Generalization of GMM to a continuum of moment conditions[*]

Marine Carrasco
CREST and The Ohio State University

Jean-Pierre Florens
GREMAQ-IDEI, Université des Sciences Sociales de Toulouse

September 16, 1999

Abstract

This paper proposes a version of the Generalized Method of Moments procedure that handles both the case where the number of moment conditions is finite and the case where there is a continuum of moment conditions. Typically, the moment conditions are indexed by an index parameter that takes its values in an interval. The objective function to minimize is then the norm of the moment conditions in a Hilbert space. The estimator is shown to be consistent and asymptotically normal. The optimal estimator is obtained by minimizing the norm of the moment conditions in the reproducing kernel Hilbert space associated with the covariance. We show an easy way to calculate this estimator. Finally, we study properties of a specification test using overidentifying restrictions. Results of this paper are useful in many instances where a continuum of moment conditions arise. Examples include efficient estimation of continuous time regression models, cross sectional models that satisfy conditional moment restrictions, as well as scalar diffusion processes.

[*] We acknowledge helpful discussions with Xiaohong Chen, Percy Deift, Christian Gouriéroux, Lars Hansen, Alberto Holly, Guido Imbens, Pierre-François Koehl, Nour Meddahi, Marcelo Navarro, Costin Protopopescu, Eric Renault, and Christine Thomas-Agnan, as well as Donald Andrews and three anonymous referees. We thank the participants of "Jeunes Economètres" Meeting 1995, the Econometric Society Winter Meetings 1996 and ESEM97, as well as the participants of seminars at the University of Chicago, Cornell, Princeton, Laval, UCLA, UCSB where earlier versions of this paper have been presented. Address correspondence to: Marine Carrasco, Crest, Bureau 2109, Timbre J360, 15 boulevard Gabriel Péri, 92245 Malakoff cedex, France. Email: carrasco@ensae.fr.

1 Introduction

In his seminal paper, Hansen (1982) has extended the method of moments to overidentified models, i.e. models in which the number of moment conditions is greater than the number of parameters. This method is now very popular and its properties are well established (see Hall (1993) or Ogaki (1993) for a survey).

The objective of this paper is to consider the extension of the Generalized Method of Moments (GMM) procedure to the case of a continuum of moment conditions together with a finite dimensional parameter. We will first examine the most general case where the moment conditions are characterized by a relation

$$E^{P_0}\left(h\left(X,\theta_0\right)\right)=0 \tag{1}$$

where X is a random element, generated by the probability P_0, θ_0 is the true value of a vector of parameters and h is a function valued in a (finite or infinite dimensional) Hilbert space. Let $(x^1, ..., x^n)$ be a sample of X (i.i.d. or with some dynamic dependence). The expectation in (1) is replaced in the estimation procedure by the empirical mean but the overidentification requires that Equation (1) become a minimization

$$\hat{\theta}_n = \arg\min_{\theta} \left\| B_n \left(\frac{1}{n} \sum_{i=1}^{n} h\left(x^i, \theta\right) \right) \right\| \tag{2}$$

where the norm is determined in the Hilbert space and B_n converges to a linear operator B. Asymptotic properties of the estimator derived from (2) are given in Section 2.

The question of optimal GMM estimation, i.e. of the optimal choice of B, is the main topic of this paper. This problem is addressed in a more specific case where the Hilbert space is the set of square integrable functions of $t \in [0, T]$. In other words, (1) is replaced by

$$E^{P_0}\left(h_t\left(X,\theta_0\right)\right)=0 \qquad \forall t \in [0,T] \tag{3}$$

where h_t is a real valued function.

First, we are going to recall the usual definition of the GMM estimator. If $T = 1$, a discretization of the interval $[0, 1]$ at the points $t = \frac{1}{m}, \frac{2}{m}, \cdots, 1$ yields m moment conditions. The $h_t(x^i)$, $t = \frac{1}{m}, \frac{2}{m}, \cdots, 1$ are stacked into a m-vector $h(x^i)$. For a given random, positive definite symmetric $m \times m$ matrix A_n, the GMM estimators associated with A_n are the solutions to the problem:

$$\hat{\theta}_n = \arg\min_{\theta} [\bar{h}_n(\theta)]' A_n [\bar{h}_n(\theta)]$$

where $\bar{h}_n$ is the m-vector with j-th element $\bar{h}_{\frac{j}{m}}(\theta) = \frac{1}{n}\sum_{i=1}^{n} h_{\frac{j}{m}}(x^i, \theta)$.

Now assume that the full continuum of moment is available. The empirical counterpart of (2) defines the following GMM estimator

$$\hat{\theta}_n = \arg\min_{\theta} \int_0^T \int_0^T \bar{h}_t(\theta)\, a_n(t,s) \bar{h}_s(\theta)\, dt\, ds \tag{4}$$

where $\bar{h}_t(\theta) = \frac{1}{n}\sum_{i=1}^{n} h_t(x^i, \theta)$ and $a_n(t,s)$ converges to $a(t,s)$ characterized by:

$$\|B\varphi\|^2 = \int_0^T \int_0^T \varphi(s)\, a(t,s)\, \varphi(t)\, ds\ dt \tag{5}$$

(4) looks like the limit of the usual GMM quadratic form as the interval between observations goes to zero. The search for an optimal GMM estimator requires an analysis of the covariance operator K defined by:

$$(Kf)(t) = \int_0^T E^{P_0}(h_s h_t)\, f(s)\, ds \tag{6}$$

Section 3 considers the estimation of K obtained by the substitution of $\frac{1}{n}\sum_{i=1}^{n} h_s(x^i,\theta)\, h_t(x^i,\theta)$ for $E^{P_0}(h_s h_t)$. Let K_n be this estimator where θ has been replaced by a first stage consistent estimate. Optimal GMM estimation is based on the use of K^{-1} which is the counterpart of the inverse of the covariance matrix in the finite dimensional framework. But K is a compact operator and is not invertible on the full reference space. We have to use a regularized estimator of K^{-1}, denoted $(K_n^{\alpha_n})^{-1}$. This operator is constructed in the following way. We first estimate the n eigenvalues, $\mu_j^{(n)}$, and eigenfunctions, $\phi_j^{(n)}$, of K_n by solving the functional equation $K_n\phi = \mu\phi$. The eigenvalues $\mu_j^{(n)}$ are perturbed by the smoothing parameter $\alpha_n \in I\!R^+$ and replaced by $\frac{\left(\mu_j^{(n)}\right)^2 + \alpha_n}{\mu_j^{(n)}}$. Then, the operator $(K_n^{\alpha_n})^{-1}$ satisfies :

$$\left((K_n^{\alpha_n})^{-1} f\right)(t) = \sum_{j=1}^{n} \frac{\mu_j^{(n)}}{\left(\mu_j^{(n)}\right)^2 + \alpha_n} \left(f, \phi_j^{(n)}\right) \phi_j^{(n)}(t)$$

where (.,.) is the usual inner product between functions defined on $[0, T]$. Definitions and properties of this estimator are studied in Section 4.

The optimal GMM estimator satisfies the following condition

$$\hat{\theta}_n = \arg\min_{\theta} \sum_{j=1}^{n} \frac{\mu_j^{(n)}}{\left(\mu_j^{(n)}\right)^2 + \alpha_n} \left(\phi_j^{(n)}, \bar{h}(\theta)\right)^2$$

Section 5 establishes its consistency and $\sqrt{n}-$asymptotic normality.

Section 6 extends to the continuous case Hansen's test for overidentifying restrictions. We give an interpretation of the speed of convergence of this test in terms of an implicit number of principal components used in finite samples.

A series of examples are analyzed in Section 7. These examples are oriented towards three basic types of results. The first question is the relation between optimal GMM and maximum likelihood estimation. It is natural that the efficiency gap between these two procedures vanishes when the number of moment conditions increases and this property is verified in some i.i.d. models, in counting processes and in some dynamic regression models. This suggests that GMM is an interesting alternative to MLE when one does not want to make distributional assumptions, for instance. Second, we consider a class of examples for which continuous GMM provides an efficient estimation method: models for which conditional moment restrictions are satisfied and scalar diffusion models. When the efficient instrument is unknown, one way to approach Chamberlain's efficiency bound is to use an infinity of moments. This paper develops the tools to implement such an approach. The third type of problem is related to tests of conditional moment restrictions. Methods suggested in this paper permit one to construct specification tests that have power against any fixed alternative to the null hypothesis. However, these tests will not have power against $1/\sqrt{n}$ local alternatives.

As illustrated by these examples continuous GMM estimation covers both cases in which t represents the time index (inference on stochastic processes observed continuously) and cases in which t is a more general index of moment conditions.

In Section 8, some concluding remarks are made. The basic definitions and properties of operators are recalled in Appendix A. Proofs are in Appendix B.

2 Consistency and asymptotic distribution of the GMM estimator

The results of this section are not restricted to a particular indexation of the moment conditions and hold under fairly general conditions. Let X be a random element (r.e.) defined on a complete probability space $(\Omega, \mathcal{F}, P_0)$ that takes its values in $(S, \mathcal{S})$. Let H be an Hilbert space with the inner product (.,.) that defines a norm $\| \, . \, \|$.

Assumption 1: The observed data $\{x^1, ..., x^n\}$ are independent realizations of the stochastic process X.

Note that independence is not crucial and we shall discuss how to relax it later on.

Assumption 2: Let h be a function on $S \times \Theta$ that takes its values in H where Θ is a compact subset of $I\!R^q$. h is a continuous function of θ.

Assumption 3: h is integrable with respect to P_0 for any θ and the equation

$$E^{P_0}(h(X, \theta)) = 0$$

has a unique solution θ_0 which is an interior point of Θ.

Assumption 4: Let B be a nonrandom bounded linear operator defined on $\mathcal{D}(B) \subset H$ valued in H. B does not depend on θ but may depend on θ_0. $E^{P_0}(h(X, \theta)) \in \mathcal{D}(B)$, $\forall \theta$.

Assumption 5: Let $N(B)$ denote the null space of B, $N(B) = \{f \in H | Bf = 0\}$. We assume that $E^{P_0}(h(X, \theta)) \in N(B)$ implies $E^{P_0}(h(X, \theta)) = 0$.

Remark 1. Assumption 5 is an identification condition implied in particular by the condition $N(B) = \{0\}$. In the finite dimensional case, this condition reduces to a full rank assumption on the weighting matrix $B'B$ and is therefore natural. In the general case and as illustrated in the following examples, $N(B) = \{0\}$ is rarely satisfied and hence is replaced by Assumption 5.

The following examples are meant to illustrate Assumption 5. Assume that $h_t(X_t, \theta) = X_t - \theta F(t)$ for any given differentiable function F, $E^{P_0}(h_t) = (\theta_0 - \theta)F(t)$. First consider the operator $B : (Bf)(t) = t \int_0^T f(s) \, ds$. B is a bounded linear operator and $N(B) = \left\{ f | \int_0^T f(s) \, ds = 0 \right\}$. Assumption 5 is satisfied $\Leftrightarrow \int_0^T F(s) \, ds \neq 0$. Consider now B a differential operator $Bf = \frac{df}{dt}(.)$. B is a linear operator that is not bounded but the example is useful since the optimal choice of B (discussed in Section 5) is not bounded. $N(B)$ is the set of constant functions. Then, Assumption 5 is satisfied $\Leftrightarrow F(t)$ is not a constant function.

Assumption 6: Let B_n be a sequence of random bounded linear operators. $B_n : \mathcal{D}(B_n) \subset H \rightarrow H$. Let $\bar{h}_n(\theta) = \frac{1}{n} \Sigma h(x^i, \theta)$. We assume that $\bar{h}_n(\theta) \in \mathcal{D}(B_n)$, $\forall \theta$ and that $Q_n = \parallel B_n \bar{h}_n(\theta) \parallel$ is a continuous function of θ.

Assumption 7: $Q_n \rightarrow Q = \parallel B E^{P_0}(h(X, \theta)) \parallel$ almost surely uniformly on Θ.

Definition 1 *The (continuous) GMM estimators $\hat{\theta}_n$ associated with B_n are defined by*

$$\hat{\theta}_n = \arg\min_{\theta} Q_n$$

Theorem 1 *Under Assumptions 1 to 7, the GMM estimator associated with B_n converges to θ_0 almost surely.*

Proof. The result follows from Theorem 3.4 of White (1994). ■

This framework encompasses at the same time GMM with a finite number of moment conditions and with a continuum of moment conditions.

i) In the case of J real moment conditions, H is taken equal to $I\!R^J$ provided with the usual Hilbert space structure so that B and B_n are $J \times J$ matrices.

ii) If the structural model specifies J moment conditions at m dates $\{t_1, ...t_m\}$, H becomes $I\!R^{Jm}$ and the analysis reduces to the previous case.

iii) In the case of a univariate moment condition indexed by $t \in [0, T]$, H is now equal to the Hilbert space of square integrable functions with respect to a given measure which can be chosen equal to the Lebesgue measure. Let $L^2[0, T]$ be this space, B_n (and B) are linear operators and $A_n = B_n^* B_n$ (where B_n^* is the adjoint of B_n) is defined through a kernel $a_n(t, s) : (A_n f)(t) = \int_0^T a_n(t, s) f(s) ds$.

iv) Finally in presence of J moment conditions indexed by $t \in [0, T]$, H will be taken equal to $(L^2[0, T])^J$. An element of H is a vector $(f_j(t))_{j=1,...,J}$ of square integrable functions. Let M be a real-valued positive definite symmetric $J \times J$-matrix with principal element m_{jk}. We define a norm by

$$\|f\| = \left[\int_0^T f'(t) M f(t) dt \right]^{1/2} = \left[\sum_{j,k=1,...,J} m_{jk} \int_0^T f_j(t) f_k(t) dt \right]^{1/2}.$$

Notice that $\|f\| \in I\!R$ and that the usual L^2-norm corresponds to a choice of $M = I_J$, the identity matrix. Assume that B is an integral operator satisfying

$$(Bf)(t) = \left(\sum_{l=1}^{J} \int_0^T b^{jl}(t, s) f_l(s) ds \right)_{j=1,...,J}.$$

Hence we have

$$\begin{aligned}\|Bf\|^2 &= \sum_{j,k=1,\ldots,J} m_{jk} \int_0^T \left(\sum_{l=1}^J \int_0^T b^{jl}(t,s) f_l(s)\, ds \right) \left(\sum_{l'=1}^J \int_0^T b^{kl'}(t,u) f_{l'}(u)\, du \right) dt \\ &= \sum_{j,k=1,\ldots,J} m_{jk} \int_0^T a^{jk}(s,u) f_j(s) f_k(u)\, ds du\end{aligned}$$

with

$$a^{jk}(s,u) = \sum_{l,l'=1,\ldots,J} \int_0^T b^{jl}(t,s)\, b^{kl'}(t,u)\, dt.$$

Equivalently $\|Bf\|^2 = (Bf, Bf) = (f, B^* Bf) = (f, Af)$ where

$$(Af)(t) = \left(\sum_{l=1}^J \int_0^1 a^{jl}(t,s) f_l(s)\, ds \right)_{j=1,\ldots,J}.$$

In order to obtain the asymptotic distribution of our estimator, it is necessary to add some extra assumptions. First we define some notation. Let $f = (f_1, ..., f_p)$ and $g = (g_1, ..., g_q)$ be elements of H^p and H^q respectively. We denote by (f, g) the $p \times q$ matrix of principal elements $(f_j, g_k)\,(j = 1, ..., p,\ k = 1, ..., q)$. Using this notation, (f, f) is a $p \times p$ matrix which will be denoted by $\|f\|^2$.

