulation study on a bivariate Gaussian IID model specified as follows:

$$x_t = \begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix} \sim N \left(\begin{pmatrix} \mu \\ \mu \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right). \quad (47)$$

The vector of parameters of interest is $\theta = (\mu, \rho)$, with $\theta_0 = (0, 0.5)$ being the true value. The CF of x_t is given by:

$$\varphi(\tau, \theta) = \exp\left(i\mu_\tau - \frac{1}{2}\sigma_\tau^2\right), \tau \in \mathbb{R}^d, \quad (48)$$

where $\mu_\tau = \mu(\tau_1 + \tau_2)$ and $\sigma_\tau^2 = \tau_1^2 + 2\rho\tau_1\tau_2 + \tau_2^2$.

We choose this simple model because it accommodates at least five approaches to estimate θ_0 : (i) the standard maximum likelihood estimator (MLE), (ii) the standard CGMM estimator, (iii) the indirect maximum likelihood based on the distribution of $y_{\tau,t} = \tau'x_t$ (IMLE), (iv) the ICGMM estimator based on the distribution of $y_{\tau,t} = \tau'x_t$, and (v) the ICGMM estimator based on a discrete subset of the continuum of moment conditions given by:

$$h_t(\tau, \theta) = \exp(i\tau'x_t) - \exp\left(i\mu_\tau - \frac{1}{2}\sigma_\tau^2\right), \tau \in \mathbb{R}^2, \quad (49)$$

where $\mu_\tau = \mu(\tau_1 + \tau_2)$ and $\sigma_\tau^2 = \tau_1^2 + 2\rho\tau_1\tau_2 + \tau_2^2$. Our goal is to compare these five estimators in a context where they are all feasible.

The MLE of θ based on the joint distribution (47) (denoted $\hat{\theta}_{MLE}$) solves the following fixed-point problem:

$$\begin{aligned} \hat{\mu}_{MLE} &= \frac{1}{2T} \sum (x_{1,t} + x_{2,t}) \text{ and} \\ \hat{\rho}_{MLE} &= \frac{\hat{\rho}_{MLE}^3 - \hat{\rho}_{MLE}^2 \frac{1}{T} \sum (x_{1,t} - \hat{\mu}_{MLE})(x_{2,t} - \hat{\mu}_{MLE}) - \frac{1}{T} \sum (x_{1,t} - \hat{\mu}_{MLE})(x_{2,t} - \hat{\mu}_{MLE})}{1 - \frac{1}{T} \sum (x_{1,t} - \hat{\mu}_{MLE})^2 - \frac{1}{T} \sum (x_{2,t} - \hat{\mu}_{MLE})^2}. \end{aligned}$$

The MLE of θ based on the distribution of $y_{\tau,t} = \tau'x_t, \tau \in [0, 1]^2$ (denoted $\hat{\theta}_{MLE}(\tau)$) is given by:

$$\begin{aligned} \hat{\mu}_{MLE}(\tau) &= \frac{1}{T(\tau_1 + \tau_2)} \sum y_{\tau,t} \text{ and} \\ \hat{\rho}_{MLE}(\tau) &= \frac{1}{2\tau_1\tau_2} \left[\frac{1}{T} \sum (y_{\tau,t} - (\tau_1 + \tau_2)\hat{\mu}_{MLE}(\tau))^2 - \tau_1^2 - \tau_2^2 \right]. \end{aligned}$$

The IMLE of θ , $\hat{\theta}_{IMLE}$, is obtained by integrating $\hat{\theta}_{MLE}(\tau)$ against an estimated optimal measure on $[0, 1]^2$.

The standard CGMM estimator of θ (denoted $\hat{\theta}_{CGMM}$) is based on the continuum of moment conditions (49). The first step CGMM estimator based on the distribution of $y_{\tau,t} = \tau'x_t, \tau \in [0, 1]^2$ (denoted $\hat{\theta}_{CGMM}(\tau)$) is obtained through the following moment function:

$$h_{\tau,t}(u, \theta) = \exp(iuy_{\tau,t}) - \exp\left(iu\mu_\tau - \frac{1}{2}u^2\sigma_\tau^2\right), u \in \mathbb{R}. \quad (50)$$

where $\mu_\tau = \mu(\tau_1 + \tau_2)$ and $\sigma_\tau^2 = \tau_1^2 + 2\rho\tau_1\tau_2 + \tau_2^2$. The CGMM objective function is evaluated using 10 quadrature points. An ICGMM estimator, $\hat{\theta}_{ICGMM,1}$, is obtained by integrating $\hat{\theta}_{CGMM}(\tau)$ against an estimated optimal measure on $[0, 1]^2$.

A first step GMM estimator, $\hat{\theta}_{GMM}(\tilde{\tau})$, can be obtained based on the vector of moment conditions given by:

$$\tilde{h}_t(\tilde{\tau}, \theta) = (h_t(\tau_1, \theta), \dots, h_t(\tau_n, \theta))',$$

where $\tilde{\tau} = (\tau'_1, \dots, \tau'_n)'$ and $\tau_i \in \mathbb{R}^2, i = 1, \dots, n$. Another ICGMM estimator, $\widehat{\theta}_{ICGMM,2}$, is obtained by integrating $\widehat{\theta}_{GMM}(\tilde{\tau})$ against an estimated optimal measure on $(\mathbb{R}^2)^n$. We consider $n = 5$ and $n = 10$ in order to gauge the sensitivity of the results to the number of discretization points.

We compare the performance of five estimators in a small scale, common random numbers Monte Carlo experiment. To start, we simulate $L = 500$ samples of size $T = 250$. From each sample indexed by $l = 1, \dots, L$, we compute $\widehat{\theta}_{MLE}^{(l)}, \widehat{\theta}_{CGMM}^{(l)}$, and $\left\{ \widehat{\theta}_{MLE}^{(l)}(\tau_i), \widehat{\theta}_{CGMM}^{(l)}(\tau_i), \widehat{\theta}_{GMM}^{(l)}(\tilde{\tau}_i) \right\}_{i=1}^{S=125}$. Each τ_i is drawn independently from the bivariate uniform distribution on $[0, 1]^2$ and the same τ_i is used to compute $\widehat{\theta}_{MLE}^{(l)}(\tau_i)$ and $\widehat{\theta}_{CGMM}^{(l)}(\tau_i)$. Each array $\tilde{\tau}_i = (\tau'_{i,1}, \dots, \tau'_{i,n})'$ is built by drawing $\tau_{i,j}, j = 1, \dots, n$ independently from a diffuse bivariate normal distribution (i.e. a distribution that puts almost uniform weights on all elements of \mathbb{R}^2). We use:

$$\tilde{\tau}_{i,j} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 25 & 0 \\ 0 & 25 \end{pmatrix} \right).$$

Let $\widehat{\Theta}$ be the $L \times S$ matrix with (l, i) element equal to $\widehat{\delta}_{l,i} = \lambda' \widehat{\theta}^{(l,i)}$, where $\lambda = (1, 1)'$ and $\widehat{\theta}^{(l,i)}$ is either $\widehat{\theta}_{MLE}^{(l)}(\tau_i), \widehat{\theta}_{CGMM}^{(l)}(\tau_i)$ or $\widehat{\theta}_{GMM}^{(l)}(\tilde{\tau}_i)$. Note that the l^{th} row of $\widehat{\Theta}$ is computed from the l^{th} sample while the i^{th} column of $\widehat{\Theta}$ is labelled by either the same τ_i or the same $\tilde{\tau}_i$. Let us define:

$$\widehat{V} = \frac{T}{SL} \left(\widehat{\Theta} - \overline{\widehat{\Theta}} \right)' \left(\widehat{\Theta} - \overline{\widehat{\Theta}} \right),$$

where $\overline{\widehat{\Theta}}$ contains the means of $\widehat{\Theta}$ column-wise. The optimal aggregating measure is estimated by:

$$\widehat{\pi}_\alpha^* = S \left(l' \widehat{V}_\alpha^{-1} l \right)^{-1} l' \widehat{V}_\alpha^{-1},$$

where $\widehat{V}_\alpha^{-1} = \left(\widehat{V}^2 + \alpha I \right)^{-1} \widehat{V}$ for an arbitrary choice $\alpha = 10^{-6}$. Finally, the feasible efficient indirect estimators from the l^{th} sample are given by:

$$\begin{aligned} \widehat{\theta}_{IMLE}^{(l)} &= \frac{1}{S} \sum_{i=1}^S \widehat{\pi}_\alpha^*(\tau_i) \widehat{\theta}_{MLE}^{(l)}(\tau_i), l = 1, \dots, L, \\ \widehat{\theta}_{ICGMM,1}^{(l)} &= \frac{1}{S} \sum_{i=1}^S \widehat{\pi}_\alpha^*(\tau_i) \widehat{\theta}_{CGMM}^{(l)}(\tau_i), l = 1, \dots, L \text{ and} \\ \widehat{\theta}_{ICGMM,2}^{(l)} &= \frac{1}{S} \sum_{i=1}^S \widehat{\pi}_\alpha^*(\tau_i) \widehat{\theta}_{GMM}^{(l)}(\tilde{\tau}_i), l = 1, \dots, L. \end{aligned}$$

where $\widehat{\pi}_\alpha^*(\tau_i)$ is estimated using the relevant values of $\widehat{\delta}_{l,i}$.

