

ON THE ASYMPTOTIC EFFICIENCY OF GMM

MARINE CARRASCO
Université de Montréal
CIREQ
and
CIRANO

JEAN-PIERRE FLORENS
Toulouse School of Economics

The efficiency of the generalized method of moment (GMM) estimator is addressed by using a characterization of its variance as an inner product in a reproducing kernel Hilbert space. We show that the GMM estimator is asymptotically as efficient as the maximum likelihood estimator if and only if the true score belongs to the closure of the linear space spanned by the moment conditions. This result generalizes former ones to autocorrelated moments and possibly infinite number of moment restrictions. Second, we derive the semiparametric efficiency bound when the observations are known to be Markov and satisfy a conditional moment restriction. We show that it coincides with the asymptotic variance of the optimal GMM estimator, thus extending results by Chamberlain (1987, *Journal of Econometrics* 34, 305–33) to a dynamic setting. Moreover, this bound is attainable using a continuum of moment conditions.

1. INTRODUCTION

The method of moments was first introduced by Pearson (1894) but was partly abandoned after Fisher (1922) demonstrated that it was inefficient relative to maximum likelihood. Later, the method of moments, which was extended into both the estimating function approach (Godambe, 1960; Durbin, 1960) and the generalized method of moments (GMM) (Hansen, 1982), became popular again because it does not require a complete specification of the model. Godambe (1960) demonstrates that the method of moments estimator based on the score is as efficient as the maximum likelihood estimator (MLE). However, using the score is not the only way for GMM estimators to reach efficiency.

The authors thank the co-editor Yuichi Kitamura and two anonymous referees for helpful suggestions. We also thank Debopam Bhattacharya and the participants of the IEE conference in honor of Arnold Zellner (Washington, DC, 2003), the Winter Meeting of the Econometric Society (San Diego, 2004), and the CIRANO and CIREQ Conference on Operator Methods (Montreal, 2004). Carrasco gratefully acknowledges partial financial support from the National Science Foundation and SSHRC. Address correspondence to Marine Carrasco, University of Montreal, Departement de Sciences Economiques, CP 6128, succ Centre Ville, Montreal, QC H3C3J7, Canada; e-mail: marine.carrasco@umontreal.ca.

This paper aims to shed new light on the efficiency of GMM estimators. To discuss efficiency, it is important to be able to handle an infinite in addition to a finite number of moment conditions. Indeed, it will become apparent later on that often the only way to reach efficiency is to use an infinite number of moment conditions. We start by giving a general framework for GMM estimation that applies indifferently to a finite and infinite number (possibly a continuum) of moment conditions. One crucial assumption is that the moment functions belong to a Hilbert space and their covariance is not invertible. The case with invertible covariance (if not finite) is not covered here. In this framework, the asymptotic variance of the GMM estimator takes the form of an inner product in a reproducing kernel Hilbert space (RKHS) associated with the covariance operator, as shown in Carrasco and Florens (2000) and Carrasco, Chernov, Florens and Renault (2007). Using results by Parzen (1959, 1970), we show that the asymptotic variance can be rewritten as an inner product in a dual space of the RKHS that is easier to handle. This provides manageable formulas to compute the asymptotic variance even in challenging cases involving the autocorrelation of the moments. Then, we apply this result to address the efficiency of GMM in two cases of interest: parametric and semiparametric.

First, we address the following simple question: Under which conditions on the set of moment restrictions is the (optimal) GMM estimator asymptotically as efficient as the MLE? To answer this question, it suffices to compare the asymptotic variance of $\hat{\theta}$ with the inverse of Fisher's information matrix. We show that the GMM estimator reaches the Cramér–Rao bound if and only if the data generating process (DGP) score belongs to the closure¹ of the set of moment conditions. Although this result may sound familiar, it generalizes former results in two dimensions: (a) the moment conditions may be autocorrelated; (b) the number of moment conditions may be infinite because we allow for both a continuum and a countably infinite number of moment conditions. Our result can be particularly useful to devise an efficient GMM estimator in the cases where MLE is either intractable or burdensome. For instance, it permits us to establish that the GMM estimator based on power functions or on the characteristic function (CF) is asymptotically efficient. In particular, it implies that the GMM estimator based on the joint CF of $(Y_t, Y_{t-1}, \dots, Y_{t-p})$ is asymptotically efficient when Y_t is Markov of order p . To our knowledge, we are the first to give a formal proof of this result. Feuerverger (1990) establishes a similar result using a grid on the argument of the CF that becomes finer and finer.² Singleton (2001) proposes efficient estimators based on the conditional CF using either a grid or the inverse Fourier transform.

Second, we consider a dynamic model where the only knowledge of the underlying model is that the observations are Markov and a conditional moment restriction is satisfied. We show that the GMM bound provides an efficiency bound in the much larger class of regular estimators. This generalizes the results proved by Chamberlain (1987) in the independent and identically distributed (i.i.d.) context, Hansen (1993) for linear models, and Wefelmeyer (1996) for moments based on the conditional mean and variance.

The rest of the paper is organized as follows. Section 2 presents a general framework for GMM estimation with a finite or infinite number of moments. Section 3 characterizes the asymptotic variance of GMM in terms of inner products in a RKHS. Section 4 investigates the (parametric) asymptotic efficiency of GMM. Section 5 gives the semiparametric efficiency bound for conditional moment restrictions. Section 6 concludes. Appendix A gives a brief review of the main definitions and properties of operators and reproducing kernel Hilbert spaces. The proofs of the main propositions are in Appendix B.

2. A GENERAL FRAMEWORK FOR GMM ESTIMATION

This section provides a general framework for GMM that permits us to handle finitely and infinitely many moment restrictions.

2.1. Assumptions

Assume that the process $\{Y_t\}_{t=-\infty}^{t=\infty}$ is stationary and ergodic with distribution F_0 . To allow for the most general setting, F_0 is not assumed to belong to a parametric family. The term Y_t may be multivariate³ so that $Y_t \in \mathbb{R}^l$. The observations are given by $\{Y_1, \dots, Y_T\}$. Let $\theta \in \Theta \subset \mathbb{R}^q$ be the parameter of interest. The relationship between F_0 and the true value of the parameter θ_0 is

$$\mathbb{E}^{F_0} [h(\tau, Y_t; \theta_0)] = 0, \quad \tau \in I,$$

where I is an index set that may be an interval or a countably finite or infinite set and $h(\tau, Y_t, \theta)$ is a function from $I \times \mathbb{R}^l \times \mathbb{R}^q$ to \mathbb{R} or \mathbb{C} . If I is countable, $h(\tau, Y_t; \theta)$ selects the τ th row of the vector h .

Inference is based on a set of moment conditions

$$\mathbb{E}^{F_0} [h(\tau, Y_t; \theta)] = 0, \quad \tau \in I. \tag{2.1}$$

We assume that θ is identified.

Assumption 2.1 (Identification condition).

$$\mathbb{E}^{F_0} [h(\tau, Y_t; \theta)] = 0, \quad \forall \tau \in I \Leftrightarrow \theta = \theta_0. \tag{2.2}$$

The number of moment conditions may be finite, as in the traditional method of moments, or infinite. The moment conditions may be countable, e.g., $l = 1$ and $h(\tau, Y_t; \theta) = Y_t^\tau - m(\tau, \theta)$, $\tau = 1, 2, \dots$ where $m(\tau, \theta) = \mathbb{E}^{F_0}(Y_t^\tau)$. Also the moment conditions may be a continuum, e.g., $h(\tau, Y_t; \theta) = \exp(i\tau'Y_t) - \psi_\theta(\tau)$, $\tau \in \mathbb{R}^l$, where $\psi_\theta(\tau) = \mathbb{E}^{F_0}[\exp(i\tau'Y_t)]$ is the characteristic function of Y_t . For efficiency considerations, we might want to base the estimation on a full set of moment conditions indexed by $\tau \in I$ even if identification is achieved from a subset of I .

Let π be a positive measure defined on I with support identical to I . Let $L^2(I, \pi)$ be the Hilbert space of complex or real-valued functions that are square integrable with respect to π . For complex-valued functions, we write

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

The inner product on this space is

$$\langle f, g \rangle = \int_I f(\tau) \overline{g(\tau)} \pi(d\tau),$$

where \bar{g} denotes the complex conjugate of g . The term $\|g\|^2 = \int_I |g(\tau)|^2 \pi(d\tau)$ denotes the norm of g in $L^2(I, \pi)$. Note that $L^2(I, \pi)$ is always separable,⁴ it has a countable basis, and hence it is isomorphic to l^2 where l^2 is the space consisting of sequences $(\alpha_i)_{i=1}^\infty$ such that $\sum_{i=1}^\infty |\alpha_i|^2 < \infty$.

Let $L^2(F_0)$ be the Hilbert space⁵ of complex-valued functions $\varphi : \mathbb{R}^l \rightarrow \mathbb{C}$ such that $\sum_{j=-\infty}^\infty \mathbb{E}^{F_0} \left(\varphi(Y_0) \overline{\varphi(Y_j)} \right) < \infty$ and

$$\langle \Phi, \Psi \rangle_{L^2(F_0)} = \sum_{j=-\infty}^\infty \mathbb{E}^{F_0} \left(\Phi(Y_0) \overline{\Psi(Y_j)} \right).$$

The asymptotic theory in Carrasco and Florens (2000) relies on a central limit theorem in a Hilbert space. It is therefore necessary that $h(\tau, Y_t; \theta)$ is a random element in $L^2(I \times \mathbb{R}^l, \pi \times F_0)$.

Assumption 2.2. $h(\tau, Y_t; \theta)$ belongs to $L^2(I \times \mathbb{R}^l, \pi \times F_0)$.

Assumption 2.2 is equivalent to $\int \sum_{j=-\infty}^\infty \mathbb{E}^{F_0} \left[h(\tau, Y_0; \theta) \overline{h(\tau, Y_j; \theta)} \right] \pi(d\tau) < \infty$. By Fubini's theorem (Billingsley, 1995), Assumption 2.2 implies that $h(\tau, Y_t; \theta)$ as a function of τ belongs to $L^2(I, \pi)$ almost surely and $h(\tau, Y_t; \theta)$ as a function of Y_t belongs to $L^2(F_0)$ almost surely. Assumption 2.2 requires that the temporal dependence between $h(\tau, Y_0; \theta)$ and $h(\tau, Y_j; \theta)$ vanishes sufficiently fast over time. For instance, if $\sum_{j=-\infty}^\infty \mathbb{E}^{F_0} \left[h(\tau, Y_0; \theta) \overline{h(\tau, Y_j; \theta)} \right] < C$ where C is some constant independent of τ and π is a probability measure, then this condition is automatically satisfied. The role of π is mainly to insure that Assumption 2.2 is satisfied. Depending on whether the number of moment restrictions is finite or infinite, the norm in $L^2(I, \pi)$ may take various forms. Examples of spaces $L^2(I, \pi)$ are as follows.

- Finite number of restrictions: Let $I = \{1, \dots, M\}$ and π be the uniform probability measure on I . Then, the norm is the euclidean norm $\|g\|^2 = (1/M) \sum_{\tau=1}^M |g(\tau)|^2$.
- Countably infinite number of restrictions: Let $I = \mathbb{N}$ and if we define π as a counting measure, i.e., $\pi(A) = \text{cardinal}(A)$, we have $L^2(I, \pi) = l^2$

the space of functions $g = (g(1), g(2), \dots)$ such that $\|g\|^2 = \sum_{\tau=0}^{\infty} |g(\tau)|^2 < \infty$. However, this space may be too small to include some moment conditions of interest. A way to enlarge this space is to choose π to be a probability measure on \mathbb{N} (e.g., a Poisson probability measure such that $\pi\{\tau\} = e^{-1}/\tau!$, $\tau = 0, 1, 2, \dots$); then $L^2(I, \pi)$ is defined as the set of functions g such that $\|g\|^2 = \sum_{\tau=0}^{\infty} |g(\tau)|^2 \pi(\tau) < \infty$ is larger than l^2 .

- Continuum of restrictions on an interval: Let $I = [0, 1]$, and π may be chosen as Lebesgue measure on $[0, 1]$; then the norm is $\|g\|^2 = \int_0^1 |g(\tau)|^2 d\tau$.
- Continuum of restrictions on \mathbb{R} : Let $Y_t \in \mathbb{R}$ and $I = \mathbb{R}$; here again π could be set equal to Lebesgue measure if it is believed that the moment functions $h(\tau, Y_t; \theta)$ are square integrable with respect to Lebesgue. In some cases, this is not realistic, and it may be useful to enlarge the space using an appropriate choice of π . For example, $h(\tau, Y_t; \theta) = \exp(i\tau Y_t) - \psi_\theta(\tau)$ is not square integrable with respect to Lebesgue but is square integrable with respect to any probability measure π on \mathbb{R} . Then, the norm is $\|g\|^2 = \int |g(\tau)|^2 \pi(d\tau)$.
- Vector of continuum of restrictions: Let $I = \{1, \dots, M\} \times \mathbb{R}$; then π could be chosen as the product measure of a uniform distribution on $\{1, \dots, M\}$ and the normal distribution.