Assumption 8: h is differentiable[1] with respect to $\theta = (\theta_j)_{j=1,\ldots,q}$, $\frac{\partial \bar{h}_n}{\partial \theta_j} \in \mathcal{D}(B_n)$, $\forall j$ and $\frac{\partial}{\partial \theta_j} E^{P_0}(h(X,\theta)) = E^{P_0}\left(\frac{\partial h}{\partial \theta_j}(X,\theta)\right) \in \mathcal{D}(B)$, $\forall j$.

Moreover the $q \times q-$ matrix $\left(BE^{P_0}\left[\frac{\partial h}{\partial \theta'}(X,\theta)\right], BE^{P_0}\left[\frac{\partial h}{\partial \theta'}(X,\theta)\right]\right) = \| BE^{P_0} \frac{\partial h}{\partial \theta'}(X,\theta) \|^2$ is positive definite and symmetric.

Assumption 9: The inner product satisfies the following differentiation rule

$$\frac{\partial}{\partial \theta'}(u(\theta), v(\theta)) = \left(\frac{\partial}{\partial \theta'} u(\theta), v(\theta) \right) + \left(u(\theta), \frac{\partial}{\partial \theta'} v(\theta) \right)$$

and B and B_n commute with the differential operator:

$$\frac{\partial}{\partial \theta'}[Bu(\theta)] = B\left[\frac{\partial}{\partial \theta'} u(\theta) \right].$$

Define $\| B \| = \sup_{\|f\| \leq 1} \| Bf \|$.

[1] We consider a function $f(\theta)$ from $\mathbb{R}^q$ to H and differentiability means Frechet differentiability. The differential in θ is a linear function from $\mathbb{R}^q$ to H which can be written $df_\theta(\lambda) = \sum_{j=1}^q \frac{\partial f}{\partial \theta_j}(\theta)\lambda_j$. Let $\frac{\partial f}{\partial \theta}$ denote the vector $\left(\frac{\partial f}{\partial \theta_j}\right)_{j=1,\ldots q}$ of elements of H.

ASSUMPTION 10: $\| B_n - B \| \to 0$ in probability.

ASSUMPTION 11: $\sqrt{n}\bar{h}_n(\theta_0) \to Y \sim \mathcal{N}(0, K)$ in distribution on H as n goes to infinity, where $\mathcal{N}$ is the Gaussian random element of H which has a zero mean and a covariance operator K. $Y \in \mathcal{D}(B)$ with probability 1.

Remark 2. Assumption 11 involves a functional convergence and is stronger than the asymptotic normality of $\sqrt{n}\left(\bar{h}_{nt_1}, ..., \bar{h}_{nt_p}\right)$ to a normal vector (where $\bar{h}_{nt} = \frac{1}{n}\sum_{i=1}^{n} h_t\left(x^i, \theta\right)$) for any finite sequence $t_1, t_2,$ ...,t_p (see Billingsley, 1968). Such a functional convergence requires a topological structure of the functional space (like Skohorod topology in the case of right continuous distribution functions) which is here the Hilbert structure (see Chen-White, 1998). Note that we directly assume asymptotic normality. An alternative approach would be to specify assumptions on the data generating process (ergodicity, mixing, ...) and on the function h in order to derive, through a functional central limit theorem, the result given in Assumption 11.

Assumptions 10 and 11 imply that $\sqrt{n}B_n\bar{h}_n(\theta_0) \to Z \sim \mathcal{N}(0, BKB^*)$. This result can be found in Corollary 5.2 of Chen-White (1992).

Remark 3. A summary of definitions and results for H-valued r.e. can be found in Chen-White (1998). The covariance operator $K : H \to H$ associated with the H-valued r.e. Y is defined as

$$Kf = E\left[\left(Y - E\left(Y\right), f\right)\left[Y - E\left(Y\right)\right]\right]$$

where $(.,.)$ corresponds to the inner product in H. In the particular case where $H = L^2\left[0, T\right]$, K satisfies

$$\left(Kf\right)\left(t\right) = \int_0^T E\left(Y_t Y_s\right) f\left(s\right) ds$$

for $f \in L^2\left[0, T\right]$. An H-valued r.e. Y has a Gaussian distribution on H if for all $f \in H$, the real valued random variable (Y, f) has a Gaussian distribution on $I\!R$.

Theorem 2 *Under Assumptions 1 to 11, the asymptotic distribution of $\hat{\theta}_n$ is given by*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{n\to\infty}{\longrightarrow} \mathcal{N}(0, V)$$

with

$$V = \left\| BE^{P_0}\frac{\partial h}{\partial \theta'}(X, \theta_0)\right\|^{-2}\left(BE^{P_0}\frac{\partial h}{\partial \theta'}(X, \theta_0), (BKB^*)BE^{P_0}\frac{\partial h}{\partial \theta'}(X, \theta_0)\right)\left\| BE^{P_0}\frac{\partial h}{\partial \theta'}(X, \theta_0)\right\|^{-2}.$$

Remark 4. If B is chosen such that BKB^* is equal to the identity operator, V reduces to $\left\|BE^{P_0}\frac{\partial h}{\partial\theta'}(X,\theta)\right\|^{-2}$. By analogy with the finite dimensional case, this should correspond to the estimator with minimal variance. This optimality result will be proved in Section 5. But, the proof of Theorem 2 will be different in that case because a normal $\mathcal{N}(0,I)$ is not well defined in a Hilbert space since the operator identity is not a nuclear operator (the sum of its eigenvalues is infinite). Intuitively, BKB^* equals the identity if B is chosen equal to $K^{-\frac{1}{2}}$. This choice is elementary in the finite dimensional case (using if necessary the generalized inverse of a matrix) but requires some care if H is a functional space.

Remark 5. The hypothesis of independence between individuals is not crucial for the proofs as long as a law of large numbers and a central limit theorem are guaranteed. Different types of dependence between individuals can be considered (see Davidson (1994)) so that we have:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}h(x^i)\rightarrow\mathcal{N}(0,\sum_{l=-\infty}^{+\infty}E^{P_0}(h(X^i)h(X^{i-l}))$$

The results of Theorems 1 and 2 will be still valid. In Theorem 2, the kernel of the covariance operator that was equal to $E^{P_o}(h_t h_s)$ becomes $\sum_{l=-\infty}^{+\infty}E^{P_0}(h_t(X^i)h_s(X^{i-l}))$ which can be considered as a function of t and s: $k(t,s)$. Then the choice of the optimal weighting function requires as before inverting the operator K: $f\rightarrow\int_0^T k(t,s)f(s)ds$. Since the covariance is usually not analytically computable, it raises the problem of its estimation. Several authors have proposed consistent positive definite estimators of covariance matrices (Newey-West (1987), Andrews (1991)) for the finite dimensional case.

3 Estimation of the covariance operator K

From now on, the reference space is $H = L^2[0,T]$ the space of all square-integrable functions defined on $[0,T]$. The basic definitions and properties of operators are given in Appendix A. In this section, we first explain how to estimate the covariance operator K, as well as its eigenvalues and eigenfunctions. Next, we give the asymptotic distribution of the estimators of the eigenvalues of K.

Consider the covariance operator with kernel:

$$k\left(t,s\right)=E^{P_0}\left(h_t\left(X,\theta\right)h_s\left(X,\theta\right)\right)\equiv E^{P_0}\left(k\left(X,t,s\right)\right).$$

Assumption 12: The covariance kernel $k(t,s)$ is an L^2 kernel.

Assumption 12 insures that K is a compact Hilbert-Schmidt operator on $L^2[0,T]$. In addition, as the kernel is symmetric, the operator is self-adjoint.

We still assume that $(x^1,...,x^n)$ is an i.i.d. sample of X. Therefore, it is natural to estimate K by K_n the integral operator

$$(K_n f)(t) = \int_0^T k_n(t,s) f(s)\, dtds$$

with kernel

$$k_n(t,s) = \frac{1}{n}\sum_{i=1}^n k\left(x^i,t,s\right) = \frac{1}{n}\sum_{i=1}^n h_t\left(x^i,\theta^0\right) h_s\left(x^i,\theta^0\right)$$

Let $\hat{\theta}_n^1$ be a $n^{1/2}-$consistent first step estimate of θ_0 such that $\hat{\theta}_n^1 - \theta_0$ is $O_p(1/\sqrt{n})$. If h_t^i denotes $h_t(x^i,\hat{\theta}_n^1)$, our estimate satisfies :

$$(K_n f)(t) = \frac{1}{n}\sum_{i=1}^n h_t^i \int_0^T h_s^i f(s)\, ds$$

The operator K_n has a degenerate kernel, and therefore, contrary to K, has a finite dimensional closed range. This space $R(K_n)$ is the space spanned by $\{h_t^i\}_{i=1,...,n}$. The number of its eigenvalues and eigenfunctions is finite and they can be computed by solving a linear system. Let $\phi^{(n)}$ and $\mu^{(n)}$ denote an eigenfunction and eigenvalue of K_n. $\phi^{(n)}$ necessarily has the form $\frac{1}{n}\sum_{i=1}^n \beta_i h_t^i$, and the equation to solve becomes:

$$\frac{1}{n}\sum_{i=1}^n h_t^i \Big[\frac{1}{n}\sum_{j=1}^n \beta_j \int_0^T h_s^j h_s^i\, ds\Big] = \mu^{(n)} \frac{1}{n}\sum_{i=1}^n \beta_i h_t^i.$$

β_i, $i = 1,...n$ and $\mu^{(n)}$ are solutions of the system of n equations:

$$\frac{1}{n}\sum_{j=1}^n \beta_j c_{ij} = \mu^{(n)}\beta_i, \quad \text{with} \quad c_{ij} = \int_0^T h_s^j h_s^i\, ds,\ i = 1,...,n.$$

The solutions $\underline{\beta} = [\beta_1,...,\beta_n]'$ and μ are the eigenvectors and eigenvalues, respectively, of the $n \times n$ matrix C of elements $\frac{1}{n}c_{ij}$. Let us denote $\{\underline{\beta}^j, \mu_j^{(n)}\}$, $j = 1,...,n$, the set of eigenvectors and eigenvalues of C. Hence, the eigenfunctions of K_n are $\phi_j^{(n)}(t) = \frac{1}{n}\underline{h_t}\underline{\beta}^j$ where $\underline{h_t} = [h_t^1, h_t^2, ..., h_t^n]$ and its eigenvalues are $\mu_j^{(n)}$. From now on, $\phi_j^{(n)}$ will denote the orthonormalized eigenfunctions associated with the eigenvalues, $\mu_j^{(n)}$, ranked in decreasing order.

In the following, we give the asymptotic distribution of the estimators of the eigenvalues assuming that $h_t(x^i,\theta_0)$ is known, that is no first step estimator is used.

Assumption 13: $E\|h\|^4 < \infty$.

Theorem 3 *Under Assumptions 1 and 13, we have*

$$\sqrt{n}\left(\mu_j^{(n)} - \mu_j\right) \to \mathcal{N}\left(0, \sigma_j^2\right)$$

with $\sigma_j^2 = var\left[(h, \phi_j)^2\right].$

When h is normal, $\sigma_j^2 = 2\mu_j^2.$ This result, in the normal case, can be found in Dauxois et al. (1982). We are now concerned with the convergence of the Hilbert-Schmidt norm, that is, the convergence of $\|K_n\|_{HS}^2 = \sum_{j=1}^{\infty}\left(\mu_j^{(n)}\right)^2 = \int_0^T\int_0^T k_n(t,s)^2\,dtds$ to $\|K\|_{HS}^2 = \sum_{j=1}^{\infty}(\mu_j)^2 = \int_0^T\int_0^T k(t,s)^2\,dtds.$

Theorem 4 *Under Assumptions 1 and 13, we have*

$$(i) \quad \|K_n - K\| = O_p\left(\frac{1}{\sqrt{n}}\right)$$

$$(ii) \ \sqrt{n}\left(\sum_{j=1}^{\infty}\left(\mu_j^{(n)}\right)^2 - \sum_{j=1}^{\infty}(\mu_j)^2\right) \to \mathcal{N}\left(0, \tau^2\right)$$

with $\tau^2 = 4var\left[\int_0^T\int_0^T k(t,s)\,k(X,t,s)\,dtds\right].$

Using a Taylor expansion, it can be shown that the $n^{1/2}-$speed of convergence and asymptotic normality obtained in Theorems 3 and 4 remain valid if θ_0 is replaced by a $n^{1/2}-$consistent first step estimate $\hat{\theta}_n^1$ provided that $k(X,t,s)$ is differentiable with respect to θ (Assumptions 2 and 8.) However, the asymptotic variances, σ_j^2 and τ^2, will be different.