The empirical distributions of the estimators are used to produce the results shown in Table 1. First, we note that the properties of the IMLE, ICGMM1 and ICGMM2 estimators do not improve as one increase the number of discretization point from 5 to 10. The results suggest that the point estimates of μ and ρ can be fairly considered unbiased. All five methods deliver 90% confidence intervals that encompass the true value of the parameters. As expected, the two estimators based on full information (i.e., MLE and CGMM) have similar biases, standard

errors and confidence intervals. Likewise, the two methods based on the reduced information set $\{y_{\tau,t} = \tau'x_t, \tau \in [0, 1]^2\}$ (i.e., the IMLE and ICGMM1) have similar performance in all respects. In particular, the standard errors of the IMLE and ICGMM1 estimators of ρ are twice as much as those of their MLE and CGMM analogues. This highlights the fact that the CGMM estimator is equivalent to the full MLE while the ICGMM1 estimator is equivalent to the reduced information MLE. The ICGMM2 estimator compares favorably to the MLE and the CGMM estimator. This suggests that the reduced information set on which the ICGMM2 is based is richer than the one pertaining to ICGMM1 for the data generating processes under consideration.

Table 1. Comparison of the MLE, CGMM, IMLE and ICGMM by Simulation.

The columns labelled "Mean", "Std. dev." and "RMSE" contain the means, standard errors and root mean square errors of the estimators, respectively. IC5 and IC95 are respectively the 5th and 95th percentile of the simulated distribution so that [IC5,IC95] is a 90% confidence interval for the true value.

T=250, L=500 and S=125.

	Mean	Std. dev.	RMSE	IC5	IC95
$\hat{\mu}_{MLE}$	0.0015	0.0571	0.0572	-0.0960	0.0940
$\hat{\rho}_{MLE}$	0.4974	0.0407	0.0408	0.4286	0.5631
$\hat{\mu}_{CGMM}$	-0.0023	0.0620	0.0620	-0.1011	0.1027
$\hat{\rho}_{CGMM}$	0.4966	0.0456	0.0457	0.4195	0.5685
n=5					
$\hat{\mu}_{IMLE}$	0.0025	0.0534	0.0535	-0.0856	0.0886
$\hat{\rho}_{IMLE}$	0.4865	0.2056	0.2060	0.1213	0.8157
$\hat{\mu}_{ICGMM,1}$	-0.0020	0.0557	0.0557	-0.0965	0.0901
$\hat{\rho}_{ICGMM,1}$	0.4873	0.2090	0.2094	0.1197	0.8576
$\hat{\mu}_{ICGMM,2}$	-0.0015	0.0526	0.0526	-0.0863	0.0848
$\hat{\rho}_{ICGMM,2}$	0.4752	0.0462	0.0525	0.4010	0.5529
n=10					
$\hat{\mu}_{IMLE}$	0.0017	0.0573	0.0573	-0.0922	0.0933
$\hat{\rho}_{IMLE}$	0.4993	0.2236	0.2236	0.1591	0.8966
$\hat{\mu}_{ICGMM,1}$	-0.0023	0.0599	0.0599	-0.0988	0.0974
$\hat{\rho}_{ICGMM,1}$	0.5159	0.2233	0.2238	0.1811	0.9133
$\hat{\mu}_{ICGMM,2}$	-0.0005	0.0567	0.0567	-0.0896	0.0960
$\hat{\rho}_{ICGMM,2}$	0.5017	0.0419	0.0420	0.4324	0.5706

7.2 Simulation study 2

The second simulation study focuses on a multivariate version of a linear model used in Cragg (1983), Newey (1993), and Kitamura, Tripathi and Ahn (2004) and Lavergne and Patilea (2008). We assume that:

$$y_t = X_t \underline{\theta}_0 + \varepsilon_t, \quad t = 1, \dots, T = 50, \quad (51)$$

where:

$$\begin{aligned} \underline{\theta}_0 &= (\theta_0, \theta_1, \theta_2, \theta_3)' = (1, 1, 1, 1)', \\ X_t &= (1, x_{1,t}, x_{2,t}, x_{3,t}), \quad \ln x_{i,t} \sim N(0, 1) \text{ for } i = 1, 2, 3, \\ \varepsilon_t | X_t &\sim N(0, \sigma_t^2) \text{ with } \sigma_t^2 = 0.2 + 0.1(X_{t1}) + 0.05(X_{t1})^2 \text{ and } \iota = (1, 1, 1, 1)'. \end{aligned}$$

and all observations are independent across t .

Assuming that the distribution of ε_t and the functional form of its variance is known, $\underline{\theta}_0$ can be efficiently estimated by feasible generalized least squares (GLS), maximum likelihood or CGMM. The feasible GLS estimator and MLE are quite easy to implement for this model. However, we choose to apply the ICGMM for illustration purposes (although the use of the ICGMM would be more justified if ε_t were specified as a stable random variable). Assuming that the variance functional form is known, the conditional CF of y_t is given by:

$$\begin{aligned}\varphi(s, \underline{\theta}_0, \beta) &\equiv E(e^{isy_t} | X_t), s \in \mathbb{R}, \\ &= \exp \left[isX_t \underline{\theta}_0 - \frac{1}{2} s^2 (\beta_0 + \beta_1 (X_t \iota) + \beta_2 (X_t \iota)^2) \right].\end{aligned}$$

The suitable moment function based on this condition CF is given by:

$$h_t^{(1)}(\tau, \underline{\theta}, \beta) = (e^{isy_t} - \varphi(s, \underline{\theta}, \beta)) \exp(ir' X_t), \quad (52)$$

where $r \in \mathbb{R}^4$ and $\tau = (s, r')' \in \mathbb{R}^5$. One advantage of the ICGMM over the feasible GLS is that the variance parameter β is jointly estimated with $\underline{\theta}_0$.

Note that Equation (51) also implies the CMR:

$$E(\varepsilon_t | X) \equiv E(y_t - X_t \underline{\theta} | X) = 0.$$

Hence, one may also consider estimating $\underline{\theta}_0$ via the continuum of unconditional moment conditions given by:

$$h_t^{(2)}(r, \underline{\theta}) = (y_t - X_t \underline{\theta}) \exp(ir' X_t), \text{ for all } r \in \mathbb{R}^4. \quad (53)$$

One possible approach to exploiting the moment function $h_t^{(2)}(r, \underline{\theta})$ is proposed by Lavergne and Patilea (2008) and it leads to the minimization of the following objective-function:

$$M_{T,b}(\underline{\theta}) = \frac{1}{2n(n-1)} \sum_{1 \leq t \neq l \leq T} (y_t - X_t \underline{\theta})(y_l - X_l \underline{\theta}) \frac{1}{b^3} \Phi \left(\frac{\tilde{X}_t - \tilde{X}_l}{b} \right), \quad (54)$$

where $\Phi(\cdot)$ is a multivariate kernel, $\tilde{X}_t = (\tilde{x}_{1,t}, \tilde{x}_{2,t}, \tilde{x}_{3,t})$ and $\tilde{x}_{i,t}$ is a standardized version of $x_{i,t}$, that is:

$$\tilde{x}_{i,t} = \frac{x_{i,t}}{\sqrt{\frac{1}{T-1} \sum_{l=1}^T (x_{i,l} - \bar{x}_i)^2}}, i = 1, 2, 3 \text{ and } t = 1, \dots, T.$$

The minimizer of $M_{T,b}(\underline{\theta})$ is called the smooth minimum distance (SMD) estimator of $\underline{\theta}_0$. Lavergne and Patilea (2008) show that this estimator is consistent and has good small sample properties. They perform a simulation study based on a version of Model (51) that contains only one regressor and find that the SMD estimator compares favorably to a feasible GLS estimator based on the knowledge of the heteroscedasticity functional form.

It might be useful to note that $h_t^{(2)}(r, \underline{\theta}) = \frac{\partial h_t^{(1)}(\tau, \underline{\theta}, \beta)}{\partial s} |_{s=0}$, which indicates that $h_t^{(2)}(r, \underline{\theta})$ contains less information than $h_t^{(1)}(\tau, \underline{\theta}, \beta)$. In particular, $h_t^{(1)}(\tau, \underline{\theta}, \beta)$ incorporates the information about the heteroscedasticity and higher order moments of $y_t - X_t \underline{\theta}$ while $h_t^{(2)}(r, \underline{\theta})$ does not. In practice, one attempts to compensate for this information loss by applying a two-step procedure. In the

first step, one obtains an estimator of the conditional variance of $h_t^{(2)}(r, \underline{\theta}_0)$ based on a preliminary consistent estimator $\tilde{\underline{\theta}}$ of $\underline{\theta}$. The true conditional variance of $h_t^{(2)}(r, \underline{\theta}_0)$ is given by:

$$\text{Var} \left(h_t^{(2)}(r, \underline{\theta}_0) | X_t \right) \equiv E \left(h_t^{(2)}(r, \underline{\theta}_0) \overline{h_t^{(2)}(r, \underline{\theta}_0)} | X_t \right) = \text{Var} (y_t - X_t \underline{\theta}_0 | X_t) = \sigma_t^2.$$

This conditional variance can be estimated nonparametrically by:

$$\hat{\sigma}_t^2 = \frac{\sum_{t=1}^t (y_t - X_t \tilde{\underline{\theta}})^2 \Phi \left(\frac{\tilde{X}_t - \tilde{X}_l}{b} \right)}{\sum_{t=1}^t \Phi \left(\frac{\tilde{X}_t - \tilde{X}_l}{b} \right)},$$

so that each $\hat{\sigma}_t^2$ is a convex combination of the squared residuals. In the second step, one estimates $\underline{\theta}_0$ based on the moment function given by:

$$h_t^{(3)}(r, \underline{\theta}) = \frac{(y_t - X_t \underline{\theta})}{\sqrt{\hat{\sigma}_t^2}} \exp(ir' X_t), \text{ for all } r \in \mathbb{R}^4, \quad (55)$$

The efficient SMD estimator of Lavergne and Patilea follows from (55) and is obtained by minimizing:

$$M_{T,b}(\underline{\theta}) = \frac{1}{2n(n-1)} \sum_{1 \leq t \neq l \leq T} \left(\frac{y_t - X_t \underline{\theta}}{\sqrt{\hat{\sigma}_t^2}} \right) \left(\frac{y_l - X_l \underline{\theta}}{\sqrt{\hat{\sigma}_l^2}} \right) \frac{1}{b^3} \Phi \left(\frac{\tilde{X}_t - \tilde{X}_l}{b} \right). \quad (56)$$

Subsequently, the same bandwidth ($b = 0.3$) is used to evaluate the objective functions (54) and (56). Also, we use the true variance σ_t^2 in Equations (55) and (56) in order to assess the full potential of the normalization of $h_t^{(2)}(r, \underline{\theta}_0)$ by its conditional variance. The simulation results appear not to be sensitive to this choice.