Under Assumption 2.2 and some other restrictions on the temporal dependence of Y_t , a functional central limit theorem guarantees that $(1/\sqrt{T}) \sum_{t=1}^T h(\cdot, Y_t; \theta)$ converges in $L^2(I, \pi)$ to a Gaussian process (denoted ξ) with mean zero and covariance operator K (see Chen and White, 1998). The covariance operator K is defined⁶ as the operator from $L^2(I, \pi)$ into $L^2(I, \pi)$ such that

$$\langle K\varphi, \varphi \rangle = \text{Var}(\langle \xi, \varphi \rangle)$$

for all φ in $L^2(\pi, I)$. In our setting, the covariance operator K takes the form of an integral operator

$$(K\varphi)(\tau) = \int_I k(\tau, \lambda) \varphi(\lambda) \pi(d\lambda) \tag{2.3}$$

for all φ in $L^2(\pi, I)$ where k is called the kernel of K . In the simplest case where $\{h(\cdot, Y_t; \theta_0)\}$ are i.i.d., we have

$$\begin{aligned} &\text{Var}(\langle h(\cdot, Y_t; \theta_0), \varphi(\cdot) \rangle) \\ &= \mathbb{E}^{F_0} \left(\int h(\tau, Y_t; \theta_0) \varphi(\tau) \pi(d\tau) \int \overline{h(\lambda, Y_t; \theta_0) \varphi(\lambda)} \pi(d\lambda) \right) \\ &\quad \times \int \left(\int \mathbb{E}^{F_0} (h(\tau, Y_t; \theta_0) \overline{h(\lambda, Y_t; \theta_0)}) \varphi(\tau) \pi(d\tau) \right) \overline{\varphi(\lambda)} \pi(d\lambda) \\ &= \langle K\varphi, \varphi \rangle. \end{aligned}$$

We see that the kernel of the covariance operator $k(\tau, \lambda) = \mathbb{E}^{F_0} (h(\tau, Y_t; \theta_0) \overline{h(\lambda, Y_t; \theta_0)})$ remains unchanged for every π . In the general case with autocorrelation, the kernel k takes the form

$$k(\tau, \lambda) = \langle h(\tau), h(\lambda) \rangle_{L^2(F_0)}, \tag{2.4}$$

where

$$\langle h(\tau), h(\lambda) \rangle_{L^2(F_0)} = \sum_{j=-\infty}^{\infty} \mathbb{E}^{F_0} \left(h(\tau, Y_0; \theta_0) \overline{h(\lambda, Y_j; \theta_0)} \right).$$

The case where the moment functions are uncorrelated can be regarded as a special case of (2.4) where $\langle h(\tau), h(\lambda) \rangle_{L^2(F_0)} = \mathbb{E}^{F_0} (h(\tau, Y_t; \theta_0) \overline{h(\lambda, Y_t; \theta_0)})$. In the case of a countably infinite number of moments $Kg(\tau) = \sum_{\lambda=1}^{\infty} k(\tau, \lambda) g(\lambda) \pi(\lambda)$, so that K is an infinite-dimensional matrix.

It is important to notice that Assumption 2.2 can be rewritten as

$$\int k(\tau, \tau) \pi(d\tau) < \infty, \tag{2.5}$$

which implies that K is nuclear and hence compact (see Kannan and Bharucha-Reid, 1970; Appendix A in the current paper). Condition (2.5) will be checked in various examples in Sections 4 and 5.1.

2.2. Optimal GMM Estimator

Let $h_T(\cdot, \theta) = \frac{1}{T} \sum_{t=1}^T h(\cdot, Y_t; \theta)$. The GMM estimators are solutions of

$$\min_{\theta} \| Bh_T(\cdot; \theta) \|^2,$$

where B is an operator from $L^2(I, \pi)$ in $L^2(I, \pi)$. Denote $K^{-1}f$ the solution g (when it exists) of the equation $Kg = f$ and $K^{-1/2} = (K^{-1})^{1/2}$ (for more details, see Carrasco, Florens, and Renault, 2007, Sect. 6.1.1). The optimal GMM estimator in the class of GMM estimators is obtained for $B = K^{-1/2}$. Let K_T be an estimator of K . In the case of a finite number of moments, we have the usual GMM objective function $h_T' K_T^{-1} h_T$ where $h_T = (h_T(1; \theta), \dots, h_T(M; \theta))'$ and K is the asymptotic covariance matrix of h_T . Then K_T^{-1} exists as long as 0 is not an eigenvalue of K_T . This simple result does not hold when I is infinite dimensional. As K is compact, the equation $Kg = f$ has a solution only for f in the range of K that is a strict subset of $L^2(I, \pi)$. Moreover, when the solution exists, it is unstable in g . When I has infinite dimension, $K^{-1/2}$ is not defined on the whole space $L^2(I, \pi)$ but on a subset denoted $\mathcal{H}(K)$ and called the reproducing kernel Hilbert space (RKHS). We use the notation $\langle g, f \rangle_K \equiv \langle K^{-1/2}g, K^{-1/2}f \rangle$ for the inner product in the RKHS associated with K . As the inverse of K is not continuous,

it needs to be stabilized by the introduction of a regularization (or smoothing) term denoted α_T . The term $(K^{\alpha_T})^{-1}$ is called the regularized inverse of K if it has the property that $(K^{\alpha_T})^{-1} f - K^{-1} f \rightarrow 0$ as α_T goes to zero, for all f in the range of K . Let $(K_T^{\alpha_T})^{-1/2} = \left((K_T^{\alpha_T})^{-1} \right)^{1/2}$ be the regularized estimate of $K^{-1/2}$. The GMM estimator is defined as

$$\hat{\theta} = \arg \min_{\theta} \left\| (K_T^{\alpha_T})^{-1/2} h_T(\cdot; \theta) \right\|^2. \tag{2.6}$$

In Carrasco and Florens (2000) and Carrasco, Chernov, Florens, and Ghysels (2007), we apply the Tikhonov regularization, which consists in defining $(K_T^{\alpha_T})^{-1/2}$ as $(K_T^2 + \alpha_T I)^{-1/2} K_T^{1/2}$. Then, the objective function in (2.6) can be rewritten in terms of products of vectors and matrices, as shown in Carrasco, Chernov, Florens, and Ghysels (2007, Prop. 3.4.). Besides Tikhonov, other forms of regularization could be used; see Kress (1999) and Carrasco, Florens, and Renault (2007). Carrasco (2012) investigates two other regularizations for the estimation of a linear instrumental variables model: one iterative method called Landweber Fridman and another consisting in using the eigenfunctions of K that are associated with the largest eigenvalues, called spectral cut-off. Another approach, frequently used when the moment conditions are countable, consists in truncating the sequence of moment conditions. This method can be regarded as a regularization scheme because it also involves a smoothing parameter which is the number of moment conditions used in the estimation. Engl, Hanke, and Neubauer (2000, Sects. 3.2 and 5.2) refer to this method as “projection” and recommend using it jointly with Tikhonov regularization when the data are noisy. Kuersteiner (2001) shows that efficiency (in a GMM sense) can be achieved by letting the number of moment conditions used in the estimation increase as the sample size increases. Donald and Newey (2001) give a method for selecting the optimal number of instruments by minimizing the mean square error of the estimator. One difference between projection and the other three regularizations mentioned earlier is that the projection requires an a priori ordering of the instruments or moment conditions, whereas the other methods are based on the spectral decomposition of the covariance operator and avoid ordering the moment conditions a priori. Another major difference is that in Kuersteiner (2001) and Donald and Newey (2001), the covariance matrix is not compact. For instance, the identity matrix is not compact. Our asymptotic results do not cover this case. It should be stressed that some assumptions necessary to apply the standard methods (in particular the assumption that the smallest eigenvalue is bounded away from zero) are not realistic in the presence of many moment conditions. The problem of the existence and stability of the solution to an equation $K\varphi = f$ arises whether K is an integral operator (case of a continuum of moment conditions) or K is an infinite-dimensional matrix (case of countable moments). The degree of difficulty is the same in both cases;

see Böttcher, Dijkstra, Langer, Dritschel, Rovnyak, and Kaashoek (1996, Lect. Ser. 1). Our method can handle the countable case as illustrated in Section 2.3.

Under some regularity conditions and an appropriate decay rate of a_T toward zero, $\hat{\theta}$ satisfies

$$\sqrt{T} \left(\hat{\theta} - \theta_0 \right) \xrightarrow{d} \mathcal{N} \left(0, \Sigma^{-1} \right), \tag{2.7}$$

where Σ is the $q \times q$ matrix with (i, j) element equal to

$$\left\langle \mathbb{E}^{F_0} \left(\nabla_{\theta_i} h(\theta) \right), \mathbb{E}^{F_0} \left(\nabla_{\theta_j} h(\theta) \right) \right\rangle_K \Big|_{\theta=\theta_0}. \tag{2.8}$$

For I finite, formula (2.8) boils down to the familiar expression

$$\left\{ \mathbb{E}^{F_0} \left(\nabla_{\theta} h \right) K^{-1} \mathbb{E}^{F_0} \left(\nabla_{\theta} h \right)' \right\}^{-1} \Big|_{\theta=\theta_0}$$

using matrix notation. Note that the asymptotic variance of $\hat{\theta}$ given by (2.8) does not depend on the choice of π except through its support (see Remark 6 in Appendix A). The result of (2.7) is shown in Carrasco and Florens (2000) for $I = [0, 1]$ and i.i.d. data and in Carrasco, Chernov, Florens, and Ghysels (2007) for $I = \mathbb{R}^p$ and time series. In the presence of a countably infinite number of moments, the result has been established by Carrasco (2012) for three regularizations. In what follows, we assume that the conditions under which result (2.7) holds are satisfied.

2.3. An Illustrative Example with Countable Moment Restrictions

Consider a countable infinite sequence of moments $h(\tau, Y_t; \theta_0)$, $\tau = 1, 2, \dots$ that are i.i.d. $(0, 1)$. If π is the counting measure on $\mathbb{R}^{\mathbb{N}}$, the associated covariance matrix is the infinite-dimensional identity matrix and obviously not compact. We can make it compact by choosing $\pi = (\pi_{\tau})$, $\tau = 1, 2, \dots$ as a probability measure on $\mathbb{R}^{\mathbb{N}}$. Let us check that for any probability measure π , the covariance operator K is nuclear (and hence compact). Let φ_{τ} denote the τ th element of an arbitrary sequence $\varphi = (\varphi_{\tau})_{\tau}$. Here K satisfies $\text{Var} \langle h, \varphi \rangle = \langle K\varphi, \varphi \rangle$ for any $\varphi \in L^2(\mathbb{R}^{\mathbb{N}}, \pi)$, and $\text{Var} \langle h, \varphi \rangle = \text{Var} \left(\sum_{\tau \in \mathbb{N}} h(\tau, Y; \theta_0) \varphi_{\tau} \pi_{\tau} \right) = \sum_{\tau \in \mathbb{N}} \mathbb{E} \left(h(\tau, Y; \theta_0)^2 \right) \varphi_{\tau}^2 \pi_{\tau}^2 = \sum_{\tau \in \mathbb{N}} \varphi_{\tau}^2 \pi_{\tau}^2 = \sum_{\tau \in \mathbb{N}} (K\varphi)_{\tau} \varphi_{\tau} \pi_{\tau}$. Hence $(K\varphi)_{\tau} = \varphi_{\tau} \pi_{\tau}$. Therefore, K is the infinite-dimensional diagonal matrix with π_{τ} as τ th diagonal element. Its eigenvalues μ_j and (orthonormalized) eigenfunctions ϕ_j satisfy $K\phi_j = \mu_j \phi_j$; hence ϕ_j is the vector of zeroes with j th element equal to $1/\sqrt{\pi_j}$ and $\mu_j = \pi_j$. The operator is nuclear because $\sum_j \mu_j = \sum_j \pi_j = 1 < \infty$. Finally, $\mathbb{E} \left(h(\tau, Y; \theta) \right)$ needs to belong to $\mathcal{H}(K)$ for θ in a neighborhood of θ_0 . From Definition A.3 given in Appendix A,

a function φ belongs to $\mathcal{H}(K)$ if and only if $\|\varphi\|_K^2 = \sum_j \langle \varphi, \phi_j \rangle^2 / \mu_j < \infty$. We have

$$\begin{aligned} \sum_j \frac{\langle \varphi, \phi_j \rangle^2}{\mu_j} &= \sum_j \frac{(\varphi_j(\phi_j)_j \pi_j)^2}{\pi_j} \\ &= \sum_j \frac{(\varphi_j \frac{1}{\sqrt{\pi_j}} \pi_j)^2}{\pi_j} \\ &= \sum_j \varphi_j^2. \end{aligned}$$

Hence, φ belongs to $\mathcal{H}(K)$ if $\sum_\tau \varphi_\tau^2 < \infty$. We see that $\mathbb{E}(h(\tau, Y; \theta))$ belongs to $\mathcal{H}(K)$ for $\theta = \theta_0$ because $\mathbb{E}(h(\tau, Y; \theta_0)) = 0$ but may or may not belong to $\mathcal{H}(K)$ for $\theta \neq \theta_0$. However, if $\mathbb{E}(h(\tau, Y; \theta))$ is continuous in θ , this condition is likely to be satisfied in a neighborhood of θ_0 .

3. NORM IN A RKHS AND VARIANCE OF GMM ESTIMATOR

Our aim is to compute the elements $\langle f, g \rangle_K$ that show up in the inverse of the asymptotic variance of the GMM estimator (eqn. (2.8)). Calculation of this variance will permit us to establish whether the estimator is efficient or not.

3.1. Norm in a RKHS

Define $L^2(h)$ as the closure of the subspace of $L^2(F_0)$ spanned by $\{h(\tau), \tau \in I\}$. It consists of linear combinations (for some integer n , real $\omega_1, \dots, \omega_n$ and indexes τ_1, \dots, τ_n) of the form

$$G_n = \sum_{l=1}^n \omega_l h(\tau_l)$$

and their limits in mean square, i.e., random variables G such that $\|G_n - G\|_{L^2(F_0)}^2 \rightarrow 0$ as n goes to infinity.