4 Properties and estimation of the inverse to K

4.1 Existence of the inverse to K

The choice of the optimal estimator is related to the inverse of the covariance operator K. Inverting K is equivalent to finding the solution Φ to a **Fredholm equation of the first kind**

$$K\Phi = f \tag{7}$$

for a given $f \in L^2[0,T]$. This equation is typically an ill-posed problem in contrast to the well-posed problems. An equation is well-posed if it has a solution Φ, not more than one for each f, and this unique solution depends continuously on f. In other words, Φ is stable with respect to small changes of f. The three conditions: existence, uniqueness and stability of the solution are not satisfied in the case of a Fredholm integral equation of the first kind. This problem is addressed in detail in Groetsch (1993). The aim of this subsection is to determine the subset of $L^2[0,T]$, for which a solution to (7) exists.

Lemma 5 (Picard's criterion) *The following conditions are necessary and sufficient for a solution of (7) to exist: (i) $f \in \overline{R(K)}$, the closure of $R(K)$, and (ii) $\sum_{j=1}^{\infty} \frac{(f,\phi_j)^2}{\mu_j^2} < \infty$. Then, any function of the form*

$$\Phi = \sum_{j=1}^{\infty} \frac{(f,\phi_j)}{\mu_j}\phi_j + \varphi$$

where $\varphi \in N(K)$ is a solution of (7).

Note that $\overline{R(K)}$ is equal to $N(K)^{\perp}$ since K is self-adjoint. We see clearly that a solution exists only for a restricted class of functions f and if it exists, it is unique only if $N(K) = \{0\}$. To enlarge the class of functions for which a type of a "generalized" solution Φ exists, we consider a least squares solution.

Definition 2 *A function $\Phi \in L^2[0,T]$ is called the* **least squares solution** *of (7) for a given $f \in L^2[0,T]$ if*

$$\| K\Phi - f \| = \inf\{\| Ku - f \| : u \in L^2[0,T]\}.$$

A least squares solution exists if and only if f lies in the dense subspace $R(K) + N(K)$ of $L^2[0,T]$. Moreover, there is a unique least squares solution of smallest norm.

Definition 3 *The mapping, denoted K^{-1}, which associates with a given $f \in R(K) + N(K)$ the unique least squares solution having smallest norm, is called* **Moore-Penrose generalized inverse** *of K and satisfies*

$$K^{-1}f = \sum_{j=1}^{\infty} \frac{(f,\phi_j)}{\mu_j}\phi_j.$$

From the expression of $K^{-1}f$ above, we see that the Moore-Penrose inverse operator is not bounded (because, in general, $R(K)$ is not closed) and that the solution $K^{-1}f$ is therefore not continuous in f. Even if we have enlarged the class of possible functions f, the existence of $K^{-1}f$ is not guaranteed. While, in discrete time, the optimal weighting matrix is always computable, here K does not admit a generalized inverse over the entire Hilbert space $L^2[0,T]$. To illustrate this point, we will consider a specific compact integral operator.

Example 1. Let W_t be a scalar Brownian motion on $[0,1]$. Its covariance kernel is given by $k\left(t,s\right) = \min\left(t,s\right) \equiv t \wedge s$. Consider the covariance operator associated with k

$$Kf\left(t\right) = \int_0^1 \left(t \wedge s\right) f\left(s\right) ds.$$

In this simple case, we can determine explicitly the inverse operator by solving the equation $Kg = f$ using two successive differentiations. The inverse operator to K is a second order differential operator, $Lf = -f''$, with domain $\mathcal{D}(K^{-1}) = \{f \in L^2[0,1] \mid f \text{ is twice differentiable}, f(0) = 0, f'(1) = 0\}$. The eigenvalues, μ_j, and eigenfunctions, ϕ_j, of K are solutions of $\int_0^1 (t \wedge s)\ \phi_j(s)ds = \mu_j\phi_j(t) \Leftrightarrow \int_0^t s\ \phi_j(s)ds + \int_t^1 t\ \phi_j(s)ds = \mu_j\phi_j(t)$. Using two successive differentiations, we see that ϕ_j is solution of a second order differential equation, $\phi_j(t) = -\mu_j\phi_j''(t)$, with boundary conditions $\phi_j(0) = 0$ and $\phi_j'(1) = 0$. Hence, the set of $\phi_j(t) = \sqrt{2}\sin(\frac{\pi j t}{2})$, associated with the eigenvalue, $\mu_j = \frac{4}{\pi^2 j^2}$, $j = 1, 3, 5, ...$, constitutes a set of orthonormal eigenfunctions. We can see that $f\left(t\right) = t$ does not satisfy Picard's criterion, while $f\left(t\right) = 2t - t^2$ does.

Picard's criterion is therefore rather restrictive. In some cases, we need only that f belongs to the domain of $K^{-\frac{1}{2}}$ instead of the domain of K^{-1}. The former one is larger than the latter. Following Wahba (1973) (see also Nashed and Wahba (1974)), we may define the square root $K^{-\frac{1}{2}}$ of K^{-1} by

$$K^{-\frac{1}{2}}f = \sum_{j=1}^{\infty} \frac{(f,\phi_j)}{\sqrt{\mu_j}}\phi_j$$

with the convention $0/0 = 0$. The domain of $K^{-\frac{1}{2}}$ is the set:

$$\mathcal{D}(K^{-\frac{1}{2}}) = \{f : f \in L^2[0,T], \sum_{j=1}^{\infty} \frac{(f,\phi_j)^2}{\mu_j} < \infty\}$$

Interestingly $\mathcal{D}(K^{-\frac{1}{2}})$ coincides with the reproducing kernel Hilbert space (RKHS) associated with K (see definition in Appendix A.)

Proposition 6 (Nashed-Wahba, 1974) *Let k be a nonnegative definite kernel and $\{\mu_j\}$, $\{\phi_j\}$ be the eigenvalues and orthonormalized eigenfunctions of K. Then,*

$$\mathcal{H}(K) = \{f : f \in L^2[0,T], \sum_{j=1}^{\infty} \frac{(f,\phi_j)^2}{\mu_j} < \infty\}$$

is the RKHS with kernel k. The inner product of $\mathcal{H}(K)$ is given by

$$(f,g)_K = \sum_{j=1}^{\infty} \frac{(f,\phi_j)(g,\phi_j)}{\mu_j} = \left(K^{-\frac{1}{2}}f, K^{-\frac{1}{2}}g\right) = \left(f, K^{-1}g\right)$$

for all $f, g \in \ \mathcal{H}(K)$.

We use the following notation $\| \ K^{-\frac{1}{2}}f \ \| = \| \ f \ \|_K$ where $\| \ . \ \|_K$ is the norm in $\mathcal{H}(K)$ associated with $(.,.)_K$. The domain of $K^{-\frac{1}{2}}$ is $\mathcal{H}(K)$. Following Nashed and Wahba (1974), this domain is extended to $\mathcal{H}(K) + \mathcal{H}(K)^{\perp}$ using the convention that $1/\sqrt{\mu_j}$ is equal to zero if μ_j is equal to zero.

4.2 Estimation of $K^{-\frac{1}{2}}$

Being unbounded, the operator $K^{-\frac{1}{2}}$ must be handled with caution. Indeed, the solution of the equation $Kg = f$ is not stable for small variation of f, which can have dramatic consequences since f will be estimated. To guarantee the stability of the solution, we are going to use the Tikhonov method of regularization, see Groetsch (1993). The idea is to replace the operator K by some nearby operator which has a bounded inverse. For $\alpha > 0$, the equation

$$\left(K^2 + \alpha I\right) g = Kf \tag{8}$$

has a unique solution for each $f \in L^2[0,T]$. Moreover, the solution depends continuously on f since the operator $(K^2 + \alpha I)$ has a bounded inverse. The Tikhonov approximation of the generalized inverse to K, K^{-1}, is given by

$$(K^{\alpha})^{-1} = \left(K^2 + \alpha I\right)^{-1} K.$$

Equation (8) is also known as the solution to the Ridge regression problem (see, eg., Golub, Health, and Wahba, 1979):

$$\min_g \|Kg - f\|^2 + \alpha \|g\|^2 .$$

The regularized inverse permits the following decomposition:

$$(K_n^{\alpha})^{-\frac{1}{2}} = \sum_{j=1}^{n} \frac{\sqrt{\mu_j^{(n)}}}{\sqrt{\mu_j^{(n)2} + \alpha}} \left(f, \phi_j^{(n)} \right) \phi_j^{(n)}.$$

Clearly the choice of α is crucial. If α is too large the approximate solution $(K_n^{\alpha})^{-\frac{1}{2}} f$ will be far away from $K^{-\frac{1}{2}} f$ and if α is too small the approximate solution will be unstable. Therefore, α will be allowed to converge to zero at a certain rate given as a function of the sample size n and will be denoted α_n in the following. Hence the GMM objective function to minimize is

$$\| (K_n^{\alpha_n})^{-\frac{1}{2}} \overline{h}(\theta) \|^2 = \sum_{j=1}^{n} \frac{\mu_j^{(n)}}{\mu_j^{(n)2} + \alpha_n} \left(\overline{h}(\theta), \phi_j^{(n)} \right)^2 \equiv \left\| \overline{h}(\theta) \right\|_{K_n^{\alpha_n}}^2 . \qquad (9)$$

where $\|.\|_{K_n^{\alpha_n}}^2$ denotes the norm in the RKHS associated with $K_n^{\alpha_n}$. The moment conditions h_t intervene only through their inner product with $\phi_j(t)$. In principal component analysis, $\{\phi_j\}$ are called principal components and represent orthogonal directions that summarize the information available in the moments h_t. As the μ_j are ranked in decreasing order, ϕ_1 is the most informative component, ϕ_2 is the second one, etc. The regularization parameter α_n is used to discard the least informative principal components, that is the one associated with the smallest eigenvalues. In discrete GMM, it is well-known that the use of too many moments tends to render the finite-sample performance poor. Here, we circumvent this problem by using the regularized method. An alternative approach would be to truncate the sum in (9), that is, to sum up to a number $m_n < n$ indexed by the sample size n. The truncation is used e.g. by DeJong-Bierens (1994). This point will be illustrated in Remark 12 and Example 7.4.

The following theorem gives some hints on the acceptable rates of α_n. But, it does not actually provide a rule to select α_n in practice. Many statistical papers present cross-validation methods for choosing the regularization parameter, see, e.g., Golub et al. (1979), Groetsch (1993), Hansen P.C. (1992). This issue will not be discussed further.

Theorem 7 *Consider f and f_n such that $\|f_n - f\| = O_p\left(\frac{1}{\sqrt{n}}\right)$.*

(i) Assume that $f \in \mathcal{H}(K) + \mathcal{H}(K)^{\perp}$. Then

$$\left\| (K_n^{\alpha_n})^{-\frac{1}{2}} f_n - K^{-\frac{1}{2}} f \right\| \to 0$$

in probability as n and $n\alpha_n^{\frac{3}{2}}$ go to infinity and α_n goes to zero.

(ii) Assume that $f \in \mathcal{D}(K^{-1})$. Then

$$\left\| (K_n^{\alpha_n})^{-1} f_n - K^{-1} f \right\| \to 0$$

in probability as n and $n\alpha_n^3$ go to infinity and α_n goes to zero.

5 Optimal estimator

5.1 Asymptotic results

We can show that choosing $B_n^{\alpha_n} = (K_n^{\alpha_n})^{-\frac{1}{2}}$ leads to the estimator of minimum variance. In such a case, the criterion to minimize is given by (9). Assumptions 4, 8 and 11 become:

Assumption 4': $E^{P_0}(h(X,\theta)) \in \mathcal{H}(K) + \mathcal{H}(K)^{\perp}$ for any $\theta \in \Theta$.

Assumption 8': $h(x,\theta)$ is differentiable with respect to $\theta = (\theta_1, ..., \theta_q)$ and $E^{P_0}\left(\frac{\partial h(X,\theta)}{\partial \theta_j}\right) = \frac{\partial}{\partial \theta_j} E^{P_0}\left(h\left(X,\theta\right)\right) \in \mathcal{D}(K^{-1})$ for any $\theta \in \Theta$.

Moreover the matrix $\left(K^{-1/2}E^{P_0}\left[\frac{\partial h}{\partial \theta'}(X,\theta)\right], K^{-1/2}E^{P_0}\left[\frac{\partial h}{\partial \theta'}(X,\theta)\right]\right)$ is positive definite and symmetric.

Remark 6. Since $(K_n^{\alpha_n})^{-1}$ has closed range, $\frac{\partial \bar{h}_n}{\partial \theta}$ belongs necessarily to $\mathcal{D}\left((K_n^{\alpha_n})^{-1}\right)$.

Assumption 11': $\sqrt{n}\overline{h_n}(\theta_0)$ converges in law to Y as n goes to infinity, where $Y \sim \mathcal{N}(0,K)$ in $L^2[0,T]$.

Assumption 14: $\left\| \bar{h}_n(\theta) - E^{P_0} h(\theta) \right\| = O_p\left(\frac{1}{\sqrt{n}}\right)$ uniformly in θ on Θ.
$\left\| \frac{\partial \bar{h}_n}{\partial \theta}(\theta) - E^{P_0}\frac{\partial h}{\partial \theta}(\theta) \right\| = O_p\left(\frac{1}{\sqrt{n}}\right)$ uniformly in θ on Θ.