Another approach to exploiting the moment functions $h_t^{(2)}(\tau, \underline{\theta})$ and $h_t^{(3)}(r, \underline{\theta})$ is by using the version of the ICGMM procedure designed for CMR models (i.e., the ICGMM2). Whether using either $h_t^{(1)}(\tau, \underline{\theta}, \beta)$, $h_t^{(2)}(\tau, \underline{\theta})$ or $h_t^{(3)}(r, \underline{\theta})$, the implementation steps of the ICGMM2 are the same. As a reminder, we summarize these steps based on $h_t^{(1)}(\tau, \underline{\theta}, \beta)$. To start, one samples moment conditions from the continuum $h_t^{(1)}(\tau, \underline{\theta}, \beta)$, $\tau \in \mathbb{R}^5$. Let $h_t^{(1)}(\tilde{\tau}, \underline{\theta}, \beta)$ be the vector of moment conditions given by:

$$\tilde{h}_t^{(1)}(\tilde{\tau}, \underline{\theta}, \beta) = \left(h_t^{(1)}(\tau_1, \underline{\theta}, \beta), \dots, h_t^{(1)}(\tau_{13}, \underline{\theta}, \beta) \right)',$$

where $\tilde{\tau} = (\tau_1, \dots, \tau_{13})$ and $\tau_j \in \mathbb{R}^5, j = 1, \dots, 13$. A GMM estimator $\hat{\underline{\theta}}_{GMM}(\tilde{\tau})$ of $\underline{\theta}_0$ can be obtained by minimizing the Euclidean norm of $\frac{1}{T} \sum_{t=1}^T \tilde{h}_t^{(1)}(\tilde{\tau}, \underline{\theta}, \beta)$ with respect to $(\underline{\theta}, \beta)$. Let $\hat{\underline{\theta}}_{GMM}^{(l)}(\tilde{\tau}_i)$, $i = 1, \dots, S = 125$ and $l = 1, \dots, L = 1000$ be copies of $\hat{\underline{\theta}}_{GMM}(\tilde{\tau})$ such that: (i) $\hat{\underline{\theta}}_{GMM}^{(l)}(\tilde{\tau}_i)$, $i = 1, \dots, S$ are computed from the same sample, and (ii) $\hat{\underline{\theta}}_{GMM}^{(l)}(\tilde{\tau}_i)$, $l = 1, \dots, L$ are computed using the same $\tilde{\tau}_i = (\tau_{i,1}, \dots, \tau_{i,13})$; each $\tilde{\tau}_i$ is formed by drawing $\tau_{i,j}$ from a normal distribution with zero mean and covariance matrix $10I_5$, where I_5 is the identity matrix of \mathbb{R}^5 . Let $\hat{\Theta}$ be the $L \times S$ matrix with (l, i) element equal to $\hat{\delta}_{l,i} = \iota' \hat{\underline{\theta}}_{GMM}^{(l)}(\tilde{\tau}_i)$ and:

$$\hat{V} = \frac{T}{SL} \left(\hat{\Theta} - \bar{\Theta} \right)' \left(\hat{\Theta} - \bar{\Theta} \right),$$

where $\widehat{\Theta}$ contains the means of $\widehat{\Theta}$ column-wise. The optimal aggregating measure is estimated by $\widehat{\pi}_\alpha^* = S \left(\iota' \widehat{V}_\alpha^{-1} \iota \right)^{-1} \iota' \widehat{V}_\alpha^{-1}$, where $\widehat{V}_\alpha^{-1} = \left(\widehat{V}^2 + \alpha I \right)^{-1} \widehat{V}$ and $\alpha = 10^{-6}$ is chosen more or less arbitrarily. The feasible efficient ICGMM2 estimator from the l^{th} sample is:

$$\widehat{\theta}_{ICGMM2}^{(l)} = \frac{1}{S} \sum_{i=1}^S \widehat{\pi}_\alpha^* (\tau_i) \widehat{\theta}_{GMM}^{(l)} (\tilde{\tau}_i), l = 1, \dots, L.$$

In total, we have five estimators for θ_0 : the basic SMD estimators (based on $h_t^{(2)}(\tau, \underline{\theta})$), the efficient SMD estimator (based on $h_t^{(3)}(r, \underline{\theta})$) and the efficient ICGMM2 estimators based on $h_t^{(1)}(\tau, \underline{\theta}, \beta)$, $h_t^{(2)}(\tau, \underline{\theta})$ and $h_t^{(3)}(r, \underline{\theta})$ respectively. Table 2 summarizes the simulation results.

Table 2. Comparison of the SMD and the ICGMM2 by simulation.

The labels "Mean", "Std. dev." and "RMSE" respectively refer to the means, standard errors and root mean square errors. IC5 and IC95 are respectively the 5th and 95th percentile of the simulated distribution so that [IC5, IC95] is a 90% confidence interval for the true value. The moment function $h_t^{(2)}(\tau, \underline{\theta})$ stems from a CMR, $h_t^{(3)}(\tau, \underline{\theta})$ is a normalized version of $h_t^{(2)}(\tau, \underline{\theta})$ while $h_t^{(1)}(\tau, \underline{\theta}, \beta)$ stems from the conditional CF. $T = 50$, $L = 1000$, $S = 125$, $n = 13$ and $b = 0.3$.

		Mean	Std. dev.	RMSE	IC5	IC95
SMD based on $h_t^{(2)}(\tau, \underline{\theta})$	$\widehat{\theta}_0$	0.9791	0.5312	0.5316	0.1029	1.8339
	$\widehat{\theta}_1$	1.0162	0.2923	0.2927	0.5514	1.5077
	$\widehat{\theta}_2$	1.0033	0.2862	0.2862	0.5456	1.4599
	$\widehat{\theta}_3$	1.0041	0.2971	0.2972	0.5089	1.5064
Efficient SMD based on $h_t^{(3)}(\tau, \underline{\theta})$	$\widehat{\theta}_0$	0.9924	0.5434	0.5434	0.0936	1.8727
	$\widehat{\theta}_1$	1.0116	0.3132	0.3134	0.5055	1.5182
	$\widehat{\theta}_2$	0.9967	0.3039	0.3039	0.5008	1.4989
	$\widehat{\theta}_3$	1.0040	0.3188	0.3188	0.4968	1.5456
ICGMM2 based on $h_t^{(2)}(\tau, \underline{\theta})$	$\widehat{\theta}_0$	0.9642	0.4754	0.4768	0.1935	1.7320
	$\widehat{\theta}_1$	1.0237	0.2305	0.2317	0.6561	1.3962
	$\widehat{\theta}_2$	1.0047	0.2284	0.2285	0.6141	1.3670
	$\widehat{\theta}_3$	1.0086	0.2147	0.2149	0.6326	1.3507
ICGMM2 based on $h_t^{(3)}(\tau, \underline{\theta})$	$\widehat{\theta}_0$	0.9837	0.4065	0.4068	0.3284	1.6215
	$\widehat{\theta}_1$	1.0178	0.1964	0.1972	0.7075	1.3442
	$\widehat{\theta}_2$	1.0027	0.1986	0.1986	0.6637	1.3211
	$\widehat{\theta}_3$	1.0052	0.1941	0.1941	0.6816	1.3139
ICGMM2 based on the conditional CF $h_t^{(1)}(\tau, \underline{\theta}, \beta)$	$\widehat{\theta}_0$	0.9892	0.1452	0.1456	0.7519	1.2163
	$\widehat{\theta}_1$	1.0129	0.0954	0.0963	0.8463	1.1638
	$\widehat{\theta}_2$	1.0026	0.0942	0.0942	0.8484	1.1630
	$\widehat{\theta}_3$	0.9988	0.0968	0.0968	0.8427	1.1616

First, we note that the RMSEs of all five estimators are roughly equal to their standard deviations, which indicates the absence of bias. Second, the ICGMM2 estimators have smaller RMSEs

compared to SMD estimators. The ICGMM estimator based on $h_t^{(1)}(\tau, \underline{\theta}, \beta)$ markedly dominates all other estimators and the second best is the ICGMM estimator based on $h_t^{(3)}(r, \underline{\theta})$. This suggests that the ICGMM makes an efficient use of the information content of the CMR. However, it should be emphasized that the computational burden associated with ICGMM is higher compared to the SMD. Note that the efficient SMD estimator does not outperform the basic SMD estimator. This stems from the fact that $h_t^{(2)}(\tau, \underline{\theta})$ and $h_t^{(3)}(\tau, \underline{\theta})$ are both built from only one CMR and $h_t^{(2)}(\tau, \underline{\theta})$ is proportional to $h_t^{(3)}(\tau, \underline{\theta})$ conditional on X_t . Arguably, the two-step SMD procedure would outperform the basic SMD in the presence of several CMRs as it would permit to exploit the information content of the cross-covariances of the CMRs.

Table 3. Assessing the performance of the ICGMM2 as the number of discretization points varies. The labels "Mean", "Std. dev." and "RMSE" respectively refer to the means, standard errors and root mean square errors. IC5 and IC95 are respectively the 5th and 95th percentile of the simulated distribution so that [IC5,IC95] is a 90% confidence interval for the true value. The moment function $h_t^{(2)}(\tau, \underline{\theta})$ stems from a CMR, $h_t^{(3)}(\tau, \underline{\theta})$ is a normalized version of $h_t^{(2)}(\tau, \underline{\theta})$ while $h_t^{(1)}(\tau, \underline{\theta}, \beta)$ stems from the conditional CF. $T = 50$, $L = 1000$, $S = 125$ and n is the number of discretization points.