Let $\mathcal{H}(K)$ denote the RKHS associated with the kernel k . It is a subspace of functions defined on I . A definition and some properties are given in Appendix A. As discussed earlier, the notations $\|f\|_K$ and $\langle f, g \rangle_K$ represent the norm and inner product in $\mathcal{H}(K)$. The following proposition gives a way to compute $\langle f, g \rangle_K$ for any $f, g \in \mathcal{H}(K)$. Let

$$\mathcal{G}_g = \left\{ G \in L^2(F_0) : g(\tau) = \langle G, h(\tau) \rangle_{L^2(F_0)}, \quad \forall \tau \in I \right\}.$$

PROPOSITION 3.1. *Assume h satisfies Assumption 2.2. Let k be a covariance kernel defined in (2.4) and $\mathcal{H}(K)$ be its associated RKHS.*

- (i) $\mathcal{H}(K)$ and $L^2(h)$ are isometrically isomorphic, i.e., there exists a one-to-one linear mapping \tilde{J} (not necessarily unique) from $\mathcal{H}(K)$ to $L^2(h)$ such that

$$\langle f, g \rangle_K = \left\langle \tilde{J}(f), \tilde{J}(g) \right\rangle_{L^2(F_0)}$$

for all $f, g \in \mathcal{H}(K)$. The term \tilde{J} is referred to as Hilbert space isomorphism from $\mathcal{H}(K)$ to $L^2(h)$.

- (ii) Let J be the Hilbert space isomorphism from $\mathcal{H}(K)$ to $L^2(h)$ that transforms $k(\cdot, \tau)$ into $h(\tau)$, i.e., $J(k(\cdot, \tau)) = h(\tau)$. We have $J(g) = G$ and $\|g\|_K^2 = \|G\|_{L^2(F_0)}^2$ where G is the unique element of $\mathcal{G}_g \cap L^2(h)$.
- (iii) Another characterization of J is as follows: $J(g) = \operatorname{arg\,min}_{G \in \mathcal{G}_g} \|G\|_{L^2(F_0)}^2$, and hence $\|g\|_K^2 = \min_{G \in \mathcal{G}_g} \|G\|_{L^2(F_0)}^2$.

The results of Proposition 3.1 are scattered in Parzen (1959, 1970) and Saitoh (1997). Although these results are not new, they have not been presented under this compact form before. In Appendix B, we provide an independent proof of this proposition that we think is useful to understand the role played by the reproducing property. Note that (i) and (ii) imply that for any $g_1, g_2 \in \mathcal{H}(K)$ and $G_i \in \mathcal{G}_i \cap L^2(h)$, $i = 1, 2$ (where \mathcal{G}_i denotes the set \mathcal{G}_g associated with g_i), we have

$$\langle g_1, g_2 \rangle_K = \langle G_1, G_2 \rangle_{L^2(F_0)}.$$

3.2. Asymptotic Variance of GMM Estimator

To calculate the asymptotic variance of the GMM estimator, we apply to equation (2.8) the results from Proposition 3.1 with $g_i = -\mathbb{E}^{F_0}(\nabla_{\theta_i} h(\tau, Y_i; \theta))|_{\theta=\theta_0}$. Define

$$\mathcal{G}_i = \left\{ G \in L^2(F_0) : -\mathbb{E}^{F_0}(\nabla_{\theta_i} h(\tau, Y_i; \theta))|_{\theta=\theta_0} = \langle G, h(\tau) \rangle_{L^2(F_0)}, \quad \forall \tau \in I \right\},$$

for $i = 1, 2, \dots, p$.

COROLLARY 3.1. *Assume that h satisfies Assumption 2.2 and the covariance operator defined in (2.3) is injective. Then, the inverse of the asymptotic variance of $\hat{\theta}$ (defined in (2.6)), Σ , has for element (i, j) :*

$$\Sigma_{ij} = \langle G_i, G_j \rangle_{L^2(F_0)}$$

$i, j = 1, \dots, q$, where $G_i \in \mathcal{G}_i \cap L^2(h)$ for $i = 1, 2, \dots, q$. Moreover for any $\tilde{G}_i \in \mathcal{G}_i$, $i = 1, 2, \dots, q$, the matrix, $\tilde{\Sigma}$, with principal element $(\tilde{G}_i, \tilde{G}_j)_{L^2(F_0)}$ satisfies the property that $\tilde{\Sigma} - \Sigma$ is nonnegative definite.

Corollary 3.1 shows that the G_i in $L^2(h)$ corresponds to the worst case (among the G_i in \mathcal{G}_i), in the sense that they correspond to the largest value of the variance of the estimator.

The condition that K is injective means that it does not admit a zero eigenvalue and $K^{-1/2}$ is well defined on $\mathcal{H}(K)$. We need this condition to define the GMM estimator as in (2.6). The injectiveness of K implies that the moment conditions are linearly independent, in the following sense: $\langle h, \omega \rangle = 0 \Rightarrow \omega = 0$ for any function ω of $L^2(I, \pi)$. To see this, we write $\int h(\lambda, Y_t; \theta_0) \omega(\lambda) \pi(d\lambda) = 0 \Rightarrow \mathbb{E}(h(\tau, Y_t; \theta) \int h(\lambda, Y_t; \theta_0) \omega(\lambda) \pi(d\lambda)) = 0 \Leftrightarrow \int \mathbb{E}(h(\tau, Y_t; \theta) \overline{h(\lambda, Y_t; \theta_0)}) \omega(\lambda) \pi(d\lambda) = 0 \Leftrightarrow K\omega = 0 \Rightarrow \omega = 0$ where the third equality is justified by Assumption 2.2 and Fubini’s theorem.

In the rest of the paper, we use Corollary 3.1 to derive conditions on (parametric and semiparametric) efficiency. However, it is worth noting that Corollary 3.1 is interesting in its own right and could be used in other circumstances, e.g., to establish the redundancy of moments as in Hall, Inoue, Jana, and Shin (2007).

Our approach is closely related to that of Hansen (1985) and Hansen, Heaton, and Ogaki (1988). Hansen (1985) considers the model

$$e = m(X, \theta),$$

where θ is the parameter of interest. There is a sequence $Z_j, j = 1, 2, \dots$ of instruments so that

$$\mathbb{E}(Z_j e) = 0 \tag{3.1}$$

for all $j = 1, 2, \dots$. Hansen derives the greatest lower bound for the asymptotic covariance matrices of GMM estimators based on any subsets of moments of type (3.1). We can derive this bound using the tools developed here.⁷ Indeed, this bound is the variance of the GMM estimator based on all the moment conditions. This is true because the variance of the GMM estimator can never increase when the number of moments increases. So formulas (2.7)–(2.8) and Corollary 3.1 provide a way to compute the efficiency bound of GMM estimators based on $h(j; \theta) = Z_j m(X, \theta), j = 1, 2, \dots$. When $\{h(j; \theta)\}$ are martingale difference sequences (m.d.s.), our method is very similar to that of Hansen (1985), as the condition satisfied by the elements of \mathcal{G}_i is the same as Hansen’s condition (4.8). When $\{h(j; \theta)\}$ are not m.d.s., both approaches are different because Hansen follows Gordin by approximating stationary ergodic processes by m.d.s. Our method is more straightforward to apply.

3.3. Role of the Measure π

The measure π has been introduced to guarantee a functional central limit theorem on $\sqrt{T}h_T(\theta_0)$. How does it affect the asymptotic variance of $\hat{\theta}$ given in (2.8)? From Definition A.2 and Proposition 3.1, we see that π does not play a role in the norm $\|\varphi\|_K$ (where φ is an arbitrary function) provided the support I remains

the same. So two equivalent measures will give the same norm $\|\varphi\|_K$. If the support I of π is changed, then different moment conditions are used and obviously the norm $\|\varphi\|_K$ is altered; otherwise $\|\varphi\|_K$ is not affected by the choice of π . This is an interesting result because it implies that the asymptotic variance of $\hat{\theta}$ does not depend on π . Section 2.3 provides an example where $\|\varphi\|_K^2$ is computed explicitly; it is equal to $\sum_j \varphi_j^2$.

In the case with a finite number of moments, changing π (but keeping the same support) corresponds to multiplying the vector of moments by a different diagonal matrix. It obviously does not change the GMM objective function and hence the value of the estimator $\hat{\theta}$ (even for finite n). Now consider the case of an infinite number of moments. In (2.6), $\hat{\theta}$ is defined as the minimum of the norm of $h_T(\cdot, \theta)$ in the RKHS associated with $K_T^{\alpha T}$. If there were no regularization, $\hat{\theta}$ would be independent of π . The effect of the regularization is not clear.

4. MLE EFFICIENCY OF GMM ESTIMATORS

In this section, we investigate the conditions under which the GMM estimator is asymptotically as efficient as MLE. When we say that a GMM estimator is efficient, we mean that its asymptotic variance is equal to the inverse of the Fisher information matrix. Godambe (1960) shows that the GMM estimator based on the score is efficient. However, this is not the only way for GMM to reach efficiency. It can also be reached using a continuum or a countably infinite number of moment conditions. We assume that the distribution of Y_t belongs to a parametric family indexed by θ ; its conditional probability density function (pdf) is denoted $f_\theta(Y_t|Y_{t-1}, Y_{t-2}, \dots, Y_{t-p})$ or $f_\theta(Y_t|Y_{t-1}^p)$. The term F_0 corresponds to the case where $\theta = \theta_0$, so that \mathbb{E}^{F_0} is replaced by \mathbb{E}^{θ_0} . We consider arbitrary functions $h(\tau, Y_t^{p+1}, \theta_0)$ that satisfy $\mathbb{E}^{\theta_0} h(\tau, Y_t^{p+1}; \theta_0) = 0$ where Y_t^{p+1} is the $(p + 1)$ -vector of random variable $Y_t^{p+1} = (Y_t, Y_{t-1}, \dots, Y_{t-p})'$. We want to provide a necessary and sufficient condition on h for the GMM estimator to be asymptotically efficient. To answer this question, it suffices to compare the asymptotic variance given in Corollary 3.1 with the inverse of Fisher's information matrix.

Assumption 4.1.

- (a) Y_t is a stationary ergodic Markov chain of order p .
- (b) Let $s_t = (\partial \ln f_\theta / \partial \theta)(Y_t|Y_{t-1}^p)|_{\theta=\theta_0}$. The information matrix $\mathbb{E}^{\theta_0} [s_t s_t']$ exists and is positive definite.
- (c) (Differentiation under the integral sign) Let $a(y_0^{T+1}, \theta) = h(\tau, y_t^{p+1}; \theta) f_\theta(y_0^{T+1})$ where $f_\theta(y_0^{T+1})$ denotes the joint density of $y_0^{T+1} = (y_0, y_{-1}, \dots, y_{-T})$. Here $a(y_0^{T+1}, \theta)$ is continuously differentiable in a neighborhood \mathcal{N} of θ_0 and

$$\int \sup_{\theta \in \mathcal{N}} \left\| \nabla_\theta a(y_0^{T+1}, \theta) \right\| dy_0^{T+1} < \infty.$$

PROPOSITION 4.1. Assume that Assumptions 2.2 and 4.1 and equations (2.7) and (2.8) hold.

(i) The inverse of the asymptotic variance of $\hat{\theta}$, the GMM estimator based on $\left\{h\left(\tau, Y_t^{p+1}; \theta\right), \tau \in I\right\}$, is given by

$$\Sigma = \langle Proj, Proj \rangle_{L^2(F_0)},$$

where $Proj$ is the orthogonal projection (in $L^2(F_0)$) of s_t on $L^2(h)$ and $L^2(h)$ the closure of the span of $\left\{h\left(\tau, Y_t^{p+1}; \theta_0\right), \tau \in I\right\}$.

(ii) Moreover, Σ coincides with Fisher’s information matrix if and only if

$$s_{ti} \in L^2(h) \quad \text{for all } i = 1, 2, \dots, q, \tag{4.1}$$

where s_{ti} is the i th element of s_t .

Consider the special case where I is finite and all finite subsets of h are linearly independent (there are no redundant moment conditions). Denote $h_t \equiv \left\{h\left(\tau, Y_t^{p+1}; \theta_0\right), \tau \in I\right\}$ the vector of moment conditions. The projection of s_t on $L^2(h)$ is equal to

$$Proj = h_t \langle h, h \rangle_{L^2(F_0)}^{-1} \langle h, s \rangle_{L^2(F_0)},$$

and Σ becomes

$$\begin{aligned} \Sigma &= \langle h, s \rangle'_{L^2(F_0)} \langle h, h \rangle_{L^2(F_0)}^{-1} \langle h, h \rangle_{L^2(F_0)} \langle h, h \rangle_{L^2(F_0)}^{-1} \langle h, s \rangle_{L^2(F_0)} \\ &= \langle h, s \rangle'_{L^2(F_0)} \langle h, h \rangle_{L^2(F_0)}^{-1} \langle h, s \rangle_{L^2(F_0)} \\ &= \sum_{j=-\infty}^{\infty} \mathbb{E}^{\theta_0} \left(s_{t-j} h'_t \right) \left(\sum_{j=-\infty}^{\infty} \mathbb{E}^{\theta_0} \left(h_t h'_{t-j} \right) \right)^{-1} \sum_{j=-\infty}^{\infty} \mathbb{E}^{\theta_0} \left(h_t s_{t-j} \right). \end{aligned}$$

We recognize the expression of the inverse of the asymptotic covariance matrix of the GMM estimator characterized by Assumption 2 and Theorem 1(a) of Hall et al. (2007), for a finite number of moment conditions.

Note that MLE efficiency is obtained only for the optimal GMM estimator, i.e., if the optimal weighting matrix/operator is used. A similar result (for a countable sequence of moments) is discussed by Tauchen (1997) in the static case and by Gallant and Long (1997) in the context of moment conditions based on the score of a seminonparametric auxiliary model, in the dynamic case. Bates and White (1993) show that the GMM estimator has minimum variance in the class of minimum discrepancy estimators. Davidson (2000) gives a way to construct efficient estimators by projection.