Theorem 7 and Assumption 14 imply

$$\left\| (K_n^{\alpha_n})^{-\frac{1}{2}} \bar{h}_n(\theta) - K^{-\frac{1}{2}} E^{P_0}(h(\theta)) \right\| \to 0 \text{ in probability,}$$

uniformly in θ, as n and $n\alpha_n^{\frac{3}{2}}$ go to infinity and α_n goes to zero.

$$\left\| (K_n^{\alpha_n})^{-1} \frac{\partial \bar{h}_n}{\partial \theta'}(\theta) - K^{-1} E^{P_0} \frac{\partial h}{\partial \theta'}(\theta) \right\| \to 0 \text{ in probability,}$$

uniformly in θ, as n and $n\alpha_n^3$ go to infinity and α_n goes to zero. These are the counterparts to Assumptions 7 and 10.

Theorem 8 *Under Assumptions 1, 2, 3, 4', 5, 6, 8', 9, 11', 12, 13, and 14, the estimator*

$$\hat{\theta}_n = \arg\min_{\theta} \left\| \bar{h}(\theta) \right\|^2_{K_n^{\alpha_n}}$$

is the optimal estimator. It satisfies

$$\hat{\theta}_n \to \theta_0 \quad \text{in probability,}$$

as n and $n\alpha_n^{\frac{3}{2}}$ go to infinity and α_n goes to zero and

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{n\to\infty}{\longrightarrow} \mathcal{N}\left(0, \left\| E^{P_0}\left(\frac{\partial h}{\partial \theta'}\right)\right\|_K^{-2}\right)$$

as n and $n\alpha_n^3$ go to infinity and α_n goes to zero.

Remark 7. This result is analogous to that obtained in the discrete case where the optimal weighting matrix is the inverse of the covariance matrix. But whereas the generalized inverse of a matrix always exists, it is not the same for an operator. For consistency of the optimal estimator, we need that $E^{P_0}h$ belongs to $\mathcal{H}(K) + \mathcal{H}(K)^{\perp}$ and for asymptotic normality, we need that $E^{P_0}\frac{\partial h(X,\theta)}{\partial \theta_j}$ belongs to $\mathcal{D}(K^{-1})$. This should restrict the field of possible applications. Nevertheless, Assumptions 4' and 8' are not as restrictive as one might think, see Subsection 7.1.

Remark 8. There are important contributions on the problem of estimating a parameter vector whose dimension increases with the sample size. This gives rise to an infinity of moment conditions. Huber (1973), Portnoy (1985), and more recently Koenker-Machado (1997) study the general linear model

$$y_i = x_i'\beta + u_i,\ i = 1, 2, ..., n. \tag{10}$$

where u_i are i.i.d. with mean zero, x_i and β are p_n-dimensional vectors. The condition $E\left[x_i\left(y_i - x_i'\beta\right)\right] = 0$ defines p_n moment conditions. The basic question concerns the asymptotic behavior of the OLS estimator $\hat{\beta}$ of β. This problem is characterized by two facts: (i) the dimension, p_n, of β increases to infinity, (ii) β can be identified only from the p_n moment conditions. On the contrary, in this paper, we consider a parameter vector θ that has a **fixed** dimension (independent of n). This is the reason why we get $\sqrt{n}-$consistency.

5.2 Efficiency

Now we address the following question: Is it more efficient to use a continuum of moment conditions instead of a discrete subsample of moment conditions? The answer is yes. Let $H = L^2[0,1]$ and $\hat{\theta}_n^{CGMM}$ the GMM estimator associated with a continuum of moment conditions h_t, $t \in [0,1]$. By Theorem 8, its asymptotic variance is $\left\| E^{\theta_0}(\frac{\partial h}{\partial \theta'}) \right\|_K^{-2}$.

Let H_N be a Hilbert space of vectors of dimension 2^N. Let $\hat{\theta}_n^{DGMM}$ denote the GMM estimator associated with 2^N moment conditions corresponding to a discretization of the interval $[0,1]$: $h_{\frac{t}{2^N}}$, $t = 1, ...2^N$. It is well known that in that case, the optimal weighting matrix is the inverse of K^N, a $2^N \times 2^N$ matrix of general element $E^{\theta_0}(h_{\frac{t}{2^N}}, h_{\frac{s}{2^N}})$, for $t, s = 1, ..., 2^N$. Then, the variance-covariance matrix of the estimator is equal to

$$\Sigma = \left[E^{\theta_0} \left(\frac{\partial h'}{\partial \theta} \right) \left(K^N \right)^{-1} E^{\theta_0} \left(\frac{\partial h}{\partial \theta'} \right) \right]^{-1},$$

where in this instance h is simply a vector of moment conditions. To be consistent with our notation, Σ should be denoted $\| E^{\theta_0}(\frac{\partial h}{\partial \theta'}) \|_{K^N}^{-2}$. A result of Parzen (1959, p.316-18) states that for any function $f \in \mathcal{H}(K)$, then

$$\| f \|_K \geq \| f \|_{K^{N+1}} \geq \| f \|_{K^N} \text{ and } \quad \| f \|_{K^N} \overset{N\to\infty}{\longrightarrow} \| f \|_K .$$

It follows that the variance of $\hat{\theta}_n^{DGMM}$ is always at least as large as that of $\hat{\theta}_n^{CGMM}$.

6 Testing overidentifying restrictions

Hansen (1982) proposed a specification test obtained by replacing θ by $\hat{\theta}_n$ in the GMM objective function. If the model is correctly specified, all the moment conditions (including the overidentifying restrictions) should be close to zero. Hansen shows that this test converges to a Chi-square distribution with $m - q$ degrees of freedom, the difference between the number of restrictions tested, m, and the number of parameters to estimate q. Here, since the number of restrictions is infinite, this statistic diverges. But an appropriate standardization leads to a statistic that is asymptotically normal.

Define

$$p_n = \sum_{j=1}^{n} \frac{\mu_j^2}{\mu_j^2 + \alpha_n}, \; q_n = 2 \sum_{j=1}^{n} \frac{\mu_j^4}{\left(\mu_j^2 + \alpha_n\right)^2}, \; z_n = \sum_{j=1}^{n} \frac{\mu_j^6}{\left(\mu_j^2 + \alpha_n\right)^3}. \tag{11}$$

Let us consider the following property:

Liapunov's condition:

$$\frac{z_n}{q_n^{3/2}} \to 0 \;\; \text{as } n \to \infty. \tag{LC}$$

Lemma 9 *(a) - Assume K has an infinite number of eigenvalues. Then Condition (LC) is satisfied for any α_n converging to zero.*

(b) - If, moreover, there are $0 < \gamma < 1$ and some positive constant c such that $p_n \sim c\alpha_n^{-\gamma}$ as n goes to infinity, then $q_n \sim d\alpha_n^{-\gamma}$ and $z_n \sim e\alpha_n^{-\gamma}$ as n goes to infinity, where d and e are some positive constants.

Example 2. If $\mu_j = \frac{1}{j}, \forall j = 1, 2, ...$ and using the development in series of $\coth(\pi x)$, we get

$$\sum_{j=1}^{\infty} \frac{\mu_j^2}{\mu_j^2 + \alpha} = \sum_{j=1}^{\infty} \frac{1}{1 + \alpha j^2} = \frac{\pi}{2}\frac{1}{\sqrt{\alpha}}\coth\left(\frac{\pi}{\sqrt{\alpha}}\right) - \frac{1}{2}$$

for a fixed α. It follows from Lemma 9 (b) that p_n, q_n, and z_n diverge all at the same speed given by $\frac{1}{\sqrt{\alpha_n}}$ as n goes to infinity.

Remark 9. When the number of eigenvalues of K is finite (case where K is degenerate), some moments are redundant and the number of factors (in the sense of the principal component analysis) is finite. Then, there is no need to penalize and the objective function should converge to a Chi-square distribution like in finite dimensional GMM. This issue will not be investigated further.

Assumption 15: $q_n\sqrt{\alpha_n} \to \infty$ as n goes to infinity.

Remark 10. Assumption 15 implies that the eigenvalues of K should not converge to zero too fast. It limits the dependence between h_t. The same type of requirement can be found in DeJong-Bierens (1994) where some examples are provided. The Brownian motion (see Example 1) will not satisfy this condition because the rate $\frac{1}{j^2}$ is too fast.

Theorem 10 *Assume K is not degenerate. Under Assumptions 1, 2, 3, 4', 5, 6, 8', 9, 11', 12 to15, we have*

$$\tau_n = \frac{\| \sqrt{n}\bar{h}_n\left(\hat{\theta}\right) \|^2_{K_n^{\alpha_n}} - \hat{p}_n}{\sqrt{\hat{q}_n}} \overset{n\to\infty}{\longrightarrow} \mathcal{N}(0,1)$$

as α_n goes to zero and $n\alpha_n^3$ goes to infinity. $\hat{p}_n$ and $\hat{q}_n$ are the counterparts of p_n and q_n where the μ_j have been replaced by their estimators $\mu_j^{(n)}$.

Remark 11. A sketch of the proof is as follows. Let us replace in $\tau_n, \sqrt{n}\bar{h}_n$ by its limit $Y \sim \mathcal{N}(0, K)$ and $\mu_j^{(n)}$ and $\phi_j^{(n)}$ by the corresponding true values μ_j and ϕ_j. The statistic τ_n becomes

$$\tilde{\tau}_n = \frac{\displaystyle\sum_{j=1}^{n} \frac{\mu_j^2}{\mu_j^2 + \alpha_n} \left[\frac{(Y, \phi_j)^2}{\mu_j} - 1 \right]}{\left[\displaystyle\sum_{l=1}^{n} \frac{2\mu_l^4}{\left(\mu_l^2 + \alpha_n\right)^2} \right]^{\frac{1}{2}}}.$$

The random real elements $\frac{(Y, \phi_j)}{\sqrt{\mu_j}}$ are i.i.d. $\mathcal{N}(0, 1)$. It follows immediately that the above expression has a zero mean and a unit variance. The asymptotic normality is deduced from the Lindeberg Feller theorem as shown in the appendix.

Remark 12. p_n can be interpreted as the number of principal components (or eigenfunctions ϕ_j) that are really used in the estimation of θ. Indeed, an intuitive argument is the following. Since α_n converges to zero, there is a $l^* > 0$ such for all $j \geq l^*, \mu_j^2 = o_p(\alpha_n)$. Hence from l^* on, the terms of the sum are negligible. If α_n were zero, then $p_n = n$. But for $\alpha_n > 0$, $p_n < n$. Assume that $\mu_j = \frac{1}{j}, \ \ j = 1, 2, ...$ From Example 2, we know that $p_n \sim \frac{\pi}{2} \frac{1}{\sqrt{\alpha_n}}$. If α_n satisfies the condition of Theorem 10, $n\alpha_n^3 \to \infty$, then a sufficient condition for the asymptotic normality of τ_n is $p_n = o_p\left(n^{\frac{1}{6}}\right)$.

7 Examples

The first three examples illustrate the link between GMM and maximum likelihood estimators (MLE). They actually show that, in some specific cases, the GMM estimator is as efficient as the MLE. This suggests that when the MLE is not available, the GMM estimator is a good candidate. The first two examples assume that the data are observed in continuous time whereas the third one assumes i.i.d. cross-sectional data. Subsection 7.4 shows how to use our method to get efficient estimators and powerful tests in a cross-sectional setting. Subsection 7.5 considers efficient estimation of a scalar diffusion when data are observed in discrete time.

7.1 Link between GMM and MLE: Example 1 (continued)

Consider the following model

$$\left\{ \begin{array}{ll} X_t^i = F(t, \theta) + u_t^i & \quad E^{P_0}\left(u_t^i\right) = 0 \\ X_0^i = 0, u_0^i = 0 & \end{array} \right. \tag{12}$$

where $\{u_t^i\}$, $i = 1, 2, ..., n$ are independent processes defined on $t \in [0, T]$ and F is a differentiable function of t and θ. F satisfies $F(0, \theta) = 0, \forall \theta$. The moment conditions are given by

$$h_t(X^i, \theta) = X_t^i - F(t, \theta).$$

First assume that $u_t^i = W_t^i$ where W_t^i is a scalar Wiener process. Then $k(t,s) = E^{P_0}(W_t^i W_s^i) = t \wedge s$. The RKHS $\mathcal{H}(K)$ consists of absolutely continuous functions f over $[0,T]$ such that (Kutoyants (1984))

$$f(0) = 0, \| f \|_K^2 = \int_0^T [\frac{d}{dt} f(t)]^2 \, dt < \infty.$$

Then, the objective function to minimize is given by

$$\| \overline{h} \|_K^2 = \int_0^T \overline{h}'(t)^2 \, dt = \int_0^T [\overline{X}_t - F(t,\theta)]'^2 \, dt$$

where $\overline{X}_t = \frac{1}{n} \sum_{i=1}^n X_t^i$ and $f'(t) = \frac{d}{dt} f(t)$. $\hat{\theta}_n^{GMM}$ is the solution of

$$\int_0^T \frac{\partial}{\partial \theta} F'(t,\theta) d\overline{X}_t - \int_0^T \frac{\partial}{\partial \theta} F'(t,\theta) F'(t,\theta) \, dt = 0. \tag{13}$$

Now we want to compare this result with maximum likelihood estimation. Note that X_t defined in (12) is the solution of the following stochastic differential equation:

$$dX_t = F'(t,\theta) \, dt + dW_t \tag{14}$$

Let μ_X and μ_W denote the measures corresponding to the process X and the Wiener process W respectively. A necessary and sufficient condition for the equivalence of the measures μ_X and μ_W is $F \in \mathcal{H}(K)$, see Kutoyants (1984). Under this condition, one can write the likelihood ratio (that is, the Radon-Nikodym derivative of μ_X with respect to μ_W) as

$$\frac{d\mu_X}{d\mu_W} = \exp \left\{ \int_0^T F'(t,\theta) d\overline{X}_t - \frac{1}{2} \int_0^T F'^2(t,\theta) \, dt \right\}.$$

As a result, $\hat{\theta}_n^{MLE}$ is solution of (13) so that $\hat{\theta}_n^{MLE} = \hat{\theta}_n^{GMM}$. It means that we succeeded in giving a GMM counterpart to the MLE in continuous time. The equivalence between GMM and MLE in this setting is not surprising since in a Gaussian model, the first moment summarizes all the information. (Here the variance is assumed to be known but the result is not modified if W_t is replaced by σW_t with an unknown positive parameter σ.)