		Mean	Std. dev.	RMSE	IC5	IC95	
n=7	ICGMM2 based on $h_t^{(2)}(\tau, \underline{\theta})$	$\hat{\theta}_0$	1.0076	0.4418	0.4419	0.3121	1.7556
		$\hat{\theta}_1$	0.9814	0.2282	0.2290	0.6177	1.3443
		$\hat{\theta}_2$	1.0102	0.2169	0.2171	0.6591	1.3821
		$\hat{\theta}_3$	0.9993	0.2173	0.2173	0.6527	1.3831
	ICGMM2 based on $h_t^{(3)}(\tau, \underline{\theta})$	$\hat{\theta}_0$	1.0244	0.3871	0.3878	0.3603	1.6350
		$\hat{\theta}_1$	0.9834	0.2083	0.2090	0.6635	1.3108
		$\hat{\theta}_2$	0.9999	0.1918	0.1918	0.6803	1.3057
		$\hat{\theta}_3$	0.9943	0.1918	0.1919	0.6655	1.3091
	ICGMM2 based on the conditional CF $h_t^{(1)}(\tau, \underline{\theta}, \beta)$	$\hat{\theta}_0$	1.0041	0.1165	0.1166	0.8156	1.1924
		$\hat{\theta}_1$	0.9971	0.0827	0.0828	0.8546	1.1349
		$\hat{\theta}_2$	1.0036	0.0797	0.0798	0.8697	1.1378
		$\hat{\theta}_3$	1.0085	0.0770	0.0775	0.8781	1.1347
n=10	ICGMM2 based on $h_t^{(2)}(\tau, \underline{\theta})$	$\hat{\theta}_0$	0.9915	0.4513	0.4513	0.2266	1.7105
		$\hat{\theta}_1$	1.0033	0.2117	0.2117	0.6425	1.3464
		$\hat{\theta}_2$	1.0047	0.2297	0.2298	0.6316	1.3904
		$\hat{\theta}_3$	1.0023	0.2290	0.2291	0.6120	1.3804
	ICGMM2 based on $h_t^{(3)}(\tau, \underline{\theta})$	$\hat{\theta}_0$	0.9946	0.3685	0.3685	0.3857	1.5870
		$\hat{\theta}_1$	1.0057	0.1856	0.1857	0.7074	1.3151
		$\hat{\theta}_2$	1.0078	0.1982	0.1984	0.6780	1.3199
		$\hat{\theta}_3$	0.9972	0.1905	0.1906	0.6958	1.3075
	ICGMM2 based on the conditional CF $h_t^{(1)}(\tau, \underline{\theta}, \beta)$	$\hat{\theta}_0$	0.9979	0.1313	0.1314	0.7785	1.2084
		$\hat{\theta}_1$	1.0031	0.0888	0.0889	0.8582	1.1503
		$\hat{\theta}_2$	1.0026	0.0865	0.0865	0.8571	1.1374
		$\hat{\theta}_3$	1.0016	0.0865	0.0865	0.8611	1.1443

We conduct a last simulation experiment aimed at assessing the sensitivity of the performance of the ICGMM2 estimator to the choice of the number of the discretization points n . Given that $h_t^{(1)}(\tau, \underline{\theta}, \beta)$ depends on seven parameters, $n = 7$ is the minimum number of discretization points that can be used to compare the estimators based on the moment functions $h_t^{(2)}(\tau, \underline{\theta})$, $h_t^{(3)}(\tau, \underline{\theta})$ and $h_t^{(1)}(\tau, \underline{\theta}, \beta)$. Table 3 presents the simulation results for $n = 7$ and $n = 10$ (see Table 2 for the case $n = 13$). Overall, there is very little variation in the performance of the ICGMM2 estimators across the different values of n considered. If one looks at the results under the microscope, we note a slight improvement in the performance of the ICGMM2 estimators based on $h_t^{(2)}(\tau, \underline{\theta})$ and $h_t^{(3)}(\tau, \underline{\theta})$ but a slight deterioration for the one based on $h_t^{(1)}(\tau, \underline{\theta}, \beta)$ as n increases. This suggests that for this particular DGP, the ICGMM2 procedure based on the CF does not require a large n for optimality, unlike the ones based on $h_t^{(2)}(\tau, \underline{\theta})$ and $h_t^{(3)}(\tau, \underline{\theta})$ which seem to require larger values of n to achieve their full potential.

8 Conclusion

The CGMM proposed by Carrasco and Florens (2000) generalizes the GMM procedure of Hansen (1982) to a continuum of moment conditions and delivers estimators that are as efficient as MLE asymptotically. Its objective function is given by the quadratic form $\int K_{\alpha T}^{-1} h(\tau, \theta) \overline{h(\tau, \theta)} \Pi(\tau) d\tau$, where θ is a finite dimensional parameter of interest (with true value θ_0), $h(\tau, \theta)$ is the difference between a theoretical CF and its empirical counterpart, τ is the Fourier transformation index, $K_{\alpha T}^{-1}$ is a regularized inverse of the empirical covariance operator associated with $h(\tau, \theta)$ and $\Pi(\tau)$ is a continuous measure on $\mathbb{R}^{\dim(\tau)}$. However, the complexity of the numerical evaluation of this objective function grows exponentially with the dimensionality of τ , which renders the CGMM estimator roughly unfeasible in high dimensional models.

We propose two methods to tackle this “*curse of dimensionality*”. In the first approach, one converts the multivariate model of interest into a continuum of auxiliary univariate models. Under certain regularity conditions, each of the auxiliary models may be used to obtain a consistent estimator of θ_0 . An indirect CGMM (ICGMM) estimator of θ_0 is then recovered as an optimal weighted average of the previous consistent estimators. In the second approach, one draws a large number of sets of n moment conditions from the continuum $\{h(\tau, \theta), \tau \in \mathbb{R}^{\dim(\tau)}\}$. Consistent first step GMM estimators of θ_0 are obtained by minimizing the Euclidean norm of the sample average of the vector formed by each set of n moment conditions. Finally, an ICGMM estimator of θ_0 is obtained by optimally combining the previous first step estimators.

The first approach may suffer from identification issues if the dimensionality reduction leads to a severe information loss. For that reason, this approach should not be used outside IID models. The second approach has a more general scope than the first and always allows the identification of θ_0 if n is large enough. We illustrate both versions of the ICGMM with two sets of Monte Carlo experiments. The first Monte Carlo study is based on a bivariate Gaussian model. The results suggest that the second version of the ICGMM is a viable alternative to the CGMM and maximum likelihood. The second Monte Carlo study focuses on a CMR that is turned into a continuum of unconditional moment restrictions. The simulation results suggest that the ICGMM outperforms the smoothed minimum distance (SMD) estimator of Lavergne and Patilea (2008). However, the SMD procedure requires less computational burden than the ICGMM.

9 Acknowledgements

The author wish to thanks Marine Carrasco, Bruno Larue, Pierre Evariste Nguimkeu, Sandy Rose and two anonymous referees for useful comments.

References

- [1] Antoine, B., and P. Lavergne (2011), “Conditional moment models under weak identification,” *Working paper available at ideas.repec.org*.
- [2] Bierens, H. (1982): “Consistent Model Specification Tests,” *Journal of Econometrics* 20, 105–134.
- [3] Carrasco, M., Chernov, M., Florens, J. P. and E. Ghysels (2007) “Efficient estimation of general dynamic models with a continuum of moment conditions,” *Journal of Econometrics* 140, 529–573.
- [4] Carrasco, M. and J. P. Florens (2000) “Generalization of GMM to a continuum of moment conditions,” *Econometric Theory* 16, 797-834.
- [5] Carrasco, M., J. P. Florens, and E. Renault (2007) “Linear Inverse Problems in Structural Econometrics: Estimation based on spectral decomposition and regularization,” in the *Handbook of Econometrics* 6, edited by J.J. Heckman and E.E. Leamer.
- [6] Cragg, J. (1983) “More efficient estimation in the presence of heteroscedasticity of unknown form,” *Econometrica* 51 (3), 751-764.
- [7] Dominguez, M. A. and I. N. Lobato (2004) “Consistent Estimation of Models defined by Conditional Moment Restriction,” *Econometrica* 72 (5), 1601–1615.
- [8] Dufour, J.-M. (1997) “Some impossibility theorems in econometrics, with applications to structural and dynamic models,” *Econometrica* 65, 1365-1389.
- [9] Gouriéroux, C. and J. Jasiak (2005) “Autoregressive Gamma Processes,” *Journal of Forecasting* 25, 129–152.
- [10] Hansen, L. (1982) “Large Sample Properties of Generalized Method of Moments Estimators,” *Econometrica* 50, 1029-1054.
- [11] Jiang, G. and J. Knight (2002) “Estimation of Continuous Time Processes via the Empirical Characteristic Function,” *Journal of Business and Economic Statistics* 20, 198-212.
- [12] Kitamura, Y., Tripathi, G., and Ahn, H. (2004) “Empirical likelihood-based inference in conditional moment restriction models,” *Econometrica* 72 (6), 1667-1714.
- [13] Knight, J. L. and J. Yu (2002) “The empirical characteristic function in time series estimation,” *Econometric Theory* 18, 691-721.