The proof of Proposition 4.1 uses the results of Corollary 3.1. It is proved in the general case where $\left\{h\left(\tau, Y_t^{p+1}; \theta\right)\right\}$ is autocorrelated. The key of the proof

consists in showing that the score s_{ti} belongs to \mathcal{G}_i . In the independent case, this is equivalent to showing

$$\mathbb{E}^\theta \left(\frac{\partial h(\tau, Y_i^{p+1}; \theta)}{\partial \theta_i} \right) \Bigg|_{\theta=\theta_0} = \text{cov}^{\theta_0} \left(s_{ti}, h(\tau, Y_i^{p+1}; \theta_0) \right). \tag{4.2}$$

This equation can be obtained by differentiating $\mathbb{E}^\theta \left(h(\tau, Y_i^{p+1}; \theta) \right) = 0$ with respect to θ . Godambe (1960) proves the efficiency of MLE in the class of method of moments estimators, by applying Cauchy–Schwarz to equation (4.2).

The following corollary to Proposition 4.1 gives a sufficient condition for efficiency. Here $L^2 \left(Y_i^{p+1} \right)$ denotes the space of functions of Y_i^{p+1} that are square integrable.

COROLLARY 4.2. *Assume that Assumptions 2.2 and 4.1 and equations (2.7) and (2.8) hold. If $\{h(\tau) : \tau \in I\}$ spans $L^2 \left(Y_i^{p+1} \right)$ (equivalently $L^2(h) = L^2 \left(Y_i^{p+1} \right)$) then the GMM estimator based on $\{h(\tau, Y_i^{p+1}; \theta)\}$ is asymptotically efficient.*

The space $L^2 \left(Y_i^{p+1} \right)$ is a separable Hilbert space and hence by definition can be approximated by a countable sequence of functions. Note that this sequence is not unique. The statement $\{h(\tau) : \tau \in I\}$ spans $L^2 \left(Y_i^{p+1} \right)$ is equivalent to saying that the family of moment functions $\{h(\tau) : \tau \in I\}$ is complete in $L^2 \left(Y_i^{p+1} \right)$. The separability property has been exploited in various papers (see Chamberlain, 1987; Gallant, and Tauchen, 1996) to show that efficiency can be approached by using a sequence of moment conditions. In what follows, we provide some examples where the GMM estimator is as efficient as the MLE. They illustrate the fact that efficiency can be reached either from an countably infinite sequence or a continuum of moment conditions.

Example 1

Let $Y_i \in \mathbb{R}$ be i.i.d. with parametric distribution F_θ . Consider

$$h(\tau, Y, \theta) = m(\tau, Y) - \mathbb{E}^\theta(m(\tau, Y)).$$

The GMM estimator based on $(h(\tau, Y, \theta))_\tau$ is as efficient as the MLE if either (a) $(m(\tau, Y_i))_\tau$ is a dense basis of $L^2(Y_i)$ or (b) if the score belongs to the closure of the linear span of $(m(\tau, Y_i))_\tau$. Dense bases of $L^2(Y_i)$ (under some mild conditions on the density of Y_i) include power functions, wavelets, indicator functions, sines, cosines, and exponential functions (see, e.g., Debnath and Mikusinski, 2005). We detail three examples where a continuum of moments is available.

(1) Indicator function
 Consider

$$h(\tau, Y, \theta) = I(Y \leq \tau) - F_\theta(\tau), \quad \tau \in \mathbb{R}. \tag{4.3}$$

First, we check condition (4.1). Let $s(Y)$ denote the score (at θ_0). Because integrals can be approximated by infinite sums, it is enough to show that there exists $\varphi(\tau)$ such that

$$s(Y) = \int h(\tau, Y, \theta_0) \varphi(\tau) d\tau.$$

Replacing, $h(\tau, Y, \theta_0)$ by its expression, we obtain

$$s(Y) = \int_Y^\infty \varphi(\tau) d\tau - \int F_{\theta_0}(\tau) \varphi(\tau) d\tau. \tag{4.4}$$

Differentiating both sides of the equality with respect to Y yields

$$s'(Y) = -\varphi(Y).$$

Taking $\varphi(Y) = -s'(Y)$, using an integration by parts and the property $\mathbb{E}^{\theta_0}(s(Y)) = 0$, we see that (4.4) is satisfied. So provided that s is differentiable with respect to Y , condition (4.1) holds.

Now we check that Assumption 2.2 (or equivalently condition (2.5)) is satisfied, which is equivalent to checking that the covariance operator K is nuclear. The kernel of K is given by

$$\begin{aligned} k(\tau, \tau) &= \mathbb{E}^{\theta_0} \left[h(\tau, \theta_0)^2 \right] \\ &= \mathbb{E}^{\theta_0} \left[(I(Y \leq \tau) - F_{\theta_0}(\tau))^2 \right] \\ &= F_{\theta_0}(\tau) - F_{\theta_0}(\tau)^2. \end{aligned}$$

As $0 \leq k(\tau, \tau) \leq 1$, we have

$$\int k(\tau, \tau) \pi(d\tau) < \infty$$

provided π is a probability measure. Hence, Assumption 2.2 is satisfied. Moreover, equations (2.7) and (2.8) hold by Carrasco and Florens (2000). Therefore, assuming Assumption 4.1 holds, Proposition 4.1 implies that the GMM estimator based on the full continuum of moments (4.3) is asymptotically efficient.

(2) Characteristic function
 Consider

$$h(\tau, Y, \theta) = \exp(i\tau Y) - \psi_\theta(\tau), \tag{4.5}$$

where $\psi_\theta(\tau)$ is the characteristic function defined as $\mathbb{E}^\theta[\exp(i\tau Y)]$.

First, we check condition (4.1). Again, it is enough to check that there exists a function φ such that

$$\begin{aligned}
 s(Y) &= \int h(\tau, Y, \theta_0) \varphi(\tau) d\tau \\
 &= \int \exp(i\tau Y) \varphi(\tau) d\tau - \int \psi_{\theta_0}(\tau) \varphi(\tau) d\tau.
 \end{aligned}$$

Let

$$\varphi(\tau) = \frac{1}{2\pi} \int e^{-i\tau Y} s(Y) dY.$$

The Fourier inversion formula implies that

$$\int \exp(i\tau Y) \varphi(\tau) d\tau = s(Y).$$

Moreover,

$$\begin{aligned}
 \int \psi_{\theta_0}(\tau) \varphi(\tau) d\tau &= \frac{1}{2\pi} \int \psi_{\theta_0}(\tau) \left(\int e^{-i\tau Y} s(Y) dY \right) d\tau \\
 &= \int \left[\frac{1}{2\pi} \int e^{-i\tau Y} \psi_{\theta_0}(\tau) d\tau \right] s(Y) dY \\
 &= \int f_{\theta_0}(Y) s(Y) dY \\
 &= \mathbb{E}^{\theta_0}(s(Y)) \\
 &= 0.
 \end{aligned}$$

Hence, condition (4.1) holds.

The kernel k of the covariance operator is given by

$$\begin{aligned}
 k(\tau, \lambda) &= \mathbb{E}^{\theta_0} [h(\tau, \theta_0) \overline{h(\lambda, \theta_0)}] \\
 &= \mathbb{E}^{\theta_0} \left[\left(e^{i\tau Y} - \psi_{\theta_0}(\tau) \right) \left(e^{-i\lambda Y} - \psi_{\theta_0}(-\lambda) \right) \right] \\
 &= \psi_{\theta_0}(\tau - \lambda) - \psi_{\theta_0}(\tau) \psi_{\theta_0}(-\lambda).
 \end{aligned}$$

Note that $|k(\tau, \tau)| \leq 2$ and hence Assumption 2.2 is satisfied provided π is a probability measure. Moreover, Equations (2.7) and (2.8) hold by Carrasco, Chernov, Florens, and Ghysels (2007). Therefore, Proposition 4.1 implies that the GMM estimator based on the full continuum of moments (4.5) is asymptotically efficient.

(3) Joint and conditional characteristic function of a Markov process

Now Y_t is a Markov time series of order p . Consider moment functions based on the joint characteristic function of Y_t^{p+1} :

$$h(\tau, Y_t^{p+1}; \theta) = e^{\tau' Y_t^{p+1}} - \mathbb{E}^{\theta} \left[e^{\tau' Y_t^{p+1}} \right] \tag{4.6}$$

or the conditional characteristic function

$$h(\tau, Y_t^{p+1}; \theta) = e^{i s' Y_{t-1}^p} \left[e^{i r Y_t} - \mathbb{E}^{\theta} \left[e^{i r Y_t} | Y_{t-1}^p \right] \right], \tag{4.7}$$

where $\tau = (s', r)'$. Using again the Fourier inversion formula, we can show that the score belongs to the span of either moment condition. Moreover, Assumption 2.2

is satisfied provided π is a probability measure (because the kernel k of K is bounded). Again, equations (2.7) and (2.8) hold by Carrasco, Chernov, Florens, and Ghysels (2007). Proposition 4.1 implies that GMM estimators based on either moment function are asymptotically efficient. This result was shown by Singleton (2001) for (4.7) using a different proof but is new for (4.6).⁸ Note that the first set of moment conditions are autocorrelated unless the data are i.i.d. On the other hand, the moment conditions (4.7) are m.d.s. and hence are uncorrelated. Therefore the second set of moment conditions may be easier to handle than the first one. If the conditional characteristic function is unknown, it may be difficult to estimate it via simulations, as it would require drawing from the conditional distribution, which is not always possible. On the other hand, simulations of the joint characteristic function are easy to perform as discussed in Carrasco, Chernov, Florens, and Ghysels (2007). Applications of the characteristic function to the estimation of diffusions are given by Singleton (2001) and Carrasco, Chernov, Florens, and Ghysels (2007), among others.

Example 2 (Efficient method of moments)

Now assume that $\{Y_t\}$ is a stationary weakly dependent process. The indirect inference method proposed by Gourieroux, Monfort, and Renault (1993) and the efficient method of moments (EMM) proposed by Gallant and Tauchen (1996) are methods designed to handle cases for which the MLE is intractable; these methods use an auxiliary model (easy to estimate) indexed by a parameter λ . The auxiliary model defines a conditional pdf $f_M(Y_t|Y_{t-1}, \dots, Y_{t-p}; \lambda)$ where M denotes the dimension of λ . Gallant and Tauchen (1996) suggest using a seminonparametric model based on a Hermite expansion as auxiliary model. Gallant and Long (1997) show that the variance of the EMM estimator converges to the Cramér–Rao efficiency bound when M and p go to infinity (see also Gallant and Nychka, 1987). Calzolari, Fiorentini, and Sentana (2004) give a condition for the efficiency of the constrained indirect inference estimator in the case where the (finitely many) moment functions are not necessarily m.d.s.s. Our results permit establishing the asymptotic efficiency of EMM in a more straightforward way. First, we can handle an infinite number of moments

$$h_t(\theta_0) = \frac{\partial \ln f}{\partial \lambda} (Y_t|Y_{t-1}, \dots, Y_{t-p}; \lambda(\theta_0))$$

(where the dimension of λ here is infinite) using appropriate π and regularization as in Carrasco (2012). Then, Proposition 4.1 implies that efficiency is achieved if the DGP score belongs to the closure of pseudo scores, which is true for the seminonparametric auxiliary model suggested by Gallant and Tauchen (1996).

5. SEMIPARAMETRIC EFFICIENCY BOUND FOR CONDITIONAL RESTRICTIONS

In an i.i.d. setting, Chamberlain (1987) shows that the asymptotic variance of the GMM estimator provides an efficiency bound in a much broader sense; i.e., it is the

lower bound for the asymptotic variance of any consistent, asymptotically normal estimator in a class of nonparametric models, where the only restriction imposed on the distribution is a set of moment conditions. In this section, we extend this result to a time-series context. Here, we do not assume a specific parametric model. The only information we have is the following set of assumptions.

Assumption 5.1. The observations are given by $(Y_{-p+1}, Y_{-p+2}, \dots, Y_{T+1})$ where $Y_t \in \mathbb{R}$ is a stationary ergodic Markov chain of order p . Let $X_t = (Y_t, \dots, Y_{t-p+1})$ be the lagged values of Y_t . The transition distribution from X_t to Y_{t+1} is unknown and denoted by $Q(x, dy)$. Here X_t has a nondegenerate invariant distribution denoted μ .

Assumption 5.2. The model is now characterized by a conditional moment restriction: $\mathbb{E}(\rho(X_t, Y_{t+1}; \theta_0) | X_t) = 0$.

Assumption 5.3. $\rho : \mathbb{R}^{p+1} \rightarrow \mathbb{R}^m$ is a continuous function of X_t, Y_{t+1} and is twice continuously differentiable with respect to θ . Denote by $\nabla_{\theta} \rho(X_t, Y_{t+1}; \theta_0)$ the $q \times m$ matrix with (i, j) element $\partial \rho_j(X_t, Y_{t+1}; \theta) / \partial \theta_i$ evaluated at $\theta = \theta_0$, where $\rho_j(X_t, Y_{t+1}; \theta)$ is the j th element of $\rho(X_t, Y_{t+1}; \theta)$. Here $\nabla_{\theta} \rho(X_t, Y_{t+1}; \theta_0)$ is bounded away from zero almost surely. Moreover, I_0 defined by

$$I_0 \equiv \mathbb{E} \left[\mathbb{E}(\nabla_{\theta} \rho(X_t, Y_{t+1}; \theta_0) | X_t) \mathbb{E}(\rho(X_t, Y_{t+1}; \theta_0) \rho(X_t, Y_{t+1}; \theta_0)' | X_t)^{-1} \times \mathbb{E}(\nabla_{\theta} \rho(X_t, Y_{t+1}; \theta_0) | X_t) \right]$$

exists and is positive definite.

First, we recall the efficiency bound in the class of GMM estimators, and we explain how to construct an estimator that reaches this bound. Second, we show that this bound coincides with the semiparametric efficiency bound.