In the case where $F(t,\theta) = t\theta,$there is a sufficient statistic and we get $\hat{\theta}_n^{MLE} = \hat{\theta}_n^{GMM} = \frac{\bar{X}_T}{T}$. In this special case, GMM does not lead to using all the data but only the last observations.

Finally if the distribution of the process $\{u_t^i\}$ is unknown, the GMM estimator involves estimating the covariance $k(t,s)$ from observations of n trajectories. If $\{u_t^i\}$ is actually Gaussian, the GMM estimator will be as efficient as the MLE estimator as n goes to infinity. For a sample size n, the estimation of $K^{-1/2}$ requires the estimation of n eigenfunctions and eigenvalues. The estimators of the first eigenfunctions and eigenvalues (in decreasing order) improve with the sample size. However, the estimation of the last eigenvalues is pretty bad for any sample size. The penalization term α_n is used to discard the last eigenvalues/eigenfunctions. As discussed in Subsection 4.2, the smallest eigenvalues correspond to the eigenfunctions that are the least informative. A few simulations, not reported here, show that the continuous GMM delivers accurate estimators of θ even in small samples. However, the estimation of the variance of these estimators, $\| \partial \overline{h}/\partial\theta \|_{K_n^{\alpha_n}}^{-2}$, is very sensitive to the choice of α_n.

7.2 Optimal GMM and MLE: The parametric i.i.d. case

Let $(x^i)_{i=1,\dots,n}$ be an i.i.d. sample and $f(x^i|\theta)$ $(\theta \in I\!R^q)$ be the density of one observation. For simplicity we assume $x^i \in [0,T] \subset I\!R$ and that all the usual regularity assumptions for maximum likelihood estimation are satisfied. (The result may be easily extended to $[0,+\infty)$ or $(-\infty,+\infty)$.) The maximum likelihood estimator $\hat{\theta}_n^{MLE}$ is then asymptotically normal and its asymptotic variance matrix is I_θ^{-1} where I_θ is the usual Fisher information matrix.

Let $F(t|\theta)$ be the c.d.f. associated with $f(.|\theta)$ and $\hat{F}_n$ is the empirical c.d.f. defined by $\hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^n I\!I(x^i \leq t)$. Intuitively, an estimator of θ will be obtained by minimizing a distance between $\hat{F}_n(.)$ and $F(.|\theta)$. Using our previous notation, this goal will be achieved by choosing

$$h_t(X,\theta) = I\!I(X \leq t) - F(t|\theta)$$

which obviously satisfies $E^{P_0}(h_t(X,\theta_0)) = 0$ for any $t \in [0,T]$. Let $\hat{\theta}_n^{GMM}$ be the optimal GMM estimator. We want to show that the asymptotic variance of $\hat{\theta}_n^{GMM}$ is also I_θ^{-1}.

It is known that $n^{-1/2}\sum_{i=1}^n (I\!I(x_i \leq t) - F(t|\theta))$ converges to a Gaussian distribution with a covariance operator characterized by the kernel:

$$k(s,t) = F(s \wedge t) - F(s)F(t).$$

Equivalently, $\sqrt{n}(\hat{F}_n \circ F^{-1}(t|\theta) - t)$ converges to a Brownian Bridge. From Theorem 8, we deduce that the asymptotic variance of $\hat{\theta}_n^{GMM}$ is J_θ^{-1} where

$$J_\theta = \left\| \frac{\partial}{\partial\theta} E^{P_0}(h) \right\|_K^2 = \left\| \frac{\partial}{\partial\theta} F(t|\theta) \right\|_K^2 .$$

Let us use a result given by Parzen (1970, page 30) (see references therein): If $k(s,t) = g(s)\ g(t)\ G(s \wedge t)$, where g and G are continuously differentiable, and $G(0) = 0$, then:

$$\|\varphi\|_K^2 = \int_0^T \left(\frac{\varphi(t)}{g(t)} \right)^{'2} \frac{1}{G'(t)} dt$$

and φ, differentiable, is an element of the RKHS if this expression is finite. In our case $g(t) = 1 - F(t|\theta)$ and $G(u) = \frac{F(u|\theta)}{1-F(u|\theta)}$. Then, under some regularity conditions, namely $\frac{\partial F}{\partial\theta} < \infty$ as $t \to 0$ and $\left(\frac{\partial F}{\partial\theta}\right)^2 \frac{F}{1-F} \to 0$ if $t \to T$, and after some manipulations, we get $J_\theta \ = \ I_\theta$. It follows that $\hat{\theta}_n^{GMM}$ is as efficient as the MLE.

7.3 Statistical inference for counting processes

Let N_t^i be an i.i.d. sample of counting processes observed between 0 and T. Each N_t^i satisfies $N_t^i = \sum_p I\!I(t \geq \tau_p^i)$ where $(\tau_p^i)_p$ is the increasing sequence of jump times of the i-th process. For counting processes, a continuous time trajectory is fully described by the knowledge of a finite number of jump times. It is therefore natural to assume continuous time observations. We assume that the model generating each N_t^i is defined by the following stochastic differential relation which characterizes a multiplicative intensity model:

$$dN_t^i = \lambda_t(\theta) Y_t^i dt + dM_t^i. \tag{15}$$

An introduction to stochastic calculus for counting processes and applications of (15) can be found, e.g., in Karr (1986). Equation (15) summarizes the following properties: The process N_t^i has a stochastic intensity, $r_t(\theta) = \lim \frac{1}{\Delta t} P(N_{t+\Delta t}^i - N_t^i = 1|\mathcal{F}_t^i)$, where $\mathcal{F}_t^i$ denotes the information generated by the past of N_t^i. This intensity is factorized into the product $\lambda_t(\theta)Y_t^i$, where $\lambda_t(\theta)$ is a deterministic function depending on the parameter and identical for all the processes and Y_t^i is a predictable observed random process (typically function of the past of N_t^i) specific to each individual in the sample. This model does not include explanatory variables and our approach would have to be extended in order to cover this case.

In this model, the information matrix satisfies (Karr, Theorem 5.19):

$$I_\theta \;=\; \int_0^T \frac{\left(\frac{\partial\lambda}{\partial\theta}\right)^2}{\lambda}\eta_t dt \quad \text{ where } \eta_t = E^{P_0}(Y_t^i).$$

The GMM estimator is characterized by the function

$$h_t(N^i,\theta) = N_t^i - \int_0^T \lambda_s(\theta) Y_s^i ds$$

which satisfies the condition $E^{P_0}\left(h_t\left(N^i,\theta_0\right)\right) = 0$.

We do not present here the practical implementation of GMM and optimal GMM and we just show that the optimal GMM estimator has the same asymptotic variance as the maximum likelihood estimator. The covariance kernel is given by

$$k(s,t) = E^{P_0}(M_s^i M_t^i) = E^{P_0}(M_{s\wedge t}^i E(M_{s\vee t}^i|\mathcal{F}_{s\wedge t}))$$

$$= E^{P_0}({M_{s\wedge t}^i}^2) = E^{P_0}(< M_{s\wedge t}^i >) = \int_0^{s\wedge t} \lambda_u(\theta)\eta_u du.$$

From Theorem 8, the asymptotic variance of the optimal GMM estimator is equal to J_θ^{-1} where

$$J_\theta \;=\; \left\| \frac{\partial}{\partial\theta} E^{P_0}(h)\right\|_K^2 = \left\| \int_0^t \frac{\partial\lambda_s(\theta)}{\partial\theta}\eta_s ds\right\|_K^2 \;=\; \int_0^T \frac{\left(\frac{\partial\lambda_t(\theta)}{\partial\theta}\eta_t\right)^2}{\lambda_t(\theta)\eta_t} dt.$$

This last equality follows from Parzen (1970), see also Section 7.2, and proves that $J_\theta = I_\theta$.

7.4 Conditional moment restrictions

Assume that $X = (Y,Z)$ be a random vector and

$$E^{P_0}\left[\rho\left(Y,Z,\theta_0\right)|Z\right] = 0 \tag{16}$$

where ρ is a known function. Two problems are of interest: (i) estimate efficiently θ_0, (ii) test consistently the conditional moment restrictions. Equation (16) implies that for any function g

$$E^{P_0}\left[g\left(Z\right)\rho\left(Y,Z,\theta_0\right)\right] = 0. \tag{17}$$

Chamberlain (1987) shows that the efficiency bound for the estimation of θ_0 corresponds to the GMM efficiency bound. Moreover, he shows that by choosing a sequence of functions $\{g_l\}$ that is complete, one can come arbitrarily close to the efficiency bound. He suggests the set of moment conditions deduced from (17) by taking the family $g_l(Z) = Z^l$, $l = 1, ..., m_n$. Newey (1990) discusses the choice of the number, m_n, of instruments. He shows that $m_n = o_p(\sqrt{n})$ is a necessary condition to have consistency and $\sqrt{n}-$ asymptotic normality of the estimator.

Our paper provides an alternative way to approach Chamberlain's efficiency bound. Lemma 1 in Bierens (1990) establishes that if

$$E^{P_0}\left[\rho(Y, Z, \theta_0)\exp(tZ)\right] = 0 \text{ for all } t \in I \tag{18}$$

for some interval $I \subset I\!R$ (except maybe a set of measure zero) then $E\left[\rho(Y, Z, \theta_0)|Z\right] = 0$. Moreover any interval I, even small, can be used. We can estimate θ using the continuum of moment conditions (18) indexed by t in I, this estimator is $\sqrt{n}-$ consistent and asymptotically normal and has a variance close to Chamberlain's efficiency bound. The main limitation of our paper is that we handle only the case with $t \in I\!R$ that is $Z \in I\!R$. The generalization to the case where $Z \in I\!R^d, d > 1$ is beyond the scope of this paper.

Now, we turn to the testing problem. In a series of papers, Bierens has proposed tests of $H_0 : P\{E[Y|Z] = f(Z, \theta_0)\} = 1$ against $H_1 : P\{E[Y|Z] = f(Z, \theta_0)\} < 1.$ Here $\rho(Y, Z, \theta_0) = Y - f(Z, \theta_0)$. Bierens (1990) shows that a test based on (18) will be consistent against all deviations from the null hypothesis. DeJong-Bierens (1994) develop a test based on a sequence of functions $\{g_l\}$ satisfying (17) where for instance $(g_1(x), g_2(x), ..., g_l(x)) = (1, \sin(x), \sin(2x), \cos(x), ...)$ with $l = 1, 2, ..., m_n$ where $m_n \to \infty$ as $n \to \infty$. Their test is a specification test using overidentifying restrictions à la Hansen similar to τ_n defined in Section 6. We can apply τ_n to the continuum of restrictions (18) indexed by $t \in I = [-3, 3]$ for instance. To be able to compare both approaches, we follow DeJong-Bierens and replace θ by $\tilde{\theta}_n$ the nonlinear least-square estimator of θ in the objective function instead of replacing θ by $\hat{\theta}_n$. Define

$$\bar{h}_n\left(t, \tilde{\theta}_n\right) = \frac{1}{n}\sum_{i=1}^{n}\rho\left(y^i, z^i, \tilde{\theta}\right)\exp\left(tz^i\right) \tag{19}$$

where $x_i = (y^i, z^i)$ is an i.i.d. sample of $X = (Y, Z)$.

Since $\tilde{\theta}_n$ is consistent and $\sqrt{n}-$ asymptotically normal regardless of α_n, we get

$$\tilde{\tau}_n = \frac{\| \sqrt{n}\bar{h}_n\left(.,\tilde{\theta}_n\right) \|^2_{K_n^\alpha} - \hat{p}_n}{\sqrt{\hat{q}_n}} \overset{n\to\infty}{\longrightarrow} \mathcal{N}(0,1)$$

as n goes to infinity, α_n goes to zero, and $n\alpha_n^2$ goes to infinity. The proof follows easily from that of Theorem 10. Notice that α_n is allowed to converge to zero faster than in Theorem 10.

In the case where $\mu_j = \frac{1}{j}$, DeJong-Bierens get the following condition on the rate of increase of m_n : $m_n = o_p\left(n^{\frac{1}{5}}\right)$. By Remark 12 and under the requirement $n\alpha_n^2 \to \infty$, we get: $p_n = o_p\left(n^{\frac{1}{4}}\right)$. Therefore, our rate is faster than that of DeJong-Bierens. But this comparison is just illustrative because eigenvalues, $\mu_j = \frac{1}{j}$, do not satisfy our Assumption 15. p_n will dictate the speed of convergence $\tilde{\tau}_n$. Our test as well as that of DeJong-Bierens has power against any fixed alternative but does not have power against $\frac{1}{\sqrt{n}}$ local alternatives to H_0 . The reason is that, as explained in DeJong-Bierens, "the sequence m_n^{-1}(or here α_n) can be viewed as a sort of window width parameter present in nonparametric regression". Hence, the small sample properties of $\tilde{\tau}_n$ might not be satisfying.