- [14] Kotchoni, R. (2012) “Applications of the Characteristic Function Based Continuum GMM in Finance,” *Computational Statistics and Data Analysis* 56 (11), 3599–3622.
- [15] Lavergne, P., and V. Patilea (2008), “Smooth minimum distance estimation and testing with conditional estimating equations: uniform in bandwidth theory,” *Working paper available at ideas.repec.org*.
- [16] Li, J., Favero, C. and Ortu F. (2012) “A spectral estimation of tempered stable stochastic volatility models and option pricing” *Computational Statistics and Data Analysis* 56 (2012) 3645–3658
- [17] Newey, W. K. (1993) “Efficient estimation of models with conditional moment restrictions,” *Handbook of Statistics* 11, G.S. Maddala, C.R. Rao and H.D. Vinod eds, 419-454.
- [18] Singleton, K. (2001) “Estimation of Affine Pricing Models Using the Empirical Characteristic Function,” *Journal of Econometrics* 102, 111-141.
- [19] Taufer, E., Leonenko, N. and M. Bee (2011) “Characteristic function estimation of Ornstein–Uhlenbeck-based stochastic volatility models” *Computational Statistics and Data Analysis* 55 (2011) 2525–2539.
- [20] Yu, J. (2004) “Empirical Characteristic Function Estimation and Its Applications,” *Econometric Reviews* 23 (2), 93-123.

Appendix: Proofs

Proof of Proposition 1: The estimator $\widehat{\theta}_{CGMM}(\tau)$ minimizes the objective function:

$$Q_{\tau,T}(\theta) = \int \widehat{h}_{\tau,T}(u, \theta) \overline{\widehat{h}_{\tau,T}(u, \theta)} \omega(u) du,$$

Assumption 1 ensures that $Q_{\tau,T}(\theta)$ is a well-defined quadratic form, θ_0 is an interior point according to Assumption 2 while Assumption 4 states that $Q_{\tau,T}(\theta)$ is three times differentiable w.r.t θ . The first order condition (FOC) of this minimization is:

$$\int \frac{\partial \widehat{h}_{\tau,T}(u, \widehat{\theta})}{\partial \theta} \overline{\widehat{h}_{\tau,T}(u, \widehat{\theta})} \omega(u) du = 0,$$

where $\widehat{\theta} \equiv \widehat{\theta}_{CGMM}(\tau)$.

The consistency and asymptotic normality of $\widehat{\theta}$ has been established by Carrasco, et al. (2007, Proposition 3.1). This allows us to consider the first order expansion of $\widehat{h}_{\tau,T}(u, \widehat{\theta})$ about θ_0 :

$$\widehat{h}_{\tau,T}(u, \widehat{\theta}) \simeq \widehat{h}_{\tau,T}(u, \theta_0) + \frac{\partial \widehat{h}_{\tau,T}(u, \theta_0)'}{\partial \theta} \left(\widehat{\theta}_{CGMM}(\tau) - \theta_0 \right).$$

Substituting back into the FOC yields:

$$\int G_{\tau,T}(u, \widehat{\theta}) \overline{\widehat{h}_{\tau,T}(u, \theta_0)} \omega(u) du + \left\langle G_{\tau,T}(\cdot, \widehat{\theta}), G_{\tau,T}(\cdot, \theta_0)' \right\rangle \left(\widehat{\theta}_{CGMM}(\tau) - \theta_0 \right) \simeq 0,$$

where $G_{\tau,T}(u, \theta) = \frac{\partial \widehat{h}_{\tau,T}(u, \theta)}{\partial \theta}$ and

$$\left\langle G_{\tau,T} \left(\cdot, \widehat{\theta} \right), G_{\tau,T} \left(\cdot, \theta_0 \right) \right\rangle = \int G_{\tau,T}(u, \widehat{\theta}) \overline{G_{\tau,T}(u, \theta_0)'} \omega(u) du$$

is the outer product of the gradient $G_{\tau,T}(u, \theta)$ integrated against the measure $\omega(\cdot)$. According the Assumption 2, $\left\langle G_{\tau,T} \left(\cdot, \widehat{\theta} \right), G_{\tau,T} \left(\cdot, \theta_0 \right) \right\rangle$ is invertible for large enough T . Hence:

$$\widehat{\theta}_{CGMM}(\tau) - \theta_0 \simeq - \left\langle G_{\tau,T} \left(\cdot, \widehat{\theta} \right), G_{\tau,T} \left(\cdot, \theta_0 \right)' \right\rangle^{-1} \int G_{\tau,T}(u, \widehat{\theta}) \overline{\widehat{h}_{\tau,T}(u, \theta_0)} \omega(u) du.$$

According to Assumption 3, the random element $\widehat{h}_{\tau,T}(u, \theta_0) = \frac{1}{T} \sum_{t=1}^T \widehat{h}_{\tau,t}(u, \theta_0)$ satisfies:

$$\sqrt{T} \widehat{h}_{\tau,T}(u, \theta_0) \rightarrow N(0, K_\tau),$$

where K_τ is the covariance operator associated with $\widehat{h}_{\tau,t}(u, \theta_0)$. If we further let:

$$G_\tau(u, \theta_0) = P \lim_{T \rightarrow \infty} \frac{\partial \widehat{h}_{\tau,T}(u, \widehat{\theta}_{CGMM}(\tau))}{\partial \theta} \text{ and } W_\tau = \langle G_\tau(\cdot, \theta_0), G_\tau(\cdot, \theta_0) \rangle,$$

then, the asymptotic covariance matrix of $\widehat{\theta}$ is given by:

$$\begin{aligned} & \lim_{T \rightarrow \infty} Var \left(\sqrt{T} \widehat{\theta}_{CGMM}(\tau) \right) \\ &= W_\tau^{-1} \left(\int G_\tau(u, \theta_0) \left[\int k_\tau(u, v) \overline{G_\tau(v, \theta_0)'} \omega(v) dv \right] \omega(u) du \right) W_\tau^{-1}. \end{aligned}$$

where $k_\tau(u, v) = \lim_{T \rightarrow \infty} Cov \left(\sqrt{T} \widehat{h}_{\tau,t}(u, \theta_0), \sqrt{T} \widehat{h}_{\tau,t}(v, \theta_0) \right)$ is the covariance operator K_τ . Hence, according to Assumption 3. Hence:

$$\begin{aligned} \lim_{T \rightarrow \infty} Var \left(\sqrt{T} \widehat{\theta}_{CGMM}(\tau) \right) &= W_\tau^{-1} \left(\frac{1}{T} \int G_\tau(u, \theta_0) \left[\overline{K_\tau G_\tau(u, \theta_0)} \right] \omega(u) du \right) W_\tau^{-1} \\ &= W_\tau^{-1} \langle G_\tau(\cdot, \theta_0), K_\tau G_\tau(\cdot, \theta_0) \rangle W_\tau^{-1}. \end{aligned}$$

This shows that $\widehat{\theta}_{CGMM}(\tau)$ is \sqrt{T} consistent and asymptotically normal under Assumptions 1 and 4. ■

Proof of Proposition 2: According to Assumptions 4 and 5, $Q_{\tau,T}(\theta)$ is twice continuously differentiable with respect to θ and τ . Furthermore, Assumption 6(i) ensures that $\frac{\partial^2 Q_{\tau,T}}{\partial \theta \partial \theta'}$ is invertible at $\widehat{\theta}_{CGMM}(\tau)$, the unique minimizer of $Q_{\tau,T}$. Hence, by the implicit function theorem, $\widehat{\theta}_{CGMM}(\tau)$ is continuously differentiable with respect to τ and we have:

$$\frac{\partial \widehat{\theta}(\tau)}{\partial \tau} = - \left[\frac{\partial^2 Q_{\tau,T} \left(\widehat{\theta}_{CGMM}(\tau) \right)}{\partial \theta \partial \theta'} \right]^{-1} \frac{\partial^2 Q_{\tau,T} \left(\widehat{\theta}_{CGMM}(\tau) \right)}{\partial \theta \partial \tau},$$

in the neighbourhood of $\widehat{\theta}_{CGMM}(\tau)$. ■

Proof of Proposition 3: According to Assumptions 1 to 4, each CGMM estimator $\widehat{\theta}_{CGMM}(\tau)$ is consistent for θ_0 . Hence for any choice of pdf π , Assumption 7 is met and for $\tau \in \Lambda \setminus \mathcal{N}$, we have:

$$\begin{aligned}
Var\left(\sqrt{T}\lambda'\widehat{\theta}_\pi\right) &= \int \int \lambda' Cov\left(\sqrt{T}\widehat{\theta}(\tau_1), \sqrt{T}\widehat{\theta}(\tau_2)\right) \lambda \pi(\tau_1) \pi(\tau_2) d\tau_1 d\tau_2 \\
&\leq \max_\tau \left[\lambda' Cov\left(T\widehat{\theta}_{CGMM}(\tau)\right) \lambda \right] \underbrace{\int \int \pi(\tau_1) \pi(\tau_2) d\tau_1 d\tau_2}_{=1}, \\
&= \max_\tau \left[\lambda' Var\left(\widehat{\theta}_{CGMM}(\tau)\right) \lambda \right], \\
&= O_p(T^{-1}), \text{ for any } \lambda \in \mathbb{R}^q, \quad q = \dim \theta.
\end{aligned}$$

This shows the consistency of $\widehat{\theta}_\pi$. The asymptotic normality of $\widehat{\theta}_\pi$ follows from the fact that each $\widehat{\theta}_{CGMM}(\tau)$ is asymptotically normal. The covariance matrix of $\widehat{\theta}_\pi$ is given by:

$$Var\left(\sqrt{T}\widehat{\theta}_\pi\right) = \int \int Cov\left(\sqrt{T}\widehat{\theta}(\tau_1), \sqrt{T}\widehat{\theta}(\tau_2)\right) \pi(\tau_1) \pi(\tau_2) d\tau_1 d\tau_2.$$

Hence, the asymptotic covariance is:

$$\begin{aligned}
\Omega_\pi &\equiv \lim_{T \rightarrow \infty} Var\left(\sqrt{T}\widehat{\theta}_\pi\right) \\
&= \int \int \left[\lim_{T \rightarrow \infty} Cov\left(\sqrt{T}\widehat{\theta}(\tau_1), \sqrt{T}\widehat{\theta}(\tau_2)\right) \right] \pi(\tau_1) \pi(\tau_2) d\tau_1 d\tau_2.
\end{aligned}$$