5.1. Bound in the Class of GMM Estimators

In this section, we are interested in the estimators of θ with minimal variance in the class of GMM estimators. The GMM estimator with minimal variance will be called optimal. It is known from Godambe (1985) and Hansen, Heaton, and Ogaki (1988, p. 867) that the optimal GMM estimator is obtained by solving the exactly identified system of equations

$$\sum_{t=1}^T h^*(X_t, Y_{t+1}; \theta) = 0, \tag{5.1}$$

where

$$h^*(X_t, Y_{t+1}; \theta) = A^*(X_t) \rho(X_t, Y_{t+1}; \theta)$$

and the optimal instrument satisfies

$$A^*(X_t) = \mathbb{E}(\nabla_{\theta} \rho(X_t, Y_{t+1}; \theta_0) | X_t) \mathbb{E}(\rho(X_t, Y_{t+1}; \theta_0) \rho(X_t, Y_{t+1}; \theta_0)' | X_t)^{-1}. \tag{5.2}$$

Then, the asymptotic variance is given by

$$\begin{aligned}
 & \left[\mathbb{E} \left(h^* (X_t, Y_{t+1}; \theta_0) h^* (X_t, Y_{t+1}; \theta_0)' \right) \right]^{-1} \\
 &= \mathbb{E} \left[\mathbb{E} (\nabla_{\theta} \rho (X_t, Y_{t+1}; \theta_0) | X_t) \mathbb{E} (\rho (X_t, Y_{t+1}; \theta_0) \rho (X_t, Y_{t+1}; \theta_0)' | X_t)^{-1} \right. \\
 &\quad \left. \times \mathbb{E} (\nabla_{\theta} \rho (X_t, Y_{t+1}; \theta_0) | X_t) \right]^{-1} \\
 &= I_0^{-1}. \tag{5.3}
 \end{aligned}$$

The unknown A^* could be replaced by a consistent nonparametric estimator to obtain a feasible estimator of θ . The resulting estimator also would be efficient (in the i.i.d. case, see Newey, 1993; for a dynamic example, see Wefelmeyer, 1996). However, estimating the conditional expectations in (5.2) may be burdensome.

We show that this bound can be attained by using a set of instruments, \mathcal{A} , that is dense in $L^2 (X_t)$, the space of functions of X_t that are square integrable.

PROPOSITION 5.1. *Assume Assumptions 5.1–5.3 hold. Let \mathcal{A} be a set of instruments $\{A (\tau, X_t) : \tau \in I\}$ in $L^2 (X_t)$ and such that the moment function*

$$h (\tau, X_t, Y_{t+1}; \theta) = A (\tau, X_t) \rho (X_t, Y_{t+1}; \theta), \quad \tau \in I \tag{5.4}$$

satisfies Assumption 2.2. If the closed linear span of \mathcal{A} denoted $\overline{\mathcal{A}}$ contains $L^2 (X_t)$, then the GMM estimator based on the moments (5.4) is optimal.

The proof of Proposition 5.1 is given in Appendix B. As the asymptotic variance of the GMM estimator decreases when adding more moment conditions, the efficiency bound (in the GMM class) is reached by the estimator that uses the richest class of instruments. Although the optimal instrument defined in (5.2) cannot be written as a finite linear combination of elements of \mathcal{A} , it belongs to $\overline{\mathcal{A}}$.

Now we give some examples of instruments that span $L^2 (X_t)$ (for the proof of denseness, see, e.g., Bierens, 1990; Stinchcombe and White, 1998). For each of these examples, we discuss the choice of π so that Assumption 2.2 is satisfied and hence the covariance operator K is nuclear. In particular, we need to check condition (2.5) with $k (\tau, \tau) = \mathbb{E} (h (\tau, X_t, Y_{t+1}; \theta) h (\tau, X_t, Y_{t+1}; \theta)')$.

- Consider the case where X_t has bounded support and $\mathcal{A} = \{\exp (\tau' X_t) : \tau \in \mathbb{R}^p\}$ satisfies the condition of Proposition 5.1. We compute the kernel k :

$$k (\tau, \tau) = \mathbb{E} \left[\exp (2\tau' X_t) \rho (X_t, Y_{t+1}; \theta_0) \rho (X_t, Y_{t+1}; \theta_0)' \right].$$

Provided π is a probability measure with existing moment generating function, condition (2.5) holds and K is nuclear.

- Again if X_t has bounded support, one can use power functions, $A (\tau, X_t) = X_t^\tau$, $\tau = 1, 2, \dots$. The kernel k satisfies

$$k (\tau, \tau) = \mathbb{E} \left(X_t^{2\tau} \rho (X_t, Y_{t+1}; \theta_0) \rho (X_t, Y_{t+1}; \theta_0)' \right),$$

$$\begin{aligned}
 |k (\tau, \tau)| &\leq \mathbb{E} \left(|X_t|^{2\tau} \rho (X_t, Y_{t+1}; \theta_0) \rho (X_t, Y_{t+1}; \theta_0)' \right) \\
 &\leq M^{2\tau} \mathbb{E} (\rho (X_t, Y_{t+1}; \theta_0) \rho (X_t, Y_{t+1}; \theta_0)'),
 \end{aligned}$$

where M is the upper bound of $|X_t|$. For condition (2.5) to be satisfied, we need to select π so that $\sum_{\tau=1,2,\dots} M^{2\tau} \pi(\tau) < \infty$. We can take, e.g., $\pi(\tau) = e^{-\tau}$. For this choice of π , Assumption 2.2 is satisfied, and our theory applies.

- When X_t is not bounded, one can use $A(\tau, X_t) = \exp(i\tau'X_t)$, $\tau \in \mathbb{R}^p$. Note that in this case

$$|k(\tau, \tau)| \leq \mathbb{E} [\rho(X_t, Y_{t+1}; \theta_0) \rho(X_t, Y_{t+1}; \theta_0)'].$$

It follows that condition (2.5) is satisfied provided π is a probability measure.

- Consider X_t scalar, not necessarily bounded, and indicator functions $A(\tau, X_t) = I(X_t < \tau)$, $\tau \in \mathbb{R}$. Here again condition (2.5) is satisfied for any probability measure π .
- One can choose an arbitrary countable family of functions that spans $L^2(X_t)$. This includes wavelets, splines, and sieves in general (see Chen, 2007).

5.2. Bound in the Class of Regular Estimators

In this section, we show that the bound (5.3) corresponds to the efficiency bound not just in the class of GMM estimators but also in the much larger class of regular estimators, and hence can be called a semiparametric efficiency bound. We adopt the usual approach for establishing the semiparametric efficiency as described in Bickel, Klaassen, Ritov, and Wellner (1993). It consists of constructing local parametric submodels for which the MLE is asymptotically normal. Among these submodels, we isolate the worse cast scenario leading to an MLE with maximal variance. Then using the convolution theorem, we show that this variance gives a bound. Although the concept of local asymptotic normality and the convolution theorem were initially introduced in an i.i.d. context, they have been extended to time series in a natural way; see various papers by Wefelmeyer (cited in the recent survey by Greenwood, Müller, and Wefelmeyer, 2004) and McNeney and Wellner (2000). Komunjer and Vuong (2010) derive the semiparametric efficiency bound for quantile models in time series. Instead of appealing to the convolution theorem, they construct global (instead of local) parametric models that encompass the true model and find the least favorable among these models. We could have adopted their approach, but the proof would have been very different and more cumbersome than the present one.

To the best of our knowledge, the efficiency in the model defined by Assumption 5.2 has not been considered yet. Assumption 5.2 includes as a special case the conditional moment restrictions of Wefelmeyer (1996). Wefelmeyer (1996) considers moments of the form

$$\rho(x, y; \theta) = \begin{pmatrix} y - m_\theta(x) \\ (y - m_\theta(x))^2 - v_\theta(x) \end{pmatrix},$$

where m_θ and v_θ are, respectively, the known conditional mean and variance of y . Our proof follows closely that of Theorem 1 in Wefelmeyer (1996). We adapt it to

our more general setting by allowing ρ to be nonlinear in y and the dimensions of ρ and θ to be arbitrary. Note that Assumption 5.2 can be rewritten as

$$\int \rho(x, y; \theta_0) Q(x, dy) = 0_m. \tag{5.5}$$

First we must construct a local submodel. Let \mathcal{G} be the set of bounded functions $G(x, y)$ that take their values in \mathbb{R}^q (where q is the dimension of θ) such that

$$\int G(x, y) Q(x, dy) = 0_q, \tag{5.6}$$

$$\int G(x, y) \rho(x, y; \theta_0)' Q(x, dy) = - \int \nabla_{\theta} \rho(x, y; \theta_0) Q(x, dy). \tag{5.7}$$

For $G \in \mathcal{G}$ and a constant $u \in \mathbb{R}^q$, we construct a transition distribution Q^{TuG} such that (5.5) holds for $Q = Q^{TuG}$ and $\theta = \theta_0 + T^{-1/2}u$. The distribution Q^{TuG} can be interpreted as a sequence of Pitman-type local alternatives applied to distributions.

PROPOSITION 5.2. *A local submodel at Q is given by the transition distribution*

$$Q^{TuG}(x, dy) = \left(1 + T^{-1/2} (G(x, y) + r_T(x, y))' u\right) Q(x, dy),$$

where $r_T(x, y)$ is a vector of \mathbb{R}^q of order $T^{-1/2}$ defined by equations (B.6) and (B.8) in Appendix B.

Let P_T denote the joint distribution of $(Y_{-p+1}, Y_{-p+2}, \dots, Y_{T+1})$ if Q is true and P_T^{TuG} if Q^{TuG} is true. The family P_T^{TuG} is a local model at Q . It includes the true model for $u = 0$. By a Taylor expansion around $u = 0$, we have

$$\begin{aligned} \ln \frac{dP_T^{TuG}}{dP_T} &= \sum_{t=1}^T \ln \left(1 + T^{-1/2} (G(X_t, Y_{t+1}) + r_T(X_t, Y_{t+1}))' u\right) \\ &= T^{-1/2} \sum_{t=1}^T (G_t + r_{Tt})' u - \frac{1}{2} T^{-1} \sum_{t=1}^T u' (G_t + r_{Tt}) (G_t + r_{Tt})' u \\ &\quad + o_{P_T}(T^{-1}) \\ &= T^{-1/2} \sum_{t=1}^T G_t' u - \frac{1}{2} T^{-1} u' \left(\sum_{t=1}^T G_t G_t' \right) u + o_{P_T}(1) \end{aligned}$$

using the notation $G_t = G(X_t, Y_{t+1})$ and $r_{Tt} = r_T(X_t, Y_{t+1})$. As a result of the law of large numbers and local asymptotic normality of Markov chains (Roussas, 1965), we have

$$\begin{aligned} T^{-1} \sum_{t=1}^T G_t G_t' &\xrightarrow{P} \mathbb{E}(GG') \quad \text{under } P_T, \\ T^{-1/2} \sum_{t=1}^T G_t &\xrightarrow{d} N(0, \mathbb{E}(GG')) \quad \text{under } P_T, \end{aligned}$$

where $\mathbb{E}(\cdot)$ denotes the expectation with respect to the invariant joint distribution of (X_t, Y_{t+1}) , denoted $\mu \otimes Q$. Estimating u in the local model corresponds to estimating θ in the original model. The true value of u is zero. The MLE of u in P_T^{TuG} has the property that

$$\hat{u} \xrightarrow{d} N\left(0, \mathbb{E}(GG')^{-1}\right) \quad \text{under } P_T,$$

and the least favorable submodel corresponds to the G that maximizes $\mathbb{E}(GG')^{-1}$ or equivalently minimizes $\mathbb{E}(GG')$. Let $\bar{\mathcal{G}}$ be the closure of \mathcal{G} in $L_2(\mu \otimes Q)$. The efficient score function $s \in \bar{\mathcal{G}}$ at Q is such that

$$s = \arg \min_{G \in \bar{\mathcal{G}}} \mathbb{E}(GG'). \tag{5.8}$$

The information bound at Q is $\mathbb{E}(ss')$.

Following Wefelmeyer (1996), we restrict our attention to the following class of regular estimators.

DEFINITION 5.1. An estimator $\hat{\theta}$ is called regular at Q with limit L if for all $G \in \mathcal{G}$ and $u \in \mathbb{R}^q$,

$$\sqrt{T} \left(\hat{\theta} - \theta_0 - T^{-1/2}u \right) \xrightarrow{d} L \quad \text{under } P_T^{TuG},$$

where L does not depend on G and u .

By the convolution theorem (Hajek, 1970), we have $L = M * N$ where M is independent of N and N is a normal with mean zero and variance $\mathbb{E}(ss')^{-1}$. Hence, for any bowl-shaped loss function (see Van der Vaart, 1998, p. 113), an estimator of θ_0 is efficient if its limit under P_T is N . This permits us to establish the following result.

PROPOSITION 5.3. Consider the class of regular estimators defined in Definition 5.1. Let A^* be as in (5.2). The efficient score at Q is given by $s(x, y) = A^*(x)\rho(x, y; \theta_0)$, and the semiparametric efficiency bound is positive and equals (5.3).

Proposition 5.3 states that the optimal GMM estimator based on moments (5.1) is asymptotically efficient in the class of regular estimators. The difference between Proposition 5.1 and Proposition 5.3 is that the first one gives the efficiency bound within the class of GMM estimators whereas the second one gives the efficiency bound within the much larger class of regular estimators. The proof of Proposition 5.3 (given in Appendix B) consists of showing that $s \in \bar{\mathcal{G}}$ and s satisfies (5.8). The semiparametric efficiency bound coincides with the GMM bound (5.3) for the following reason. According to Corollary 3.1, a GMM estimator based on moments $A\rho$ has an asymptotic variance equal to the inverse of $\mathbb{E}(GG')$ where G satisfies $\mathbb{E}(G) = 0$ and

$$\mathbb{E}(A\nabla_{\theta}\rho) = \mathbb{E}(GA\rho). \tag{5.9}$$

Note that if A is complete, then (5.9) is equivalent to $\mathbb{E}(\nabla_{\theta}\rho|x) = \mathbb{E}(G\rho|x)$. So for the choice of A described in Proposition 5.1, we have equivalence between (5.9) and (5.7), and the asymptotic variances coincide.