7.5 Efficient estimation of a scalar diffusion model

Consider a scalar diffusion process

$$dx_t = \mu(x_t,\theta)dt + \sigma(x_t,\theta)dW_t \tag{20}$$

where μ and σ are known functions of the parameter of interest θ. Assume $\{x_t\}$ is stationary and strong mixing on $I\!R$. $\{x_t\}$ is observed at discrete time-points $t = 1, ..., T$, and T goes to infinity. Let $\mathcal{A}$ be the infinitesimal generator for $\{x_t\}$, $\mathcal{A}$ can be represented as (see Hansen-Scheinkman, 1995):

$$\mathcal{A}\phi = \mu\phi' + \frac{1}{2}\sigma^2\phi''.$$

Let $\mathcal{D}$ be the domain of $\mathcal{A}$. One wants to estimate θ using moment conditions proposed by Hansen-Scheinkman (1995):

$$\begin{aligned} E^{P_0}\left[\mathcal{A}\phi\left(x_t\right)\right] &= 0, \text{ for all } \phi \in \mathcal{D} \\ E^{P_0}\left[\phi\left(x_t\right)\mathcal{A}\phi\left(x_{t+1}\right) - \phi\left(x_{t+1}\right)\mathcal{A}\phi\left(x_t\right)\right] &= 0, \text{ for all } \phi \in \mathcal{D}. \end{aligned} \tag{21}$$

Both (21) and (21) might be necessary to the identification of the model. Conley et al. (1997, Appendix C) show that an efficient choice of ϕ in (21) is $\phi = \partial \ln q / \partial \theta$, where q is the stationary distribution of $\{x_t\}$. However, there is no result on the efficient choice of ϕ in (21). We suggest to use as test functions, $\phi_\delta(x) = \Phi\left(x/\delta\right)$, where Φ is the standard normal cumulative function. The estimation of model (20) will be based on a continuum of moment conditions:

$$h_\delta(x_t, x_{t+1}) = \phi_\delta(x_t)\mathcal{A}\phi_\delta\left(x_{t+1}\right) - \phi_\delta\left(x_{t+1}\right)\mathcal{A}\phi_\delta\left(x_t\right)$$

where $\delta > 0$ belongs to a well-chosen interval I. The estimates based on a full interval will be more efficient than those based on a few values of δ. Here $\{x_t\}$ are not independent, see Remark 5.

8 Conclusions and directions for future research

We achieved our goal of providing a framework that encompasses both the case where a discrete number of moment conditions are available and the case where a full interval is available. However, the generalization to a continuum of moment conditions is not as straightforward as expected. The determination of the optimal operator relies on inverting a covariance operator. But while the generalized inverse of a matrix always exists, the generalized inverse of a compact operator exists only for a subset of $L^2[0,T]$, the reproducing kernel Hilbert space (RKHS) with kernel the covariance between moment conditions. We give an estimator of the covariance operator and suggest the use of the method of regularization to guarantee the stability of the inverse.

There are numerous limitations to our analysis. We consider only cases where the index parameter, t, belongs to $I\!R$. However, for many applications, t belongs to $I\!R^d$, $d > 1$, see Subsection 7.4. Extension to $t \in I\!R^d$ should not be particularly difficult. It would be also of particular interest to examine the case where θ is a function. Moment conditions involving a function are frequent in constrained models. Fields of application include auctions, instrumental variables, etc. However, this nonparametric setting will complicate the proofs and alter the speed of convergence.

Appendix A: Definitions and properties of operators

This section summarizes useful results on operators. Most definitions and results reported here can be found in Dunford and Schwartz (1963) or in Hochstadt (1973).

Let $L^2[0,T]$ be the space of all square-integrable functions defined on the closed interval $[0,T]$ that take on real values. $L^2[0,T]$ with the inner product

$$(f,g) = \int_0^T f(t)g(t)\,dt$$

forms a Hilbert space. We denote the norm by

$$\|f\| = \left\{ \int_0^T f(t)^2\,dt \right\}^{1/2}$$

The extension to $(L^2[0,T])^J$ is straightforward but we simplify the exposition by imposing $J=1$.

An operator K assigns to an element f in $L^2[0,T]$ a new element Kf in $L^2[0,T]$. The operator needs not be defined on the full space $L^2[0,T]$ and in this case, its domain $\mathcal{D}(K)$ must be determined (for instance the differential operator $Kf = \frac{df}{dt}$ is defined only for the differentiable functions). Let $N(K)$ be the nullspace of the K, $N(K) = \{f \in L^2[0,T] \mid Kf = 0\}$, and $R(K)$ be the range of K, $R(K) = \{f \mid Kg = f, \text{ some } g \in L^2[0,T]\}$.

Definition 4 *The operator K is said to be* **linear** *if it satisfies*

$$K(\alpha f + \beta g) = \alpha K f + \beta K g$$

for any scalars α and β, and any functions f and g in $L^2[0,T]$.

An operator K is said to be **bounded** if for some constant $M > 0$, that may depend on f, we have

$$\|Kf\| \leq M\,\|f\|$$

for all f in $L^2[0,T]$. The greatest bound of all M is called the **norm** of K and denoted $\|K\|$. Another way of defining $\|K\|$ is by

$$\|K\| = \sup_{\|f\| \leq 1} \|Kf\|$$

An operator is continuous if and only if it is bounded.

Definition 5 *Let K be a bounded linear operator on $L^2[0,T]$. Let $\{f_n\}$ be an infinite uniformly bounded sequence in $L^2[0,T]$. K is said to be* **compact** *if from the sequence $\{Kf_n\}$ one can extract a subsequence $\{Kf_{n_l}\}$ that is a Cauchy sequence.*

Definition 6 *With K, we can associate the adjoint K^* that is defined by*

$$(Kf, g) = (f, K^* g) .$$

An operator is said to be **self-adjoint** *if $K = K^*$.*

Definition 7 *If for some μ*

$$K\phi = \mu\phi \tag{22}$$

has solutions other than $\phi = 0$, we shall call μ an **eigenvalue** *of K and the solutions of (22)* **eigenfunctions***.*

Lemma 11 *Let K be a self-adjoint operator on $L^2[0,T]$ then all eigenvalues of K are real.*

Lemma 12 *Let K be a* **nonnegative definite** *operator, that is,*

$$(Kf, f) \geq 0 \ \ \textit{for all } f \in L^2[0,T],$$

then all the eigenvalues of K are nonnegative.

Lemma 13 *Let K be a compact, self-adjoint operator on $L^2[0,T]$ then the set of its eigenvalues $\{\mu_j\}$ is countable and its eigenfunctions $\{\phi_j\}$ can be orthonormalized. Moreover, any function f in $L^2[0,T]$ can be represented as*

$$f = \sum_{j=1}^{\infty} (f, \phi_j)\, \phi_j + f_0,$$

where f_0 is a suitable element of the nullspace of K ($Kf_0 = 0$). It follows that

$$Kf = \sum_{j=1}^{\infty} \mu_j\, (f, \phi_j)\, \phi_j,$$

where μ_j is repeated according to its order of multiplicity.

If moreover K is nonnegative, the eigenvalues can be ordered as a decreasing sequence

$$\mu_1 \geq \mu_2 \geq ...$$

We consider an **integral** operator

$$Kf(t) = \int_0^T k(t,s) f(s) \, ds,$$

where $f \in L^2[0,T]$ and its **kernel** $k(t,s) : [0,T] \times [0,T] \to I\!R$. Notice that K is self-adjoint if k is symmetric $(k(t,s) = k(s,t))$.

Lemma 14 *Let $k(t,s)$ be an L^2 kernel, that is, $k(t,s)$ satisfies*

$$\int_0^T \int_0^T k(t,s)^2 \, dtds < \infty, \tag{23}$$

then the associated integral operator K is a compact operator of $L^2[0,T]$.

Lemma 15 *Let K be a compact self-adjoint integral operator with an L^2 kernel $k(t,s)$ and $\{\mu_j\}$ the set of eigenvalues. Then*

$$\int_0^T \int_0^T k(t,s)^2 \, dtds = \sum_{j=1}^{\infty} \mu_j^2.$$

Operators for which (23) holds are referred to as **Hilbert-Schmidt operators**. Notice that if $k(t,s)$ is continuous on $[0,T] \times [0,T]$, then (23) is necessarily satisfied. We denote the Hilbert-Schmidt norm as

$$\|K\|_{HS} = \left(\sum_{j=1}^{\infty} \mu_j^2 \right)^{1/2}.$$

Definition 8 (Parzen, 1970) *Every symmetric nonnegative definite kernel k defining an operator K possesses a unique* **reproducing kernel Hilbert space** *(RKHS) denoted $\mathcal{H}(K)$ defined as follows:*

(1) $\mathcal{H}(K)$ is a Hilbert space with inner product $(.,.)_K$,

(2) $k(.,t) \in \mathcal{H}(K)$, $\forall \ t \in [0,T]$,

(3) Reproducing property: $(k(.,t),f)_K = f(t)$, $f \in \mathcal{H}(K)$, $t \in [0,T]$.

9 Appendix B: Technical proofs

Proof of Theorem 2. A mean value expansion of $\overline{h_n}(\hat{\theta}_n)$ about θ_0 gives

$$\overline{h_n}(\hat{\theta}_n) = \overline{h_n}(\theta_0) + \frac{\partial \overline{h_n}}{\partial \theta'}(\bar{\theta})(\hat{\theta}_n - \theta_0)$$

where $\bar{\theta}$ is on the line segment joining $\hat{\theta}_n$ and θ_0. Differentiating the objective function with respect to θ yields to

$$\left(B_n\frac{\partial\overline{h_n}}{\partial\theta'}(\hat{\theta}_n), B_n\overline{h_n}(\hat{\theta}_n)\right) = 0$$

$$\Leftrightarrow \left(B_n\frac{\partial\overline{h}_n}{\partial\theta'}(\hat{\theta}_n), B_n\{\overline{h_n}(\theta_0) + \frac{\partial\overline{h}_n}{\partial\theta'}(\bar{\theta})(\hat{\theta}_n - \theta_0)\}\right) = 0$$

by the first order condition. Then by linearity of the operator, we obtain

$$(\hat{\theta}_n - \theta_0) = -\left(B_n\frac{\partial\overline{h_n}}{\partial\theta'}(\hat{\theta}_n), B_n\frac{\partial\overline{h}_n}{\partial\theta'}(\bar{\theta})\right)^{-1}\left(B_n\frac{\partial\overline{h_n}}{\partial\theta'}(\hat{\theta}_n), B_n\overline{h_n}(\theta_0)\right)$$

Assumption 8 implies the invertibility of the first matrix for n large. Since $\hat{\theta}_n \xrightarrow{\text{P}} \theta_0$ and then $\bar{\theta} \xrightarrow{\text{P}} \theta_0$, we have by Slutsky's Theorem and Assumption 11:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -\left(BE^{P_0}\frac{\partial h}{\partial\theta'}(\theta_0), BE^{P_0}\frac{\partial h}{\partial\theta'}(\theta_0)\right)^{-1}\left(BE^{\theta_0}\frac{\partial h}{\partial\theta'}(\theta_0), Z\right) + o_p(1)$$

where Z is defined in Remark 2. We have

$$\left(BE^{P_0}\frac{\partial h}{\partial\theta'}(\theta_0), Z\right) \sim \mathcal{N}\left(0, \left(BE^{P}\frac{\partial h}{\partial\theta'}(\theta_0), (BKB^*)BE^{P_0}\frac{\partial h}{\partial\theta'}(\theta_0)\right)\right)$$

The desired result follows.

Proof of Theorem 3. We consider μ_j as a function of F the c.d.f. of P_0. Let $\mu_j = U(F)$ and equivalently $\mu_j^{(n)} = U(F_n)$ where F_n is the empirical c.d.f.. A first order Taylor development in the sense of Frechet leads to

$$\mu_j^{(n)} - \mu_j = DU_F(F_n - F) + \varepsilon(F_n - F)\|F_n - F\| \tag{24}$$

The norm is the sup norm. The term $\varepsilon(F_n - F)$ converges to zero and $\sqrt{n}\|F_n - F\|$ is bounded. Let DU_F be the derivative of U in F. Then

$$\sqrt{n}\left(\mu_j^{(n)} - \mu_j\right) = \sqrt{n}DU_F(F_n - F) + o_p(1). \tag{25}$$

In order to compute the leading term, we differentiate the relation

$$\int E^F(k(X,t,s))\,\phi_j(s)\,ds = \mu_j\phi_j(t)$$

with respect to F, ϕ_j, and μ_j. E^F denotes the expectation taken with respect to F. If $\tilde{F}, \tilde{\phi}_j$ and $\tilde{\mu}_j$ are the corresponding differential elements, we get

$$\int E^{\tilde{F}} \left(k \left(X, t, s \right) \right) \phi_j \left(s \right) ds + \int E^F \left(k \left(X, t, s \right) \right) \widetilde{\phi}_j \left(s \right) ds = \mu_j \widetilde{\phi}_j \left(t \right) + \widetilde{\mu}_j \phi_j \left(t \right) \tag{26}$$

Multiplying (26) by $\phi_j(t)$ and integrating with respect to t, we obtain, using $\int \phi_j(t)^2 dt = 1$,

$$\widetilde{\mu}_j = DU_F \left(\tilde{F} \right) = \int \int E^{\tilde{F}} \left(k \left(X, t, s \right) \right) \phi_j \left(t \right) \phi_j \left(s \right) dtds.$$

From (25), we obtain

$$\sqrt{n} \left(\mu_j^{(n)} - \mu_j \right) = \sqrt{n} \frac{1}{n} \sum_{i=1}^{n} \left[\int \int k \left(x_i, t, s \right) \phi_j \left(t \right) \phi_j \left(s \right) dtds - \mu_j \right] + o_p \left(1 \right).$$

The result follows.

Proof of Theorem 4. (i) Since $\|K_n - K\| \leq \|K_n - K\|_{HS}$, we have

$$\begin{aligned} \|K_n - K\|^2 &\leq \int \int \left[k_n \left(t, s \right) - k \left(t, s \right) \right]^2 dtds \\ &= \int \int \left[\frac{1}{n} \sum_{i=1}^{n} k \left(x_i, t, s \right) - k \left(t, s \right) \right]^2 dtds \\ &= \frac{1}{n^2} \sum_{i,j} \int \int \left(k \left(x_i, t, s \right) - k \left(t, s \right) \right) \left(k \left(x_j, t, s \right) - k \left(t, s \right) \right) dtds. \end{aligned}$$

This expression is a U-statistic such that

$$E \left[\int \int \left(k \left(x_i, t, s \right) - k \left(t, s \right) \right) \left(k \left(x_j, t, s \right) - k \left(t, s \right) \right) dtds | x_j \right] = 0.$$

Using Serfling's theorem (1980, p.194), this U-statistic converges to a mixture of Chi-square distributions at the speed n. This implies the result.