Note that:

$$\sqrt{T}\left(\widehat{\theta}(\tau_i) - \theta_0\right) \simeq -W_{\tau_i}^{-1} \int G_{\tau_i, T}(u, \widehat{\theta}) \overline{\sqrt{T}\widehat{h}_{\tau_i, T}(u, \theta_0)} \omega(u) du, \text{ for } i = 1, 2.$$

Hence:

$$\begin{aligned}
&\lim_{T \rightarrow \infty} Cov\left(\sqrt{T}\widehat{\theta}(\tau_1), \sqrt{T}\widehat{\theta}(\tau_2)\right) \\
&= W_{\tau_1}^{-1} \left(\int G_{\tau_1}(u, \theta_0) \left[\int k_{\tau_1, \tau_2}(u, v) \overline{G_{\tau_2}(v, \theta_0)} \omega(v) dv \right] \omega(u) du \right) W_{\tau_2}^{-1},
\end{aligned}$$

where $k_{\tau_1, \tau_2}(u, v) = \lim_{T \rightarrow \infty} Cov\left(\overline{\sqrt{T}\widehat{h}_{\tau_1, t}(u, \widehat{\theta})}, \overline{\sqrt{T}\widehat{h}_{\tau_2, t}(v, \widehat{\theta})}\right)$ is the kernel of the cross-covariance operator K_{τ_1, τ_2} . Thus:

$$\begin{aligned}
&\lim_{T \rightarrow \infty} Cov\left(\sqrt{T}\widehat{\theta}(\tau_1), \sqrt{T}\widehat{\theta}(\tau_2)\right) \\
&= W_{\tau_1}^{-1} \left(\frac{1}{T} \int G_{\tau_1}(u, \theta_0) \left[\overline{K_{\tau_1, \tau_2} G_{\tau_2}(u, \theta_0)} \right] \omega(u) du \right) W_{\tau_2}^{-1}, \\
&= W_{\tau_1}^{-1} \langle G_{\tau_1}(\cdot, \theta_0), K_{\tau_1, \tau_2} G_{\tau_2}(\cdot, \theta_0) \rangle W_{\tau_2}^{-1}.
\end{aligned}$$

Finally:

$$\Omega_\pi = \int \int \left[W_{\tau_1}^{-1} \langle G_{\tau_1}(\cdot, \theta_0), K_{\tau_1, \tau_2} G_{\tau_2}(\cdot, \theta_0) \rangle W_{\tau_2}^{-1} \right] \pi(\tau_1) \pi(\tau_2) d\tau_1 d\tau_2.$$

■

Proposition 4: The ideal measure $\pi_\lambda^*(\cdot)$ solves:

$$\pi_\lambda^* = \arg \min_{\pi} \int \int g_\lambda(\tau_1, \tau_2) \pi(\tau_1) \pi(\tau_2) d\tau_1 d\tau_2,$$

subject to:

$$\int \pi(\tau_1) d\tau_1 = 1,$$

where $g_\lambda(\tau_1, \tau_2) = \lambda' Cov(\widehat{\theta}_{CGMM}(\tau), \widehat{\theta}_{CGMM}(\tau)) \lambda$. The Lagrangian for this problem is given by:

$$\mathcal{L}(\pi) = \int \int g_\lambda(\tau_1, \tau_2) \pi(\tau_1) \pi(\tau_2) d\tau_1 d\tau_2 + \mu_\lambda \left(1 - \int \pi(\tau_1) d\tau_1 \right),$$

where μ_λ is a Lagrange multiplier. The first order condition for this problem is obtained by differentiating $\mathcal{L}(\pi)$ with respect to $\pi(\tau_1)$:

$$\int g_\lambda(\tau, \tau_1) \pi(\tau_1) d\tau_1 = \mu_\lambda J(\tau),$$

where $J(\tau) = 1$ for all τ in $\Lambda \setminus \mathfrak{N}$.

Let V_λ be the linear operator with kernel $\nu_\lambda(\tau_1, \tau_2)$. The first order condition becomes:

$$V_\lambda \pi(\tau) = \mu_\lambda J(\tau). \quad (57)$$

An exact solution to this problem exists only if the constant function $J(\tau)$ lies entirely in the range of V_λ . Otherwise, an approximate solution can be constructed based on the projection of $J(\tau)$ onto the range of V_λ . As compact a covariance operator, V_λ has a discrete nonnegative spectrum with orthogonal eigenfunctions. Let $\phi_{\lambda,j}(\tau)$ be the eigenfunction of V_λ associated with the eigenvalue $\nu_{\lambda,j}$. Then any function $f(\tau)$, can be decomposed as:

$$f(\tau) = \sum_{j=1}^{\infty} \left(\int \phi_{\lambda,j}(\tau_1) f(\tau_1) d\tau \right) \phi_{\lambda,j}(\tau) + \widetilde{f}(\tau),$$

where \widetilde{f} is a residual such that $V_\lambda \widetilde{f}(\tau) = 0$. Hence, we have $J(\tau) = J_0(\tau) + \widetilde{J}(\tau)$, where

$$J_0(\tau) = \sum_{j=1}^{\infty} \left(\int \phi_{\lambda,j}(\tau_1) d\tau_1 \right) \phi_{\lambda,j}(\tau), \text{ and } \widetilde{J}(\tau) = 1 - J_0(\tau).$$

The approximate solution of (57) with minimal norm is given by:

$$\begin{aligned} \pi_\lambda^*(\tau) &= \mu_\lambda V_\lambda^{-1} J_0(\tau), \\ &= \mu_\lambda \sum_{j=1}^{\infty} \left(\int \phi_{\lambda,j}(\tau_1) d\tau_1 \right) V_\lambda^{-1} \phi_{\lambda,j}(\tau), \\ &= \mu_\lambda \sum_{j=1}^{\infty} \left(\int \phi_{\lambda,j}(\tau_1) d\tau_1 \right) \frac{1}{\nu_{\lambda,j}} \phi_{\lambda,j}(\tau). \end{aligned}$$

The Lagrange multiplier is identified using the constraint $\int \pi(\tau) d\tau = 1$. This yields:

$$\mu_\lambda^* = \left[\sum_{j=1}^{\infty} \frac{1}{\nu_{\lambda,j}} \left(\int \phi_{\lambda,j}(\tau) d\tau \right)^2 \right]^{-1} = \left(\int V_\lambda^{-1} J_0(\tau) d\tau \right)^{-1}.$$

We substitute this for μ_λ in $\pi_\lambda^*(\tau)$ to obtain:

$$\begin{aligned} \pi_\lambda^*(\tau) &= \left[\sum_{j=1}^{\infty} \frac{1}{\nu_{\lambda,j}} \left(\int \phi_{\lambda,j}(\tau_1) d\tau_1 \right)^2 \right]^{-1} \sum_{j=1}^{\infty} \frac{1}{\nu_{\lambda,j}} \left(\int \phi_{\lambda,j}(\tau_1) d\tau_1 \right) \phi_{\lambda,j}(\tau). \\ &= \mu_\lambda^* V_\lambda^{-1} J_0(\tau). \end{aligned}$$

At the solution $\pi_\lambda^*(\tau)$, we have:

$$\begin{aligned} \text{Var} \left(\lambda' \widehat{\theta}_{\pi_\lambda^*} \right) &= \int \left[\int g_\lambda(\tau_1, \tau_2) \pi_\lambda^*(\tau_1) d\tau_1 \right] \pi_\lambda^*(\tau_2) d\tau_2 = \int [V_\lambda \pi_\lambda^*(\tau_2)] [\pi_\lambda^*(\tau_2)] d\tau_2, \\ &= \mu_\lambda^{*2} \int [V_\lambda V_\lambda^{-1} J_0(\tau_2)] [V_\lambda^{-1} J_0(\tau_2)] d\tau_2 = \mu_\lambda^{*2} \int [J_0(\tau_2)] [V_\lambda^{-1} J_0(\tau_2)] d\tau_2, \\ &= \mu_\lambda^{*2} \int \left[\sum_{j=1}^{\infty} \left(\int \phi_{\lambda,j}(\tau_1) d\tau_1 \right) \phi_{\lambda,j}(\tau_2) \right] \left[\sum_{j=1}^{\infty} \frac{1}{\nu_{\lambda,j}} \left(\int \phi_{\lambda,j}(\tau_1) d\tau_1 \right) \phi_{\lambda,j}(\tau_2) \right] d\tau_2, \\ &= \mu_\lambda^{*2} \sum_{j=1}^{\infty} \frac{1}{\nu_{\lambda,j}} \left(\int \phi_{\lambda,j}(\tau_1) d\tau_1 \right)^2 \underbrace{\int \phi_{\lambda,j}^2(\tau_2) d\tau_2}_{\equiv 1} = \mu_\lambda^{*2} \mu_\lambda^{*-1} = \mu_\lambda^*. \end{aligned}$$