Moreover, the semiparametric efficiency bound is attainable by GMM based on the moments discussed in Proposition 5.1.

6. CONCLUSION

We started this paper by characterizing the asymptotic variance of GMM estimators in terms of an inner product in a RKHS. Then, we used this characterization to derive two important results on efficiency:

1. We gave a necessary and sufficient condition for the GMM estimator to be as efficient as MLE.
2. We derived the semiparametric efficiency bound in a class of regular estimators where the only knowledge of the process is that (a) it is Markov, (b) it satisfies a conditional moment condition. It turns out that this bound coincides with the GMM bound as it is true for i.i.d. data.

This paper sheds new light on how to reach efficiency. One way is to use the true score (in parametric models, Sect. 4) or the efficient score (in semiparametric models, Sect. 5). The other way is to use a set of an infinity (countable or not) of moment restrictions that is sufficiently rich to encompass the score or efficient score. When the score is unknown or intractable, the only way for GMM to be as efficient as MLE is to use an infinity of moment conditions, which boils down to nonparametrically estimating the unknown score. The same is true for semiparametric models; the econometricians concerned with efficiency usually nonparametrically estimate the optimal instrument (Newey, 1993). A less well-known way to reach the same efficiency is to use an infinity of moment restrictions, e.g., a continuum of moments based on the exponential function. Although the second approach does not require any nonparametric procedure, it still requires the introduction of a smoothing parameter to be able to invert the infinitely dimensional covariance operator.

Recently, there has been increasing interest in alternative methods to GMM based on entropy, such as the continuous updating estimator (Hansen, Heaton, and Yaron, 1996), the empirical likelihood estimator (Imbens, 1997), and the exponential tilting estimator (Kitamura and Stutzer, 1997). As the asymptotic variance of these estimators is the same as that of GMM, the efficiency results we gave here also apply to these estimators.

NOTES

1. Note that when the set is infinite dimensional, the closure of the set may contain the score even if the individual moment conditions are very different from the score.
2. But Feuerverger does not give the rate at which the grid should become finer.

3. This framework covers the case $Y_t = (Z_t, Z_{t-1}, \dots, Z_{t-l+1})'$ where Z_t is a univariate time series.
4. See Brezis (2002, Thm. 4.13).
5. The space $L^2(F_0)$ coincides with $L^2(Y)$, the Hilbert space of square integrable functions with respect to F_0 only for the uncorrelated case.
6. See Kuo (1975).
7. If $Z_j = Z_{t-j}$ are the lagged values of some random variable Z , the observations $Y_t = (X_t, Z_{t-1}, Z_{t-2}, \dots)'$ are infinite dimensional, and this example does not satisfy our assumption $Y_t \in \mathbb{R}^l$. A modification of our approach to handle infinite-dimensional random variables would be necessary. On the other hand, if $Z_j = Z^j$ for some scalar random variable Z , then $Y_t = (X_t, Z_t)' \in \mathbb{R}^2$ and we can handle this case by setting $\tau = j \in \mathbb{N}$.
8. Feuerverger (1990) claims the result, but his proof is heuristic.

REFERENCES

- Bates, C. & H. White (1993) Determination of estimators with minimum asymptotic covariance matrices. *Econometric Theory* 9, 633–648.
- Berlinet, A. & C. Thomas-Agnan (2004) *Reproducing Kernel Hilbert Spaces in Probability & Statistics*. Kluwer Academic Publishers.
- Bickel, P., C. Klaassen, Y. Ritov, & J. Wellner (1993) *Efficient and Adaptive Estimation for Semiparametric Models*. Springer-Verlag.
- Bierens, H. (1990) A consistent conditional moment test of functional form. *Econometrica* 58, 1443–1458.
- Billingsley, P. (1995) *Probability and Measure*. Wiley.
- Böttcher, A., A. Dijkstra, H. Langer, M. Dritschel, J. Rovnyak, & M. Kaashoek (1996) *Lectures on Operator Theory and Its Applications*. Fields Institute Monographs. American Mathematical Society.
- Brezis, H. (2002) *Analyse Fonctionnelle: Theorie et Applications*. Dunod.
- Calzolari, G., G. Fiorentini, & E. Sentana (2004) Constrained indirect estimation. *Review of Economic Studies* 71, 945–973.
- Carrasco, M. (2012) A regularization approach to the many instruments problem. *Journal of Econometrics* 170, 383–398.
- Carrasco, M., M. Chernov, J.P. Florens, & E. Ghysels (2007) Efficient estimation of general dynamic models with a continuum of moment conditions. *Journal of Econometrics* 140, 529–573.
- Carrasco, M. & J.P. Florens (2002) Simulation based method of moments & efficiency. *Journal of Business & Economic Statistics* 20, 482–492.
- Carrasco, M., J.P. Florens, & E. Renault (2007) Linear inverse problems in structural econometrics: Estimation based on spectral decomposition and regularization. In J.J. Heckman & E.E. Leamer (eds.), *Handbook of Econometrics*, vol. 6B, pp. 5633–575. Elsevier.
- Chamberlain, G. (1987) Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics* 34, 305–334.
- Chen, X. (2007) Large sample sieve estimation of semi-nonparametric models. In J.J. Heckman & E.E. Leamer (eds.), *Handbook of Econometrics*, vol. 6B, pp. 5549–5632. Elsevier.
- Chen, X. & H. White (1998) Central limit and functional central limit theorems for Hilbert space-valued dependent processes. *Econometric Theory* 14, 260–284.
- Davidson, R. (2000) Efficiency and robustness in a geometrical perspective. In P. Marriott & M. Salmon (eds.), *Applications of Differential Geometry to Econometrics*, pp. 151–183. Cambridge University Press.
- Debnath, L. & P. Mikusinski (2005) *Introduction to Hilbert Spaces with Applications*. Academic Press.
- Donald, S. & W. Newey (2001) Choosing the number of instruments. *Econometrica* 69, 1161–1191.

- Durbin, J. (1960) Estimation of parameters in time-series regression models. *Journal of the Royal Statistical Society, Series B* 22, 139–153.
- Engl, H., M. Hanke, & A. Neubauer (2000) *Regularization of Inverse Problems*. Kluwer Academic Publishers.
- Feuerverger, A. (1990) An efficiency result for the empirical characteristic function in stationary time-series models. *Canadian Journal of Statistics* 18, 155–161.
- Fisher, R.A. (1922) On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London, Series A* 222, 309–368.
- Gallant, A.R. & J.R. Long (1997) Estimating stochastic differential equations efficiently by minimum chi-squared. *Biometrika* 84, 125–141.
- Gallant, A.R. & D. Nychka (1987) Semi-nonparametric maximum likelihood estimation, *Econometrica* 55, 363–390.
- Gallant, A.R. & G. Tauchen (1996) Which moments to match? *Econometric Theory* 12, 657–681.
- Godambe, V.P. (1960) An optimum property of regular maximum likelihood estimation. *Annals of Mathematical Statistics* 31, 1208–1212.
- Godambe, V.P. (1985) The foundations of finite sample estimation in stochastic processes. *Biometrika* 72, 419–428.
- Gourieroux, C., A. Monfort, & E. Renault (1993) Indirect inference. *Journal of Applied Econometrics* 8, S85–S118.
- Greenwood, P., U. Müller, & W. Wefelmeyer (2004) An introduction to efficient estimation for semiparametric time series. In M.S. Mikulin, N. Balakrishnan, M. Mesbak, & N. Limnios (eds.), *Parametric & Semiparametric Models with Applications to Reliability, Survival Analysis, and Quality of Life*, pp. 253–272. Birkhauser.
- Hajek, J. (1970) A characterization of limiting distributions of regular estimates. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 14, 323–330.
- Hall, A., A. Inoue, K. Jana, & C. Shin (2007) Information in generalized method of moments estimation and entropy-based moment selection. *Journal of Econometrics* 138, 488–512.
- Hansen, L. (1982) Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029–1054.
- Hansen, L. (1985) A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators. *Journal of Econometrics* 30, 203–238.
- Hansen, L. (1993) Semiparametric efficiency bounds for linear time-series models. In Peter C.B. Phillips (ed.), *Models, Methods, and Applications of Econometrics: Essays in Honor of A.R. Bergstrom*, pp. 253–271. Blackwell.
- Hansen, L., J. Heaton & M. Ogaki (1988) Efficiency bounds implied by multiperiod conditional moment restrictions. *Journal of the American Statistical Association* 83, 863–871.
- Hansen, L., J. Heaton, & A. Yaron (1996) Finite-sample properties of some alternative GMM estimators. *Journal of Business & Economic Statistics* 14, 262–280.
- Imbens, G.W. (1997) One-step estimators for over-identified generalized method of moment models. *Review of Economic Studies* 64, 359–383.
- Kannan, D. & A.T. Bharucha-Reid (1970) Note on covariance operators of probability measures on a Hilbert space. *Proceedings of the Japan Academy* 46, 124–129.
- Kitamura, Y. & M. Stutzer (1997) An information-theoretic alternative to generalized method of moments estimation. *Econometrica* 65, 861–874.
- Komunjer, I. & Q. Vuong (2010) Semiparametric efficiency bound in time-series models for conditional quantiles. *Econometric Theory* 26, 383–405.
- Kress, R. (1999) *Linear Integral Equations*. Springer-Verlag.
- Kuersteiner, G. (2001) Optimal instrumental variables estimation for ARMA models. *Journal of Econometrics* 104, 359–405.
- Kuo, H.-H. (1975) *Gaussian Measures in Banach Spaces*. Springer-Verlag.
- McNeney, B. & J. Wellner (2000) Application of convolution theorems in semiparametric models with non-i.i.d. data. *Journal of Statistical Planning and Inference* 91, 441–480.

Newey, W. (1993) Efficient estimation of models with conditional moment restrictions. In G.S. Maddala, C.R. Rao, & H.D. Vinod (eds.), *Handbook of Statistics*, vol. 11, pp. 419–454. Elsevier.

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

Parzen, E. (1959) Statistical Inference on time series by Hilbert space methods, I. Technical report 23, Applied Mathematics & Statistics Laboratory, Stanford. Reprinted in *Time Series Analysis Papers* (1967) Holden-Day.

Parzen, E. (1970) Statistical inference on time series by RKHS methods. In: R. Pyke (ed.) *12th Biennial Seminar Canadian Mathematical Congress Proceedings*, pp. 1-37. Canadian Mathematical Society.

Pearson, K. (1894) Contribution to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London*, Series A 185, 71–110.

Roussas, G.G. (1965) Asymptotic inference in markov processes. *Annals of Mathematical Statistics* 36, 987–992.

Saitoh, S. (1997) *Integral Transforms, Reproducing Kernels and Their Applications*. Longman.

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

Stinchcombe, M. & H. White (1998) Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory* 14, 295–325.

Stout, W. (1974) *Almost Sure Convergence*. Academic Press.

Tauchen, G. (1997) New Minimum Chi-Square Methods in Empirical Finance. In: D. Kreps & K. Wallis, (eds.) *Advances in Econometrics, Seventh World Congress*, pp. 279–317. Cambridge University Press.

van der Vaart, A.W. (1998) *Asymptotic Statistics*. Cambridge University Press.

Wefelmeyer, W. (1996) Quasi-likelihood models and optimal inference. *Annals of Statistics* 24, 405–422.

APPENDIX A: Useful Definitions and Results on Operators

For a review on operators and reproducing kernel Hilbert space (RKHS), see Saitoh (1997), Berlinet and Thomas-Agnan (2004), and Carrasco, Florens, and Renault (2007). Here, we recall only the basic results needed in this paper.

DEFINITION A.1. A function $k : I \times I \rightarrow \mathbb{C}$ is called a positive type function if $\sum_{i=1}^n \sum_{j=1}^n a_i \bar{a}_j k(\tau_i, \tau_j) \geq 0$ for any integer n , $(a_1, \dots, a_n) \in \mathbb{C}^n$ and $(\tau_1, \dots, \tau_n) \in I^n$.

There are various definitions of the RKHS associated with k .

DEFINITION A.2. Let $k : I \times I \rightarrow \mathbb{C}$ be a positive type function. Then, $\mathcal{H}(K)$, the RKHS associated with k , is defined as the Hilbert space of functions f on I with inner product $\langle \cdot, \cdot \rangle_K$ that satisfies the following properties:

- (i) $k(\cdot, \tau) \in \mathcal{H}(K)$,
- (ii) (reproducing property) $f(\tau) = \langle f, k(\cdot, \tau) \rangle_K$.

Remark 1. If k is a covariance kernel then it is necessarily of a positive type.

Remark 2. The reproducing property implies that $\{k(\cdot, \tau) : \tau \in I\}$ spans $\mathcal{H}(K)$. Indeed $\langle f, k(\cdot, \tau) \rangle_K = 0$ for all $\tau \in I \Rightarrow f = 0$.

Remark 3. The definition A.2 has been written without reference to an integral operator and for that matter to a measure π . It means that $\mathcal{H}(K)$ is invariant to the choice of π introduced in Section 2 provided the support I remains the same.