(ii) As for Theorem 3, let $\|K\|_{HS}^2 = V(F)$ and $\|K_n\|_{HS}^2 = V(F_n)$. The first order Taylor expansion gives

$$\|K_n\|_{HS}^2 - \|K\|_{HS}^2 = DV_F \left(\widetilde{F} \right) + \varepsilon \left(\widetilde{F} \right) \left\| \widetilde{F} \right\|$$

with $\widetilde{F} = F_n - F$ and $DV_F \left(\widetilde{F} \right) = 2 \int E^F \left(k \left(X, t, s \right) \right) E^{\tilde{F}} \left(k \left(X, t, s \right) \right) dtds$. As before we have $\sqrt{n} \varepsilon \left(\widetilde{F} \right) \left\| \widetilde{F} \right\| = o_p(1)$, therefore

$$\sqrt{n} \left(\|K_n\|_{HS}^2 - \|K\|_{HS}^2 \right) = \sqrt{n} \frac{1}{n} \sum_{i=1}^{n} 2 \left[\int \int k \left(x_i, t, s \right) k \left(t, s \right) dtds - \int \int k^2 \left(t, s \right) dtds \right] + o_p \left(1 \right)$$

This yields the result.

Proof of Theorem 7. (i) Let $B = K^{-\frac{1}{2}}$, $B_n = K_n^{-\frac{1}{2}}$, $B^{\alpha_n} = \left(K^2 + \alpha_n I\right)^{-\frac{1}{2}} K^{\frac{1}{2}}$, and $B_n^{\alpha_n} = \left(K_n^2 + \alpha_n I\right)^{-\frac{1}{2}} K_n^{\frac{1}{2}}$. We want to show:

$$\|B_n^{\alpha_n} f_n - Bf\| \to 0$$

in probability, as n and $n\alpha_n^{\frac{3}{2}}$ go to infinity, α_n goes to zero. The proof has three steps:

1) $\|B^{\alpha_n} f - Bf\| \to 0$ as $\alpha_n \to 0$.

2) $\|B_n^{\alpha_n} f - B^{\alpha_n} f\| \to 0$ as $\alpha_n \to 0$ and $n\alpha_n^{3/2}$ goes to infinity.

3) $\|B_n^{\alpha_n} f_n - B_n^{\alpha_n} f\| \to 0$ as $\alpha_n \to 0$ and $n\sqrt{\alpha_n}$ goes to infinity.

Proofs of Steps 1) and 3) draw from Groetsch (1993, p.84-88).

Step 1) uses the Fourier decompositions:

$Bf = \sum_{j=1}^{\infty} \frac{1}{\sqrt{\mu_j}} \left(f, \phi_j\right) \phi_j$,

$B^{\alpha_n} f = \sum_{j=1}^{\infty} \frac{\sqrt{\mu_j}}{\sqrt{\mu_j^2 + \alpha_n}} \left(f, \phi_j\right) \phi_j$,

$\|B^{\alpha_n} f - Bf\|^2 = \sum_{j=1}^{\infty} \left[\frac{\sqrt{\mu_j^2+\alpha_n} - \mu_j}{\sqrt{\mu_j}\sqrt{\mu_j^2+\alpha_n}}\right]^2 \left(f, \phi_j\right)^2 \leq \sum_{j=1}^{\infty} \frac{1}{\mu_j} \left(f, \phi_j\right)^2 < \infty$,

since $f \in \mathcal{H}(K) + \mathcal{H}(K)^{\perp}$. We may, in passing to the limit as $\alpha_n \to 0$, interchange the limit and the summation so that $\|B^{\alpha_n} f - Bf\| \to 0$ as $\alpha_n \to 0$.

Step 3) follows from

$$\|B_n^{\alpha_n} f_n - B_n^{\alpha_n} f\| \leq \|B_n^{\alpha_n}\| \|f_n - f\|.$$

The second term on the right hand-side is $O_p\left(\frac{1}{\sqrt{n}}\right)$ by assumption. The first term is bounded by $1/\alpha^{\frac{1}{4}}$ for n large because of the following result:

$$\begin{aligned}
\|B^{\alpha_n}\|^2 &= \left\|\left(K^2 + \alpha_n I\right)^{-\frac{1}{2}} K^{\frac{1}{2}}\right\|^2 \\
&= \left(\left(K^2 + \alpha_n I\right)^{-\frac{1}{2}} K^{\frac{1}{2}}, \left(K^2 + \alpha_n I\right)^{-\frac{1}{2}} K^{\frac{1}{2}}\right) \\
&= \left(\left(K^2 + \alpha_n I\right)^{-\frac{1}{2}}, \left(K^2 + \alpha_n I\right)^{-\frac{1}{2}} K\right) \\
&\leq \underbrace{\left\|\left(K^2 + \alpha_n I\right)^{-\frac{1}{2}}\right\|}_{\leq \frac{1}{\sqrt{\alpha_n}}} \underbrace{\left\|\left(K^2 + \alpha_n I\right)^{-\frac{1}{2}} K\right\|}_{\leq 1}.
\end{aligned}$$

Therefore $\frac{1}{n\sqrt{\alpha_n}} \to 0$ implies $\|B_n^{\alpha_n} f_n - B_n^{\alpha_n} f\| \to 0$.

Step 2)

$$\begin{aligned}
\|B_n^{\alpha_n} f - B^{\alpha_n} f\| &\leq \left\|\left(K_n^2 + \alpha_n I\right)^{-\frac{1}{2}} K_n^{\frac{1}{2}} f - \left(K_n^2 + \alpha_n I\right)^{-\frac{1}{2}} K^{\frac{1}{2}} f\right\| \quad (A) \\
&+ \left\|\left(K_n^2 + \alpha_n I\right)^{-\frac{1}{2}} K^{\frac{1}{2}} f - \left(K^2 + \alpha_n I\right)^{-\frac{1}{2}} K^{\frac{1}{2}} f\right\|. \quad (B)
\end{aligned}$$

$$(A) \leq \underbrace{\left\| \left(K_n^2 + \alpha_n I \right)^{-\frac{1}{2}} \right\|}_{\leq \frac{1}{\sqrt{\alpha_n}}} \underbrace{\left\| K_n^{\frac{1}{2}} f - K^{\frac{1}{2}} f \right\|}_{= O_p\left(\frac{1}{\sqrt{n}} \right)}.$$

(A) goes to zero if $n\alpha_n$ goes to infinity.

$$\begin{aligned}
(B) \quad \leq & \quad \left\| (K_n^2 + \alpha_n I)^{-\frac{1}{2}} K^{\frac{1}{2}} f - (K_n^2 + \alpha_n I)^{-\frac{1}{2}} K (K^2 + \alpha_n I)^{-\frac{1}{2}} K^{\frac{1}{2}} f \right\| \quad (1) \\
+ & \quad \left\| (K_n^2 + \alpha_n I)^{-\frac{1}{2}} K_n (K^2 + \alpha_n I)^{-\frac{1}{2}} K^{\frac{1}{2}} f - (K^2 + \alpha_n I)^{-\frac{1}{2}} K^{\frac{1}{2}} f \right\| \quad (2) \\
+ & \quad \left\| (K_n^2 + \alpha_n I)^{-\frac{1}{2}} K (K^2 + \alpha_n I)^{-\frac{1}{2}} K^{\frac{1}{2}} f - (K_n^2 + \alpha_n I)^{-\frac{1}{2}} K_n (K^2 + \alpha_n I)^{-\frac{1}{2}} K^{\frac{1}{2}} f \right\|. \quad (3)
\end{aligned}$$

$$(1) \leq \underbrace{\left\| \left(K_n^2 + \alpha_n I \right)^{-\frac{1}{2}} K \right\|}_{\leq 1} \underbrace{\left\| \left(K^{-\frac{1}{2}} - B^{\alpha_n} \right) f \right\|}_{= O(\sqrt{\alpha_n})}. \text{ for } n \text{ large.}$$

$$\begin{aligned}
(2) \quad = & \quad \left\| \left(B_n^{\alpha} - K_n^{-\frac{1}{2}} \right) K_n^{\frac{1}{2}} (K^2 + \alpha I)^{-\frac{1}{2}} K^{\frac{1}{2}} f \right\| \\
\leq & \quad \underbrace{\left\| K_n^{\frac{1}{2}} \left(K^2 + \alpha I \right)^{-\frac{1}{2}} K^{\frac{1}{2}} \right\|}_{\leq 1} \underbrace{\left\| \left(B_n^{\alpha} - K_n^{-\frac{1}{2}} \right) f \right\|}_{= O(\sqrt{\alpha_n})}. \text{ for } n \text{ large.}
\end{aligned}$$

$$(3) \leq \underbrace{\left\| \left(K_n^2 + \alpha_n I \right)^{-\frac{1}{2}} \right\|}_{\leq \frac{1}{\sqrt{\alpha_n}}} \underbrace{\| K_n - K \|}_{= O_p\left(\frac{1}{\sqrt{n}} \right)} \underbrace{\| B^{\alpha_n} f \|}_{\leq 1/\alpha_n^{1/4} \| f \|}.$$

We actually have

$$\| B_n^{\alpha_n} f_n - B^{\alpha_n} f \| = O_p \left(\frac{1}{\sqrt{n}} \frac{1}{\alpha_n^{3/4}} \right).$$

$(1), (2)$, and (3) converge to zero as long as α goes to zero and $n\alpha_n^{3/2}$ goes to infinity.

(ii) can be proved similarly.

Proof of Theorem 8. The consistency follows directly from Theorem 7. Consider now the asymptotic distribution. The proof of Theorem 2 is valid only if B is a bounded operator. Here $B = K^{-\frac{1}{2}}$ is not bounded. We use the beginning of the proof of Theorem 2 up to the point where Slutsky's Theorem is mentioned. Using the fact $B_n^{\alpha_n}$ is self-adjoint and $B_n^{\alpha} B_n^{\alpha} = (K_n^{\alpha_n})^{-1}$, we want to show that

$$\left((K_n^{\alpha_n})^{-1} \frac{\partial \overline{h_n}}{\partial \theta'}(\hat{\theta}_n), \sqrt{n} \overline{h_n}(\theta_0) \right) \rightarrow \mathcal{N} \left(0, \left\| E^{P_0} \left(\frac{\partial h}{\partial \theta'} \right) \right\|_K^2 \right).$$

This result will be proved in two steps:

Step 1: Show that

$$\left((K_n^{\alpha_n})^{-1} \frac{\partial \overline{h_n}}{\partial \theta'}(\hat{\theta}_n), \sqrt{n}\overline{h_n}(\theta_0) \right) \overset{D}{\to} \left(K^{-1} E^{P_0} \frac{\partial h}{\partial \theta'}(\theta_0), Y \right).$$

Step 2: Show that

$$\left(K^{-1} E^{P_0} \frac{\partial h}{\partial \theta'}(\theta_0), Y \right) \sim \mathcal{N}\left(0, \left\| E^{P_0}\left(\frac{\partial h}{\partial \theta'} \right) \right\|_K^2 \right).$$

Step 1:

$$\begin{aligned} & \left((K_n^{\alpha_n})^{-1} \frac{\partial \overline{h_n}}{\partial \theta'}(\hat{\theta}_n), \sqrt{n}\overline{h_n}(\theta_0) \right) \\ = \quad & \left((K_n^{\alpha_n})^{-1} \frac{\partial \overline{h_n}}{\partial \theta'}(\hat{\theta}_n) - K^{-1} E^{P_0} \frac{\partial h}{\partial \theta'}(\theta_0), \sqrt{n}\overline{h_n}(\theta_0) \right) & (27) \\ & + \left(K^{-1} E^{P_0} \frac{\partial h}{\partial \theta'}(\theta_0), \sqrt{n}\overline{h_n}(\theta_0) \right). & (28) \end{aligned}$$

We have

$$(27) \le \underbrace{\left\| (K_n^{\alpha_n})^{-1} \frac{\partial \overline{h_n}}{\partial \theta'}(\hat{\theta}_n) - K^{-1} E^{P_0} \frac{\partial h}{\partial \theta'}(\theta_0) \right\|}_{=o_p(1)} \underbrace{\left\| \sqrt{n}\overline{h_n}(\theta_0) \right\|}_{=O_p(1)}$$

by Theorem 7 and

$$(28) \overset{D}{\to} \left(K^{-1} E^{P_0} \frac{\partial h}{\partial \theta'}(\theta_0), Y \right)$$

by definition of the convergence in a Hilbert space.

Step 2: Note that $\left(K^{-1} E^{P_0} \frac{\partial h}{\partial \theta'}(\theta_0), Y \right) = \left(E^{P_0} \frac{\partial h}{\partial \theta'}(\theta_0), Y \right)_K$ where $(.,.)_K$ denotes the inner product in the RKHS. As Kailath (1971) argues, this is not a real inner product but the notation is convenient. It is usually referred to as congruence inner product. Since $k(t,s) = E^{P_0}(Y_t Y_s)$, we have $\left(E^{P_0} \frac{\partial h}{\partial \theta'}, Y \right)_K \sim \mathcal{N}\left(0, \left\| E^{P_0}\left(\frac{\partial h}{\partial \theta'} \right) \right\|_K^2 \right)$, see for instance Parzen (1970).

Taking the proof of Theorem 2 where we left it and using the results of step 1 and step 2, it follows that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{n \to \infty}{\longrightarrow} \mathcal{N}(0, V_K)$$

where $V_K = \| E^{P_0}(\frac{\partial h}{\partial \theta'}) \|_K^{-2}$.