■

Proof of Proposition 5: We recall that:

$$\widehat{V}_\lambda = \frac{T}{SL} \left(\widehat{\Theta}_\lambda - \bar{\Theta}_\lambda \right)' \left(\widehat{\Theta}_\lambda - \bar{\Theta}_\lambda \right),$$

with the (i, j) element of \widehat{V}_λ given by:

$$\widehat{g}_\lambda(\tau_i, \tau_j) = \frac{T}{SL} \sum_{l=1}^L \lambda' \left(\widehat{\theta}(\tau_i, l) - \bar{\theta}(\tau_i, l) \right) \left(\widehat{\theta}(\tau_j, l) - \bar{\theta}(\tau_j, l) \right)' \lambda,$$

and $\underline{f} = (f(\tau_1), \dots, f(\tau_S))'$, where $\widehat{\theta}(\tau_i, l)$ is IID across l . We have:

$$\begin{aligned} \left(\widehat{V}_\lambda \underline{f} \right)_i &= \sum_{j=1}^S \widehat{g}_\lambda(\tau_i, \tau_j) f(\tau_j), \\ &= \sum_{j=1}^S \left(\frac{T}{SL} \sum_{l=1}^L \lambda' \left(\widehat{\theta}(\tau_i, l) - \bar{\theta}(\tau_i, l) \right) \left(\widehat{\theta}(\tau_j, l) - \bar{\theta}(\tau_j, l) \right)' \lambda \right) f(\tau_j), \\ &= \frac{T}{L} \sum_{l=1}^L \lambda' \left(\widehat{\theta}(\tau_i, l) - \bar{\theta}(\tau_i, l) \right) \left(\frac{1}{S} \sum_{j=1}^S \left(\widehat{\theta}(\tau_j, l) - \bar{\theta}(\tau_j, l) \right)' f(\tau_j) \right) \lambda, \end{aligned}$$

Note that $\tau_i, i = 1, \dots, S$ are drawn from the multivariate uniform distribution of Λ . By the Law of Large Numbers, as S goes to infinity, we have:

$$\frac{1}{S} \sum_{j=1}^S \left(\widehat{\theta}(\tau_j, l) - \bar{\theta}(\tau_j, l) \right)' f(\tau_j) \rightarrow \int_{\Lambda} \left(\widehat{\theta}(\tau, l) - \bar{\theta}(\tau, l) \right)' f(\tau) d\tau.$$

Hence:

$$\left(\widehat{V}_{\lambda} \underline{f} \right)_i \rightarrow \frac{T}{L} \sum_{l=1}^L \lambda' \left(\widehat{\theta}(\tau_i, l) - \bar{\theta}(\tau_i, l) \right) \left(\int_{\Lambda} \left(\widehat{\theta}(\tau, l) - \bar{\theta}(\tau, l) \right)' f(\tau) d\tau \right) \lambda,$$

and we have:

$$\begin{aligned} & \left(\widehat{V}_{\lambda} \underline{f} \right)_i - \int_{\Lambda} \left(\frac{T}{L} \sum_{l=1}^L \lambda' \left(\widehat{\theta}(\tau_i, l) - \bar{\theta}(\tau_i, l) \right) \left(\widehat{\theta}(\tau, l) - \bar{\theta}(\tau, l) \right)' \lambda \right) f(\tau) d\tau \\ &= O_p(S^{-1/2}). \end{aligned}$$

Next, $\widehat{\theta}(\tau_i, l)$ is IID across $l = 1, \dots, L$. Hence, a Law of Large Numbers applies as L goes to infinity:

$$\begin{aligned} & \int_{\Lambda} \left(\frac{T}{L} \sum_{l=1}^L \lambda' \left(\widehat{\theta}(\tau_i, l) - \bar{\theta}(\tau_i, l) \right) \left(\widehat{\theta}(\tau, l) - \bar{\theta}(\tau, l) \right)' \lambda \right) f(\tau) d\tau \\ & - \int_{\Lambda} \lambda' Cov \left(\sqrt{T} \widehat{\theta}(\tau_i, l), \sqrt{T} \widehat{\theta}(\tau, l) \right) \lambda f(\tau) d\tau = O_p(L^{-1/2}). \end{aligned}$$

This shows that

$$\left(\widehat{V}_{\lambda} \underline{f} \right)_i - \int_{\Lambda} \lambda' Cov \left(\sqrt{T} \widehat{\theta}(\tau_i, l), \sqrt{T} \widehat{\theta}(\tau, l) \right) \lambda f(\tau) d\tau = O_p(S^{-1/2}) + O_p(L^{-1/2}).$$

Because of the linearity of the operator and the bilinearity of the covariance, this convergence result is not sequential. This means that S can go to infinity first and L second or vice versa. ■

Proof of Proposition 6: Let $f = f_0 + \tilde{f}$, where f_0 is the projection of f onto the range of V_{λ} . Then, $V_{\lambda} f = V_{\lambda} f_0$ and we have:

$$\begin{aligned} & \left\| \widehat{V}_{\lambda, \alpha}^{-1} \underline{f} - V_{\lambda, \alpha}^{-1} f \right\| \\ &= \left\| (\widehat{V}_{\lambda}^2 + \alpha I)^{-1} \widehat{V}_{\lambda} \underline{f} - (V_{\lambda}^2 + \alpha I)^{-1} V_{\lambda} f \right\|, \\ &= \left\| (\widehat{V}_{\lambda}^2 + \alpha I)^{-1} \widehat{V}_{\lambda} \underline{f} - (\widehat{V}_{\lambda}^2 + \alpha I)^{-1} V_{\lambda} f + (\widehat{V}_{\lambda}^2 + \alpha I)^{-1} V_{\lambda} f - (V_{\lambda}^2 + \alpha I)^{-1} V_{\lambda} f \right\|, \\ &\leq \left\| (\widehat{V}_{\lambda}^2 + \alpha I)^{-1} \left(\widehat{V}_{\lambda} \underline{f} - V_{\lambda} f \right) \right\| + \left\| \left((\widehat{V}_{\lambda}^2 + \alpha I)^{-1} - (V_{\lambda}^2 + \alpha I)^{-1} \right) V_{\lambda} f \right\|. \end{aligned}$$

The first term satisfies:

$$\begin{aligned} \left\| (\widehat{V}_{\lambda}^2 + \alpha I)^{-1} \left(\widehat{V}_{\lambda} \underline{f} - V_{\lambda} f \right) \right\| &= \underbrace{\left\| (\widehat{V}_{\lambda}^2 + \alpha I)^{-1} \right\|}_{\leq \alpha^{-1}} \underbrace{\left\| \widehat{V}_{\lambda} \underline{f} - V_{\lambda} f \right\|}_{= O_p(L^{-1/2}) + O_p(S^{-1/2})}, \\ &= O_p(\alpha^{-1} L^{-1/2}) + O_p(\alpha^{-1} S^{-1/2}). \end{aligned}$$

For the second term, we have:

$$\left\| \left((\widehat{V}_\lambda^2 + \alpha I)^{-1} - (V_\lambda^2 + \alpha I)^{-1} \right) V_\lambda f \right\| = \left\| (\widehat{V}_\lambda^2 + \alpha I)^{-1} (V_\lambda^2 - \widehat{V}_\lambda^2) (V_\lambda^2 + \alpha I)^{-1} V_\lambda f \right\|,$$

where we used the fact that $B^{-1} - A^{-1} = B^{-1}(A - B)A^{-1}$. Hence:

$$\begin{aligned} & \left\| \left((\widehat{V}_\lambda^2 + \alpha I)^{-1} - (V_\lambda^2 + \alpha I)^{-1} \right) V_\lambda f \right\| \\ &= \underbrace{\left\| (\widehat{V}_\lambda^2 + \alpha I)^{-1} \right\|}_{\leq \alpha^{-1}} \underbrace{\left\| (V_\lambda^2 - \widehat{V}_\lambda^2) \right\|}_{=O_p(L^{-1/2}) + O_p(S^{-1/2})} \underbrace{\left\| (V_\lambda^2 + \alpha I)^{-1/2} \right\|}_{\leq \alpha^{-1/2}} \underbrace{\left\| (V_\lambda^2 + \alpha I)^{-1/2} V_\lambda f \right\|}_{\rightarrow 1}, \\ &= O_p(\alpha^{-3/2} L^{-1/2}) + O_p(\alpha^{-3/2} S^{-1/2}) \end{aligned}$$

The second term dominates the first one. Hence the result.

Proof of Proposition 7: Let $J(\tau) = 1$ for all τ and J_0 be the projection of J onto the range of V_λ . Assume that J_0 belongs to the range of V_λ^ϵ , for some $\epsilon > 1$. We have:

$$\begin{aligned} (V_{\lambda,\alpha}^{-1} - V_\lambda^{-1}) J &\equiv (V_\lambda^2 + \alpha I)^{-1} V_\lambda J - V_\lambda^{-1} J, \\ &= (V_\lambda^2 + \alpha I)^{-1} V_\lambda J_0 - V_\lambda^{-1} J_0, \\ &= \sum_{j=1}^{\infty} \left[\frac{\nu_j}{\alpha + \nu_j^2} - \frac{1}{\nu_j} \right] \langle J_0, \phi_j \rangle \phi_j, \\ &= \sum_{j=1}^{\infty} \left(\frac{\nu_j^2}{\alpha + \nu_j^2} - 1 \right) \frac{\langle J_0, \phi_j \rangle}{\nu_j} \phi_j. \end{aligned}$$

Next, we take the norm:

$$\begin{aligned} \|(V_{\lambda,\alpha}^{-1} - V_\lambda^{-1}) J\| &= \left(\sum_{j=1}^{\infty} \left(\frac{\nu_j^2}{\alpha + \nu_j^2} - 1 \right)^2 \frac{|\langle J_0, \phi_j \rangle|^2}{\nu_j^2} \right)^{1/2}, \\ &= \left(\sum_{j=1}^{\infty} \nu_j^{2\epsilon-2} \left(\frac{\nu_j^2}{\alpha + \nu_j^2} - 1 \right)^2 \frac{|\langle J_0, \phi_j \rangle|^2}{\nu_j^{2\epsilon}} \right)^{1/2}, \\ &\leq \left(\sum_{j=1}^{\infty} \frac{|\langle J_0, \phi_j \rangle|^2}{\nu_j^{2\epsilon}} \right)^{1/2} \sup_{1 \leq j \leq \infty} \nu_j^{\epsilon-1} \left| \frac{\nu_j^2}{\alpha + \nu_j^2} - 1 \right|. \end{aligned}$$