In the GMM setting, we use the RKHS machinery in connection with a covariance operator. Consider an integral operator K defined as

$$K : L^2(I, \pi) \rightarrow L^2(I, \pi),$$

$$f \rightarrow g(\tau) = \int_I k(\tau, \lambda) f(\lambda) \pi(d\lambda), \tag{A.1}$$

The operator K is self-adjoint if $\langle Kf, g \rangle = \langle f, Kg \rangle$, which is satisfied if $k(\tau, \lambda) = \overline{k(\lambda, \tau)}$. An operator K on $L^2(I, \pi)$ is said to be nuclear (or trace-class) if for any complete orthonormal system $\{e_j\}$ of $L^2(I, \pi)$, $\sum_j \langle Ke_j, e_j \rangle < \infty$. The integral operator K defined in (A.1) is nuclear if

$$\int k(\tau, \tau) \pi(d\tau) < \infty.$$

A nuclear operator is compact and hence has a countable spectrum; moreover, the sum of its eigenvalues is finite. A nuclear operator is Hilbert–Schmidt (for a definition of Hilbert–Schmidt operators see Carrasco, Florens, and Renault, 2007). Let $\{\phi_j, \mu_j\}$, $j = 1, 2, \dots$ be the eigenfunctions and eigenvalues of K . An alternative characterization of the RKHS in terms of the spectral decomposition of K is as follows.

DEFINITION A.3. *The RKHS associated with a self-adjoint Hilbert–Schmidt operator K on $L^2(I, \pi)$ is defined as the Hilbert space*

$$\mathcal{H}(K) = \left\{ f \in L^2(I, \pi) \mid \sum_j \frac{|\langle f, \phi_j \rangle|^2}{\mu_j} < \infty \right\}$$

with inner product

$$\langle f, g \rangle_K = \sum_j \frac{\langle f, \phi_j \rangle \overline{\langle g, \phi_j \rangle}}{\mu_j},$$

where the convention $1/0 = 0$ is adopted if some of the μ_j are equal to zero.

Remark 4. A covariance operator is necessarily self-adjoint.

Remark 5. $\mathcal{H}(K)$ coincides with the range of $K^{1/2}$. Moreover, if 0 is not an eigenvalue of K , then

$$\langle f, g \rangle_K = \left\langle K^{-1/2} f, K^{-1/2} g \right\rangle.$$

Remark 6. The eigenvalues μ_j and eigenfunctions ϕ_j depend on the measure π . In spite of this, both Definitions A.2 and A.3 define the same space and the same inner product, provided k is square integrable (This is satisfied if K is the covariance operator of h and h satisfies Assumption 2.2.) with respect to π . A representation of the inner product in $\mathcal{H}(K)$ that is independent of the arbitrary measure π is given in Proposition 3.1.

APPENDIX B: Proofs

Proof of Proposition 3.1.

Proof of (i). It suffices to show that there exists a correspondence J that satisfies the conditions. We are going to construct such a correspondence. As mentioned in Appendix A, the reproducing property implies that $\mathcal{H}(K)$ is spanned by $\{k(\cdot, \tau) : \tau \in I\}$. Hence, each element in $\mathcal{H}(K)$ is the limit in $L^2(I, \pi)$ of functions g_n such that

$$g_n = \sum_{j=1}^n \alpha_j k(\cdot, \tau_j)$$

for some integer n , real constants $\alpha_1, \dots, \alpha_n$, and points τ_1, \dots, τ_n in I . It is useful to introduce the notation $L_n(k(\cdot, \tau) : \tau \in I)$ for the space of all elements g_n , i.e., the space of all finite linear combinations of elements $k(\cdot, \tau)$, $\tau \in I$.

On the other hand, each element of $L^2(h)$ is the limit in $L^2(F_0)$ of functions G_n such that

$$G_n = \sum_{j=1}^n \beta_j h(\tau_j)$$

for some integer n , real β_1, \dots, β_n , and points τ_1, \dots, τ_n in I . Again, it is convenient to define $L_n(h(\tau) : \tau \in I)$ as the space of all G_n , the space of all finite linear combinations of elements $h(\tau)$, $\tau \in I$.

We construct the correspondence J between $L_n(k(\cdot, \tau) : \tau \in I)$ and $L_n(h(\tau) : \tau \in I)$ as follows:

$$J(k(\cdot, \tau_j)) = h(\tau_j),$$

$$J\left(\sum_{j=1}^n \alpha_j k(\cdot, \tau_j)\right) = \sum_{j=1}^n \alpha_j h(\tau_j)$$

for any integer n , real $\alpha_1, \dots, \alpha_n$, and points τ_1, \dots, τ_n in I . We want to show that this correspondence is one-to-one and linear and preserves the inner product.

(a) J is one-to-one:

Let $g_n = \sum_{j=1}^n \alpha_j k(\cdot, \tau_j)$ and $J(g_n) = \sum_{j=1}^n \alpha_j h(\tau_j)$. We need to show that two different representations of an element

$$g_n = \sum_{j=1}^n \alpha_j k(\cdot, \tau_j) = \sum_{j=1}^n \alpha'_j k(\cdot, \tau'_j)$$

lead to the same value

$$J(g_n) = \sum_{j=1}^n \alpha_j h(\tau_j) = \sum_{j=1}^n \alpha'_j h(\tau'_j).$$

Denote

$$\sum_{i=1}^n \beta_i k(\cdot, \tau_i) \equiv \sum_{j=1}^n \alpha_j k(\cdot, \tau_j) - \sum_{j=1}^n \alpha'_j k(\cdot, \tau'_j).$$

Then it suffices to show that

$$\sum_{i=1}^n \beta_i k(\cdot, \tau_i) = 0 \Rightarrow \sum_{i=1}^n \beta_i h(\tau_i) = 0.$$

It follows from the fact that

$$\begin{aligned} 0 &= \left\| \sum_{i=1}^n \beta_i k(\cdot, \tau_i) \right\|_K^2 \\ &= \left\langle \sum_{i=1}^n \beta_i k(\cdot, \tau_i), \sum_{l=1}^n \beta_l k(\cdot, \tau_l) \right\rangle_K \\ &= \sum_{i,l} \beta_i \beta_l \langle k(\cdot, \tau_i), k(\cdot, \tau_l) \rangle_K \\ &= \sum_{i,l} \beta_i \beta_l k(\tau_i, \tau_l) \end{aligned}$$

by the reproducing property. Moreover, $k(\tau_i, \tau_l) = (h(\tau_i), h(\tau_l))_{L^2(F_0)}$ by the definition of k . Therefore,

$$\begin{aligned} 0 &= \sum_{i,l} \beta_i \beta_l k(\tau_i, \tau_l) \\ &= \sum_{i,l} \beta_i \beta_l \langle h(\tau_i), h(\tau_l) \rangle_{L^2(F_0)} \\ &= \left\| \sum_{i=1}^n \beta_i h(\tau_i) \right\|_{L^2(F_0)}^2, \end{aligned}$$

where the last equality follows from Assumption 2.2 and Fubini’s theorem. It implies that $\sum_{i=1}^n \beta_i h(\tau_i) = 0$.

(b) J is linear:

Let a be an arbitrary constant, $g_{1n} = \sum_{j=1}^n \alpha_{1j} k(\cdot, \tau_j)$ and $g_{2m} = \sum_{j=1}^m \alpha_{2j} k(\cdot, \tau_j)$. We have

$$\begin{aligned} J(ag_{1n} + g_{2m}) &= J\left(\sum_{j=1}^{\max(m,n)} (a\alpha_{1j} + \alpha_{2j}) k(\cdot, \tau_j)\right) \\ &= \sum_{j=1}^{\max(m,n)} (a\alpha_{1j} + \alpha_{2j}) h(\tau_j) \\ &= aJ(g_{1n}) + J(g_{2m}), \end{aligned}$$

where $\alpha_{1j} = 0$ for $j > n$ and $\alpha_{2j} = 0$ for $j > m$.

(c) J preserves the inner product:

Let g_{1n} and g_{2n} be defined as in (b) and $G_{1n} = \sum_{j=1}^n \alpha_{1j} h(\tau_j)$ and $G_{2n} = \sum_{j=1}^n \alpha_{2j} h(\tau_j)$. We have

$$\begin{aligned} \langle g_{1n}, g_{2n} \rangle_K &= \left\langle \sum_{j=1}^n \alpha_{1j} k(\cdot, \tau_j), \sum_{l=1}^n \alpha_{2l} k(\cdot, \tau_l) \right\rangle_K \\ &= \sum_{j,l} \alpha_{1j} \alpha_{2l} \langle k(\cdot, \tau_j), k(\cdot, \tau_l) \rangle_K \\ &= \sum_{j,l} \alpha_{1j} \alpha_{2l} k(\tau_j, \tau_l) \\ &= \sum_{j,l} \alpha_{1j} \alpha_{2l} (h(\tau_j), h(\tau_l))_{L^2(F_0)} \\ &= \langle G_{1n}, G_{2n} \rangle_{L^2(F_0)}. \end{aligned}$$

Hence, we proved that J is an isomorphism between $L_n(k(\cdot, \tau) : \tau \in I)$ and $L_n(h(\tau) : \tau \in I)$. Now, we proceed to extend J to $\mathcal{H}(K) = \overline{U_n L_n(k(\cdot, \tau) : \tau \in I)}$ (the closure of the union of all $L_n(k(\cdot, \tau) : \tau \in I)$). For $g = \lim_{n \rightarrow \infty} g_n$, define $J(g) = \lim_{n \rightarrow \infty} J(g_n)$. This extension is an isomorphism between $\mathcal{H}(K)$ and $L^2(h)$.

Proof of (ii). From the proof of (a), it is immediate to see that $G \in \mathcal{G}_g \cap L^2(h)$. Now we proceed to prove the uniqueness of G . Let \tilde{G} be another element of $\mathcal{G}_g \cap L^2(h)$. We have $g(\tau) = \langle G, h(\tau) \rangle_{L^2(F_0)} = \langle \tilde{G}, h(\tau) \rangle_{L^2(F_0)}$ for all $\tau \in I$. Hence, $\langle G - \tilde{G}, h(\tau) \rangle_{L^2(F_0)} = 0$ for all $\tau \in I \Rightarrow G - \tilde{G} \in L^2(h)^\perp \Rightarrow G - \tilde{G} = 0$. This proves uniqueness.

Proof of (iii). First note that an element of \mathcal{G}_g can not be orthogonal to $h(\tau)$ for all τ . Define a general element of \mathcal{G}_g , $G = G_0 + G_1$ where $G_0 = \sum_l \omega_l h(\tau_l)$ belongs to $\mathcal{G}_g \cap L^2(h)$. Here $G \in \mathcal{G}_g$ implies

$$\langle G_1, h(\tau) \rangle_{L^2(F_0)} = 0, \quad \forall \tau. \tag{B.1}$$

Next we compute the norm of G and show that it is minimal for $G_1 = 0$.

$$\|G\|_{L^2(F_0)}^2 = \|G_0\|_{L^2(F_0)}^2 + \|G_1\|_{L^2(F_0)}^2 + 2\langle G_0, G_1 \rangle_{L^2(F_0)}.$$

The condition (B.1) implies that

$$\begin{aligned} \sum_l \omega_l \langle G_1, h(\tau_l) \rangle_{L^2(F_0)} &= 0 \\ \Rightarrow \langle G_0, G_1 \rangle_{L^2(F_0)} &= 0. \end{aligned}$$

Hence

$$\|G\|_{L^2(F_0)}^2 = \|G_0\|_{L^2(F_0)}^2 + \|G_1\|_{L^2(F_0)}^2 \geq \|G_0\|_{L^2(F_0)}^2.$$

(c) \Leftrightarrow (b) follows because $G = G_0$ has minimal norm. ■

Proof of Corollary 3.1. The first statement follows immediately from Proposition 3.1 and equation (2.8). To prove the second statement, let us write $\tilde{G}_i = G_{i0} + G_{i1}$ where

$G_{i0} \in L^2(h)$ and $G_{i1} \in L^2(h)^\perp, i = 1, 2, \dots, q$. Here $\tilde{G}_i \in \mathcal{G}_i \Rightarrow G_{i0} \in \mathcal{G}_i \cap L^2(h)$. From Proposition 3.1, we know that G_{i0} is unique, and therefore $\Sigma_{ij} = \langle G_{i0}, G_{j0} \rangle_{L^2(F_0)}$.

$$\begin{aligned} \tilde{\Sigma}_{ij} &= \langle G_{i0} + G_{i1}, G_{j0} + G_{j1} \rangle_{L^2(F_0)} \\ &= \langle G_{i0}, G_{j0} \rangle_{L^2(F_0)} + \langle G_{i1}, G_{j1} \rangle_{L^2(F_0)} \\ &= \Sigma_{ij} + \langle G_{i1}, G_{j1} \rangle_{L^2(F_0)} \end{aligned}$$

as the cross-products are zero. It follows that $\tilde{\Sigma}_{ij} - \Sigma_{ij} \geq 0$ from the fact that the matrix with elements $\langle G_{i1}, G_{j1} \rangle_{L^2(F_0)}$ is nonnegative definite. ■

Proof of Proposition 4.1.

(a) By Corollary 3.1, we need to find $G_i \in \mathcal{G}_i \cap L^2(h)$. The proof of (a) proceeds in two steps. First, show that $s_{ti} = \nabla_{\theta_i} \ln f_{\theta}(Y_t | Y_{t-1}^p) |_{\theta=\theta_0}$ belongs to the set \mathcal{G}_i . Second, show that $G_i - s_{ti} \in L^2(h)^\perp$.