Recall that as a compact operator, the largest eigenvalue ν_1 of V_λ is bounded. We need to find an equivalent to:

$$\sup_{0 \leq \nu \leq \nu_1} \nu^{\epsilon-1} \left(1 - \frac{\nu^2}{\alpha + \nu^2} \right) = \sup_{0 \leq \lambda \leq \nu_1^2} \lambda^{\frac{\epsilon-1}{2}} \left(1 - \frac{1}{\alpha/\lambda + 1} \right). \quad (58)$$

Case where $\epsilon \leq 3/2$: We apply another change of variables $x = \alpha/\lambda$ to (58):

$$\sup_{x \geq 0} \frac{\alpha^{\epsilon/2-1/2}}{x^{\epsilon/2-1/2}} \left(1 - \frac{1}{x+1} \right).$$

An equivalent to (58) is given by $\alpha^{\epsilon/2-1/2}$, provided that $\sup_{x \geq 0} \frac{1}{x^{\epsilon/2-1/2}} \left(1 - \frac{1}{x+1}\right)$ is bounded. We study the properties of

$$g(x) \equiv \frac{1}{x^{\epsilon/2-1/2}} \left(1 - \frac{1}{x+1}\right).$$

Note that $g(x)$ is continuous and therefore bounded on any interval of $(0, +\infty)$. It goes to 0 at $+\infty$ (for any $\epsilon > 1$). For the limit at 0, we apply l'Hopital's rule and obtain

$$g(x) \underset{0}{\sim} \frac{1}{\left(\frac{\epsilon-1}{2}\right) x^{\frac{\epsilon-3}{2}}} = 0,$$

provided $\epsilon < 3/2$. For $\epsilon = 3/2$, we have

$$g(x) \underset{0}{\sim} \frac{1}{\left(\frac{\epsilon-1}{2}\right)}.$$

Hence $g(x)$ is bounded on \mathbb{R}^+ for all $\epsilon \leq 3/2$.

Case where $\epsilon > 3/2$: We rewrite (58) as

$$\sup_{0 \leq \lambda \leq \nu_1^2} \alpha \lambda^{\frac{\epsilon-3}{2}} \frac{\left(1 - \frac{1}{\alpha/\lambda+1}\right)}{\alpha/\lambda}.$$

The term $\lambda^{\frac{\epsilon-3}{2}} \frac{\left(1 - \frac{1}{\alpha/\lambda+1}\right)}{\alpha/\lambda}$ is the product of an increasing function of λ , namely $\lambda^{\frac{\epsilon-3}{2}}$ (which is bounded because λ is bounded) and a function of the form $\left(1 - \frac{1}{x+1}\right)/x$. It is easy to show using the l'Hopital's rule that $\left(1 - \frac{1}{x+1}\right)/x$ is bounded on \mathbb{R}^+ :

$$\left(1 - \frac{1}{x+1}\right)/x \underset{+\infty}{\sim} 0 \text{ and } \left(1 - \frac{1}{x+1}\right)/x \underset{0}{\sim} 1.$$

Hence the rate of (58) is given by α . By putting everything together, we have:

$$\|(V_{\lambda, \alpha}^{-1} - V_{\lambda}^{-1}) J\| = O\left(\alpha^{\min(1, \frac{\epsilon-1}{2})}\right).$$

■

Proof of Proposition 8: Let $\widehat{\underline{\theta}} = \left(\widehat{\theta}(\tau_1), \dots, \widehat{\theta}(\tau_S)\right)'$ and τ_1, \dots, τ_S be S draws from the multivariate uniform distribution on Λ . We have:

$$\frac{1}{S} \sum_{i=1}^S \widehat{\pi}_{\lambda, \alpha}^*(\tau_i) \widehat{\theta}(\tau_i) = \left(\frac{1}{S} \ell' \widehat{V}_{\lambda, \alpha}^{-1} \ell\right)^{-1} \left(\frac{1}{S} \ell' \widehat{V}_{\lambda, \alpha}^{-1} \widehat{\underline{\theta}}\right).$$

According to Propositions 6 and 7, we have:

$$\left(\widehat{V}_{\lambda, \alpha}^{-1} \widehat{\underline{\theta}}\right)_i - V_{\lambda}^{-1} \widehat{\theta}(\tau_i) = O_p\left(\alpha^{-3/2} L^{-1/2}\right) + O_p\left(\alpha^{-3/2} S^{-1/2}\right) + O\left(\alpha^{\min(1, \frac{\epsilon-1}{2})}\right),$$

as S and L go to infinity. Hence, we have:

$$\begin{aligned} \iota' \widehat{V}_{\lambda, \alpha}^{-1} \widehat{\theta} - \int V_{\lambda}^{-1} \widehat{\theta}_{CGMM}(\tau) d\tau &= \frac{1}{S} \sum_{i=1}^S \left[\left(\widehat{V}_{\lambda, \alpha}^{-1} \widehat{\theta} \right)_i - V_{\lambda}^{-1} \widehat{\theta}(\tau_i) \right] \\ &\quad + \frac{1}{S} \sum_{i=1}^S V_{\lambda}^{-1} \widehat{\theta}(\tau_i) - \int V_{\lambda}^{-1} \widehat{\theta}_{CGMM}(\tau) d\tau \end{aligned}$$

with, $\frac{1}{S} \sum_{i=1}^S \left[\left(\widehat{V}_{\lambda, \alpha}^{-1} \widehat{\theta} \right)_i - V_{\lambda}^{-1} f(\tau_i) \right] = O_p(\alpha^{-3/2} L^{-1/2}) + O_p(\alpha^{-3/2} S^{-1/2}) + O(\alpha^{\min(1, \frac{\epsilon-1}{2})})$ and $\frac{1}{S} \sum_{i=1}^S V_{\lambda}^{-1} \widehat{\theta}(\tau_i) - \int V_{\lambda}^{-1} \widehat{\theta}_{CGMM}(\tau) d\tau = O(S^{-1/2})$. The later convergence rate stems from the Law of Large Numbers applied to a uniform distribution. In total, we have:

$$\iota' \widehat{V}_{\lambda, \alpha}^{-1} \widehat{\theta} - \int V_{\lambda}^{-1} f(\tau) d\tau = O_p(\alpha^{-3/2} L^{-1/2}) + O_p(\alpha^{-3/2} S^{-1/2}) + O(\alpha^{\min(1, \frac{\epsilon-1}{2})}).$$

Similarly, we have:

$$\begin{aligned} \iota' \widehat{V}_{\lambda, \alpha}^{-1} \iota - \int V_{\lambda}^{-1} J_0(\tau) d\tau &= \frac{1}{S} \sum_{i=1}^S \left[\left(\widehat{V}_{\lambda, \alpha}^{-1} \iota \right)_i - V_{\lambda}^{-1} J_0(\tau_i) \right] \\ &\quad + \frac{1}{S} \sum_{i=1}^S V_{\lambda}^{-1} J_0(\tau_i) - \underbrace{\int V_{\lambda}^{-1} J_0(\tau) d\tau}_{S^{-1/2}}, \\ &= O_p(\alpha^{-3/2} L^{-1/2}) + O_p(\alpha^{-3/2} S^{-1/2}) + O(\alpha^{\min(1, \frac{\epsilon-1}{2})}) \end{aligned}$$

Putting everything together, we have:

$$\begin{aligned} &\frac{1}{S} \sum_{i=1}^S \widehat{\pi}_{\lambda, \alpha}^*(\tau_i) \widehat{\theta}(\tau_i) - \int \pi_{\lambda}^*(\tau) \widehat{\theta}_{CGMM}(\tau) d\tau \\ &= \frac{1}{S} \sum_{i=1}^S \widehat{\pi}_{\lambda, \alpha}^*(\tau_i) \widehat{\theta}(\tau_i) - \left(\int V_{\lambda}^{-1} J_0(\tau) d\tau \right)^{-1} \int V_{\lambda}^{-1} \widehat{\theta}_{CGMM}(\tau) d\tau, \\ &= \left(\frac{1}{S} \iota' \widehat{V}_{\lambda, \alpha}^{-1} \iota \right)^{-1} \left(\frac{1}{S} \iota' \widehat{V}_{\lambda, \alpha}^{-1} \widehat{\theta} \right) - \left(\int V_{\lambda}^{-1} J_0(\tau) d\tau \right)^{-1} \int V_{\lambda}^{-1} \widehat{\theta}_{CGMM}(\tau) d\tau, \\ &= \left(\int V_{\lambda}^{-1} J_0(\tau) d\tau \right)^{-1} \left[\left(\frac{1}{S} \iota' \widehat{V}_{\lambda, \alpha}^{-1} \widehat{\theta} \right) - \int V_{\lambda}^{-1} \widehat{\theta}_{CGMM}(\tau) d\tau \right] \\ &\quad + \left[\left(\frac{1}{S} \iota' \widehat{V}_{\lambda, \alpha}^{-1} \iota \right)^{-1} - \left(\int V_{\lambda}^{-1} J_0(\tau) d\tau \right)^{-1} \right] \int V_{\lambda}^{-1} \widehat{\theta}_{CGMM}(\tau) d\tau \\ &\quad + \left[\left(\frac{1}{S} \iota' \widehat{V}_{\lambda, \alpha}^{-1} \iota \right)^{-1} - \left(\int V_{\lambda}^{-1} J_0(\tau) d\tau \right)^{-1} \right] \left[\left(\frac{1}{S} \iota' \widehat{V}_{\lambda, \alpha}^{-1} \widehat{\theta} \right) - \int V_{\lambda}^{-1} \widehat{\theta}_{CGMM}(\tau) d\tau \right]. \end{aligned}$$

The rate of the first and second term is $O_p(\alpha^{-3/2} L^{-1/2}) + O_p(\alpha^{-3/2} S^{-1/2}) + O(\alpha^{\min(1, \frac{\epsilon-1}{2})})$ while the third term is of higher order. ■