We consider the general case where $\{h(\tau, Y_t^{p+1})\}$ is autocorrelated, as the case where $\{h(\tau, Y_t^{p+1})\}$ is a m.d.s. is a special case. The covariance operator has for kernel

$$\begin{aligned} k(\tau_1, \tau_2) &= \sum_{j=-\infty}^{\infty} \mathbb{E}^{F_0} \left[h(\tau_1, Y_t^{p+1}) h(\tau_2, Y_{t-j}^{p+1}) \right] \\ &= \sum_{j=-\infty}^{\infty} \mathbb{E}^{F_0} \left[h(\tau_1, S^j X) h(\tau_2, X) \right] \\ &\equiv \langle h, h \rangle_{L^2(F_0)}, \end{aligned}$$

where $X_t = Y_t^{p+1}$ and S is a measure preserving shift operator as defined in Stout (1974, p. 171). To show that s_{ti} belongs to \mathcal{G}_i , we need to establish the following conditions. This equality appears in Newey and McFadden, 1994, in the independent case. It has been proved for the dynamic case by Hall et al., 2007, independently to us. Hall et al. assume that X_t is α -mixing instead of being Markov.

$$\begin{aligned} -\mathbb{E}^\theta \left[\nabla_{\theta_i} h(\tau, Y_0^{p+1}, \theta) \right] &= \langle s, h \rangle_{L^2(F_0)} \\ &= \sum_{j=-\infty}^{\infty} \mathbb{E}^\theta \left[\nabla_{\theta_i} \ln f_{\theta}(Y_j | Y_{j-1}, Y_{j-2}, \dots, Y_{j-p}) h(\tau, Y_0^{p+1}) \right] \\ &= \sum_{j=-\infty}^0 \mathbb{E}^\theta \left[\nabla_{\theta_i} \ln f_{\theta}(Y_j | Y_{j-1}, Y_{j-2}, \dots, Y_{j-p}) h(\tau, Y_0^{p+1}) \right] \\ &\equiv \sum_{j=-\infty}^0 \gamma(j). \end{aligned}$$

By Assumption 4.1(b) and the fact that s is a m.d.s., s belongs to $L^2(F_0)$. Because s and h belong to $L^2(F_0)$, we have by Cauchy-Schwarz $\langle s, h \rangle_{L^2(F_0)} < \infty$, and hence the infinite sum $\sum_{j=-\infty}^{\infty} \gamma(j)$ is necessarily finite. No mixing condition is needed here.

We use the fact that $\mathbb{E}^\theta \left[h \left(\tau, Y_0^{p+1}, \theta \right) \right] = 0$ for all θ . Differentiating with respect to θ_i gives

$$\begin{aligned} \nabla_{\theta_i} \mathbb{E}^\theta \left[h \left(\tau, Y_0^{p+1}, \theta \right) \right] = 0 &\iff \nabla_{\theta_i} \int h \left(\tau, y_0^{p+1}, \theta \right) \\ &\times f_\theta \left(y_0, y_{-1}, \dots, y_{-T+1}, y_{-T} \right) dy_0^{T+1} = 0. \end{aligned}$$

Interchanging the order of differentiation and integration as allowed by Assumption 4.1(c), we obtain

$$\begin{aligned} \int \nabla_{\theta_i} h \left(\tau, y_0^{p+1}, \theta \right) f_\theta \left(y_0^{T+1} \right) dy_0^{T+1} + \int h \left(\tau, y_0^{p+1}, \theta \right) \nabla_{\theta_i} f_\theta \left(y_0^{T+1} \right) dy_0^{T+1} &= 0 \\ \iff \mathbb{E}^\theta \left(\nabla_{\theta_i} h \left(\tau, Y_0^{p+1}, \theta \right) \right) \\ + \int h \left(\tau, y_0^{p+1}, \theta \right) \nabla_{\theta_i} \ln f_\theta \left(y_0^{T+1} \right) f_\theta \left(y_0^{T+1} \right) dy_0^{T+1} &= 0. \end{aligned}$$

Using $\ln f_\theta \left(y_0, y_{-1}, \dots, y_{-T+1}, y_{-T} \right) = \sum_{j=-T}^0 \ln f_\theta \left(y_j | y_{j-1}, \dots, y_{-T} \right)$, the second equality can be rewritten in the following way:

$$\begin{aligned} -\mathbb{E}^\theta \left(\nabla_{\theta_i} h \left(\tau, Y_0^{p+1}, \theta \right) \right) \\ = \sum_{j=-T}^0 \int h \left(\tau, y_0^{p+1}, \theta \right) \nabla_{\theta_i} \ln f_\theta \left(y_j | y_{j-1}, \dots, y_{-T} \right) f_\theta \left(y_0^{T+1} \right) dy_0^{T+1} \end{aligned} \tag{B.2}$$

$$\xrightarrow{T \rightarrow \infty} \sum_{j=-\infty}^0 \gamma \left(j \right). \tag{B.3}$$

This proves $s_{ii} \in \mathcal{G}_i$.

Let G_i be the unique element of $\mathcal{G}_i \cap L^2(h)$. We have

$$\begin{aligned} \langle G_i - s_{ii}, h(\tau) \rangle_{L^2(F_0)} &= \langle G_i, h(\tau) \rangle_{L^2(F_0)} - \langle s_{ii}, h(\tau) \rangle_{L^2(F_0)} \\ &= g(\tau) - g(\tau) \\ &= 0, \quad \text{for all } \tau \in I. \end{aligned}$$

It follows that $G_i - s_{ii} \in L^2(h)^\perp$. Hence G_i is the orthogonal projection of s_{ii} on $L^2(h)$.

(b) The GMM estimator reaches the Cramer–Rao efficiency bound if the inverse of its asymptotic variance equals the information matrix, i.e., if the (i, j) element of Σ satisfies

$$\Sigma_{ij} = \mathbb{E}^\theta \left[s_{ii} s'_{ij} \right] \Big|_{\theta=\theta_0}.$$

A necessary and sufficient condition is that $G_i = s_{ii}$, which is equivalent to $s_{ii} \in L^2(h)$. ■

Proof of Proposition 5.1. We apply Corollary 3.1 to compute the inverse of the asymptotic variance of the GMM estimator $\hat{\theta}$ based on the moments described in Proposition 5.1: $\Sigma = \mathbb{E} \left(GG' \right)$ where G is the unique element of $\mathcal{G} \cap L^2(h)$; $G \in \mathcal{G}$ requires that

$$\begin{aligned} -\mathbb{E} \left(\nabla_{\theta} h \left(\tau, X_t, Y_{t+1} \right) \right) &= \mathbb{E} \left(Gh \left(\tau, X_t, Y_{t+1} \right) \right) \quad \text{for all } \tau \in I, \\ \iff -\mathbb{E} \left(A \left(X_t, \tau \right) \nabla_{\theta} \rho \left(X_t, Y_{t+1} \right) \right) &= \mathbb{E} \left(GA \left(X_t, \tau \right) \rho \left(X_t, Y_{t+1} \right) \right) \quad \forall \tau \in I, \end{aligned} \tag{B.4}$$

where the dependence in θ_0 is omitted to simplify notation. As \mathcal{A} is complete in $L^2(X_t)$, equation (B.4) is equivalent to

$$-\mathbb{E}(\nabla_{\theta} \rho(X_t, Y_{t+1}) | X_t) = \mathbb{E}(G \rho(X_t, Y_{t+1}) | X_t). \tag{B.5}$$

We see that equation (B.5) is satisfied for $G = h^* = A^*(X_t) \rho(X_t, Y_{t+1})$ where A^* is defined in (5.2). It remains to check that h^* belongs to $L^2(h)$, which is true again because $\bar{\mathcal{A}}$ contains $L^2(X_t)$. It follows from Corollary 3.1 that the asymptotic variance of $\hat{\theta}$ is $\mathbb{E}(h^* h^{*\prime})$, which coincides with that of the optimal GMM. ■

Proof of Proposition 5.2. We look for r_T such that (5.5) holds. Let $\bar{\rho}(x, y)$ be an m -vector:

$$\begin{aligned} \bar{\rho}(x, y) &= \rho(x, y; \theta_0) I \left[|\rho(x, y; \theta_0)| \leq T^{1/4} \right] \\ &\quad - \int \rho(x, y; \theta_0) I \left[|\rho(x, y; \theta_0)| \leq T^{1/4} \right] Q(x, dy) \end{aligned}$$

and

$$r_T(x, y) = a(x) \bar{\rho}(x, y), \tag{B.6}$$

where $a(x)$ is a $q \times m$ matrix that will be characterized later on.

We see that r_T has conditional mean equal to zero. Choose θ_{Tu} between θ_0 and $\theta_0 + T^{-1/2}u$ such that

$$\rho(x, y; \theta_0 + T^{-1/2}u) = \rho(x, y; \theta_0) + T^{-1/2} \nabla_{\theta} \rho(x, y; \theta_{Tu})' u. \tag{B.7}$$

Condition (5.5) for $Q = Q^{TuG}$ and $\theta = \theta_0 + T^{-1/2}u$ reads

$$\int \rho(x, y; \theta_0 + T^{-1/2}u) Q^{TuG}(x, dy) = 0.$$

Replacing ρ by its expansion (B.7) and Q^{TuG} by its expression, we obtain

$$\begin{aligned} T^{-1/2} \int [\nabla_{\theta} \rho(x, y; \theta_{Tu}) - \nabla_{\theta} \rho(x, y; \theta_0)]' u Q(x, dy) \\ + T^{-1/2} \int \rho(x, y; \theta_0) r_T(x, y)' u Q(x, dy) \\ + T^{-1} \int \nabla_{\theta} \rho(x, y; \theta_{Tu})' u (G(x, y) \\ + r_T(x, y))' u Q(x, dy) \\ = 0. \end{aligned}$$

Simplifying by $T^{-1/2}$ and u , it is enough to find r_T that satisfies

$$\begin{aligned} \int [\nabla_{\theta} \rho(x, y; \theta_{Tu}) - \nabla_{\theta} \rho(x, y; \theta_0)] Q(x, dy) \\ + \int r_T(x, y) \rho(x, y; \theta_0)' Q(x, dy) \\ + T^{-1/2} \int (G(x, y) \\ + r_T(x, y)) u' \nabla_{\theta} \rho(x, y; \theta_{Tu}) Q(x, dy) \\ = 0. \end{aligned}$$

Then, replacing $r_T(x, y)$ by $a(x)\bar{p}(x, y)$, we obtain

$$\begin{aligned} & \int [\nabla_{\theta}\rho(x, y; \theta_{Tu}) - \nabla_{\theta}\rho(x, y; \theta_0)] Q(x, dy) + \int a(x)\bar{p}(x, y)\rho(x, y; \theta_0)' Q(x, dy) \\ & \quad + T^{-1/2} \int (G(x, y) + a(x)\bar{p}(x, y))u' \nabla_{\theta}\rho(x, y; \theta_{Tu}) Q(x, dy) \\ & = 0. \end{aligned}$$

It suffices to select $a(x)$ such that

$$\begin{aligned} a(x) = - \left\{ \int [\nabla_{\theta}\rho(x, y; \theta_{Tu}) - \nabla_{\theta}\rho(x, y; \theta_0)] Q(x, dy) \right. \\ \left. + T^{-1/2} \int G(x, y)u' \nabla_{\theta}\rho(x, y; \theta_{Tu}) Q(x, dy) \right\} \\ \times \left[\int \bar{p}(x, y) \left(\rho(x, y; \theta_0)' + T^{-1/2}u' \nabla_{\theta}\rho(x, y; \theta_{Tu}) \right) Q(x, dy) \right]^{-1}. \end{aligned} \tag{B.8}$$

The term $D = \int \bar{p}(x, y) \left(\rho(x, y; \theta_0)' + T^{-1/2}u' \nabla_{\theta}\rho(x, y; \theta_{Tu}) \right) Q(x, dy)$ is bounded away from zero by Assumption 5.3.. It follows that $a(x) = O(T^{-1/2})$ and that Q^{TuG} satisfies (5.5). Moreover it is immediate to check that Q^{TuG} is positive for T large enough and integrates to one. Therefore, it defines a valid local submodel. ■

Proof of Proposition 5.3. We need to show that $s \in \bar{\mathcal{G}}$ and s satisfies (5.8).

We have

$$\int s(x, y) Q(x, dy) = A^*(x) \int \rho(x, y; \theta_0) Q(x, dy) = 0$$

and

$$\begin{aligned} & \int s(x, y)\rho(x, y; \theta_0)' Q(x, dy) \\ & = A^*(x) \int \rho(x, y; \theta_0)\rho(x, y; \theta_0)' Q(x, dy) \\ & = \int \nabla_{\theta}\rho(x, y; \theta_0) Q(x, dy) \left[\int \rho(x, y; \theta_0)\rho(x, y; \theta_0)' Q(x, dy) \right]^{-1} \\ & \quad \times \int \rho(x, y; \theta_0)\rho(x, y; \theta_0)' Q(x, dy) \\ & = \int \nabla_{\theta}\rho(x, y; \theta_0) Q(x, dy). \end{aligned}$$

Hence, $s \in \mathcal{G}$. Moreover, by Assumption 5.3., $\mathbb{E}(ss')$ is well defined and positive, and hence $s \in L_2(\mu \otimes Q)$.

It remains to prove that s minimizes $\mathbb{E}(GG')$. Consider another element G of $\bar{\mathcal{G}}$ and denote $\tilde{G} = G - s$. Element G satisfies $\mathbb{E}(G|x) = 0$, so it follows that $\mathbb{E}(\tilde{G}|x) = 0$. Similarly, we have $\mathbb{E}(G\rho|x) = \mathbb{E}(\nabla_{\theta}\rho|x)$, which implies that $\mathbb{E}(\tilde{G}\rho|x) = 0$. From the expression of s , it follows that $\mathbb{E}(\tilde{G}s') = \mathbb{E}(s\tilde{G}') = 0$. Now consider

$$\begin{aligned}\mathbb{E}(GG') &= \mathbb{E}\left(\left(\tilde{G} + s\right)\left(\tilde{G} + s\right)'\right) = \mathbb{E}\left(\tilde{G}\tilde{G}'\right) + \mathbb{E}(ss') + \mathbb{E}\left(\tilde{G}s'\right) + \mathbb{E}\left(s\tilde{G}'\right) \\ &= \mathbb{E}\left(\tilde{G}\tilde{G}'\right) + \mathbb{E}(ss') \geq \mathbb{E}(ss').\end{aligned}$$

Hence, $\mathbb{E}(GG')$ is minimal for $G = s$. Finally, note that $s = h^*$ and $\mathbb{E}(ss') = I_0$. This concludes the proof. ■