We address now the issue of the optimality. For any linear bounded operator B, the variance, V, of the GMM estimator is given in Theorem 2. To simplify notations, θ is omitted in the expression of h. In order to show that $V - V_K$ is definite positive, we shall show that

$$\forall z \in I\!R^K,$$
$$z' \left[\left(BE^{P_0}\tfrac{\partial h}{\partial \theta'}, (BKB^*)BE^{P_0}\tfrac{\partial h}{\partial \theta'} \right) - \left\| BE^{P_0}\tfrac{\partial h}{\partial \theta'} \right\|^2 \left\| K^{-\frac{1}{2}} E^{P_0}\tfrac{\partial h}{\partial \theta'} \right\|^{-2} \left\| BE^{P_0}\tfrac{\partial h}{\partial \theta'} \right\|^2 \right] z \geq 0$$

Let f denote $E^{P_0}\frac{\partial h}{\partial \theta'}$ and $u = Bfz$. The above inequality can be rewritten:

$$\begin{gathered} (u, (BKB^*)u) - (u, Bf) \left\| K^{-\frac{1}{2}} f \right\|^{-2} (Bf, u) \geq 0 \\ \Leftrightarrow (B^*u, KB^*u) - (B^*u, f) \left\| K^{-\frac{1}{2}} f \right\|^{-2} (f, B^*u) \geq 0 \\ \Leftrightarrow (v, Rv) \geq 0 \end{gathered}$$

where $v = B^*u$ and $R : g \rightarrow Kg - f \left\| K^{-\frac{1}{2}} f \right\|^{-2} (f, g)$. It is easy to check that $RK^{-1}R = R$ and that R is self-adjoint. It follows that

$$(v, Rv) = (v, RK^{-1}Rv) = \left\| K^{-\frac{1}{2}} Rv \right\|^2 \geq 0$$

since Rv belongs to $\mathcal{H}(K)$.

Proof of Lemma 9. (a) - First we rewrite q_n and z_n in a convenient way.

$$q_n = 2\sum_{j=1}^{n} \frac{\mu_j^4}{\left(\mu_j^2 + \alpha_n\right)^2} = 2\sum_{j=1}^{n} \frac{1}{\left(1 + \frac{\alpha_n}{\mu_j^2}\right)^2}, \tag{29}$$

$$z_n = \sum_{j=1}^{n} \frac{\mu_j^6}{\left(\mu_j^2 + \alpha_n\right)^3} = \sum_{j=1}^{n} \frac{1}{\left(1 + \frac{\alpha_n}{\mu_j^2}\right)^3}. \tag{30}$$

From (29) and (30), we have

$$\frac{z_n}{q_n} < \frac{1}{2}, \text{ and } \frac{z_n}{q_n^{3/2}} < \frac{1}{2q_n^{1/2}}.$$

Since $q_n \rightarrow \infty$ as n goes to infinity as long as α_n goes to zero, (LC) is satisfied.

(b) - Define

$$\bar{p}_n(\alpha) = \sum_{j=1}^{n} \frac{\mu_j^2}{\mu_j^2 + \alpha} = \sum_{j=1}^{n} \frac{1}{1 + \frac{\alpha}{\mu_j^2}} = \frac{1}{\alpha} \sum_{j=1}^{n} \frac{1}{\frac{1}{\alpha} + \frac{1}{\mu_j^2}}$$

and

$$\bar{p}_{\infty}(\alpha)=\sum_{j=1}^{\infty}\frac{\mu_j^2}{\mu_j^2+\alpha}=\frac{1}{\alpha}\sum_{j=1}^{\infty}\frac{1}{\frac{1}{\alpha}+\frac{1}{\mu_j^2}}$$

$\alpha\bar{p}_n(\alpha)$ converges uniformly in α to $\alpha\bar{p}_{\infty}(\alpha)$ as n goes to infinity since

$$\left|\alpha\bar{p}_n(\alpha)-\alpha\bar{p}_{\infty}(\alpha)\right|\leq\sum_{j=n+1}^{\infty}\frac{1}{\frac{1}{\alpha}+\frac{1}{\mu_j^2}}\leq\sum_{j=n+1}^{\infty}\mu_j^2$$

This sum converges to 0 because by Assumption 12, $\sum_{j=1}^{\infty}\mu_j^2<\infty$. Then $\bar{p}_n(\alpha)$ converges uniformly to $\bar{p}_{\infty}(\alpha)$ and by assumption, we have

$$\bar{p}_n(\alpha)\sim c\alpha^{-\gamma}$$

Define equivalently

$$\bar{q}_n(\alpha)=2\sum_{j=1}^{n}\frac{\mu_j^4}{\left(\mu_j^2+\alpha\right)^2},\ \text{ and }\ \bar{z}_n(\alpha)=\sum_{j=1}^{n}\frac{\mu_j^6}{\left(\mu_j^2+\alpha\right)^3}.$$

Let $x=\frac{1}{\alpha}$, we define

$$\begin{aligned}\bar{p}^*(x)&=&\bar{p}_n\left(\frac{1}{x}\right)=x\sum_{j=1}^{n}\frac{1}{x+\frac{1}{\mu_j^2}}\sim cx^{\gamma},\\ \bar{q}^*(x)&=&\bar{q}_n\left(\frac{1}{x}\right)=2x^2\sum_{j=1}^{n}\frac{1}{\left(x+\frac{1}{\mu_j^2}\right)^2},\ \text{and}\ \bar{z}^*(x)=\bar{z}_n\left(\frac{1}{x}\right)=x^3\sum_{j=1}^{n}\frac{1}{\left(x+\frac{1}{\mu_j^2}\right)^3}.\end{aligned}$$

Differentiating p_n with respect to x leads to the identity

$$\bar{q}^*(x)=2\bar{p}^*(x)-2x\frac{d\bar{p}^*(x)}{dx}$$

Therefore, q_n and p_n diverge at the same rate. Differentiating q_n with respect to x leads to

$$\bar{z}^*(x)=\frac{\bar{q}^*(x)}{2}-\frac{x}{4}\frac{d\bar{q}^*(x)}{dx}.$$

Hence, z_n and q_n diverge at the same speed.

Proof of Theorem 10. Let P_n denote the projection that associates to an operator K the operator $\tilde{K}_n$ defined by the first n eigenfunctions and eigenvalues of K. We are going to prove our result in three steps:

Step 1: Show that

$$\frac{1}{\sqrt{q_n}}\left\{\parallel\left(K_n^{\alpha_n}\right)^{-1/2}\sqrt{n}\bar{h}_n\left(\hat{\theta}_n\right)\parallel-\parallel P_n\left(K^{\alpha_n}\right)^{-1/2}Y\parallel\right\}\xrightarrow{P}0$$

where $Y \sim \mathcal{N}(0, K)$ in H.

Step 2: Show that

$$\frac{1}{\sqrt{q_n}} \left\{ \| \; P_n \left(K^{\alpha_n}\right)^{-1/2} Y \; \|^2 - p_n \right\} \rightarrow \mathcal{N}(0, 1).$$

Step 3: Show that

$$\hat{p}_n - p_n \xrightarrow{P} 0, \quad \text{and} \quad \hat{q}_n - q_n \xrightarrow{P} 0.$$

Step 1: We have

$$\begin{aligned}
& \frac{1}{\sqrt{q_n}} \left\{ \| \; \left(K_n^{\alpha_n}\right)^{-1/2} \sqrt{n} \bar{h}_n \left(\hat{\theta}_n\right) \; \| - \| \; P_n \left(K^{\alpha_n}\right)^{-1/2} Y \; \| \right\} \\
\leq \; & \frac{1}{\sqrt{q_n}} \left\| \left(K_n^{\alpha_n}\right)^{-1/2} \sqrt{n} \bar{h}_n \left(\hat{\theta}_n\right) - P_n \left(K^{\alpha_n}\right)^{-1/2} Y \right\| \\
\leq \; & \frac{1}{\sqrt{q_n}} \left\{ \left\| \left(K_n^{\alpha_n}\right)^{-1/2} \left(\sqrt{n} \bar{h}_n \left(\hat{\theta}_n\right) - Y \right) \right\| + \left\| \left(\left(K_n^{\alpha_n}\right)^{-1/2} - P_n \left(K^{\alpha_n}\right)^{-1/2} \right) Y \right\| \right\} \\
\leq \; & \frac{1}{\sqrt{q_n}} \left\{ \left\| \left(K_n^{\alpha_n}\right)^{-1/2} \right\| \left\| \sqrt{n} \bar{h}_n \left(\hat{\theta}_n\right) - Y \right\| + \left\| \left(K_n^{\alpha_n}\right)^{-1/2} - P_n \left(K^{\alpha_n}\right)^{-1/2} \right\| \|Y\| \right\}.
\end{aligned}$$

We have

$$\left\| K_n^{\alpha_n - \frac{1}{2}} \right\| \leq \frac{1}{\alpha_n^{1/4}} \quad \text{and} \quad \left\| \sqrt{n} \bar{h}_n \left(\hat{\theta}_n\right) - Y \right\| = O_p(1)$$

by Assumption 11' and Theorem 8 with $n\alpha_n^3 \rightarrow \infty$. Moreover, we have

$$\|Y\| = O_p(1) \quad \text{and} \quad \left\| \left(K_n^{\alpha_n}\right)^{-1/2} - P_n \left(K^{\alpha_n}\right)^{-1/2} \right\| \leq \|P_n\| \left\| B_n^{\alpha_n} - B^{\alpha_n} \right\|$$

with

$$\|P_n\| \leq 1 \quad \text{and} \quad \left\| B_n^{\alpha_n} - B^{\alpha_n} \right\| = O_p \left(\frac{1}{\sqrt{n}} \frac{1}{\alpha_n^{3/4}} \right).$$

by the proof of Theorem 7. Therefore the result is established under the conditions $n\alpha_n^{3/2} \rightarrow \infty$ and $q_n \sqrt{\alpha_n} \rightarrow \infty$. The first one is necessarily satisfied under the condition, $n\alpha_n^3 \rightarrow \infty$, and the second one is not true in general and is imposed in Assumption 15.

Note that Step 1 implies

$$\frac{1}{\sqrt{q_n}} \left\{ \| \; \left(K_n^{\alpha_n}\right)^{-1/2} \sqrt{n} \bar{h}_n \left(\hat{\theta}_n\right) \; \|^2 - \| \; P_n \left(K^{\alpha_n}\right)^{-1/2} Y \; \|^2 \right\} \xrightarrow{P} 0$$

Step 2: First we define some notations

$$\frac{1}{\sqrt{q_n}} \left\{ \| \; P_n \left(K^{\alpha_n}\right)^{-1/2} Y \; \|^2 - p_n \right\} = \sum_{j=1}^{n} X_{nj}$$

with

$$X_{nj}=\frac{\frac{\mu_j^2}{\mu_j^2+\alpha_n}\left[\frac{(Y,\phi_j)^2}{\mu_j}-1\right]}{\left[\sum_{l=1}^n\frac{2\mu_l^4}{\left(\mu_l^2+\alpha_n\right)^2}\right]^{\frac{1}{2}}}$$

$\frac{(Y,\phi_j)}{\sqrt{\mu_j}}$ are independent $\mathcal{N}(0,1)$ (see for instance Shorack-Wellner (1986), p.15), therefore the X_{nj} are independent with $E\left(X_{nj}\right)=0$ and $V\left(X_{nj}\right)=\sigma_{nj}^2$. $S_n=\sum_{j=1}^n X_{nj}$ satisfies $E\left(S_n^2\right)=\sum_{j=1}^n\sigma_{nj}^2=1$. Liapunov's Theorem states that a sufficient condition for $S_n\rightarrow\mathcal{N}(0,1)$ is the so-called Liapunov's condition

$$\lim_{n\rightarrow\infty}\sum_{j=1}^n E\left|X_{nj}\right|^3=0$$

See for instance Davidson (1994, p. 373). Liapunov's condition is given by Equation (LC) which is necessarily satisfied (see Lemma 9 (a)).

Step 3:

$$\begin{aligned}\hat{p}_n-p_n&=\sum_{j=1}^n\left(\frac{\mu_j^{(n)2}}{\mu_j^{(n)2}+\alpha_n}-\frac{\mu_j^2}{\mu_j^2+\alpha_n}\right)\\&=\alpha_n\sum_{j=1}^n\frac{\mu_j^{(n)2}-\mu_j^2}{\left(\mu_j^{(n)2}+\alpha_n\right)\left(\mu_j^2+\alpha_n\right)}\\&\leq\frac{\alpha_n}{\alpha_n^2}\sum_{j=1}^n\left(\mu_j^{(n)2}-\mu_j^2\right)\\&\leq O_p\left(\frac{1}{\alpha_n\sqrt{n}}\right)\end{aligned}$$

by Theorem 4. Therefore, $\hat{p}_n-p_n\rightarrow 0$ if $\alpha_n\sqrt{n}\rightarrow\infty$.

$$\begin{aligned}\frac{1}{2}\left(\hat{q}_n-q_n\right)&=\sum_{j=1}^n\left(\frac{\mu_j^{(n)4}}{\left(\mu_j^{(n)2}+\alpha_n\right)^2}-\frac{\mu_j^4}{\left(\mu_j^2+\alpha_n\right)^2}\right)\\&=\sum_{j=1}^n\left(\frac{\mu_j^{(n)2}}{\mu_j^{(n)2}+\alpha_n}-\frac{\mu_j^2}{\mu_j^2+\alpha_n}\right)\left(\frac{\mu_j^{(n)2}}{\mu_j^{(n)2}+\alpha_n}+\frac{\mu_j^2}{\mu_j^2+\alpha_n}\right)\\&\leq\left(\hat{p}_n-p_n\right)\sup_j\left(\frac{\mu_j^{(n)2}}{\mu_j^{(n)2}+\alpha_n}+\frac{\mu_j^2}{\mu_j^2+\alpha_n}\right)\\&\leq 2\left(\hat{p}_n-p_n\right)\end{aligned}$$

Hence, $\hat{q}_n-q_n\rightarrow 0$ if $\alpha_n\sqrt{n}\rightarrow\infty$.

To complete the proof, we need to check that the three steps are enough to guarantee the result. Let A denote $\parallel K_n^{\alpha-\frac{1}{2}}\sqrt{n}\bar{h}_n\left(\hat{\theta}\right)\parallel^2$. A Taylor expansion of $\frac{A-\hat{p}_n}{\sqrt{\hat{q}_n}}$ around p_n and q_n gives

$$\frac{A-\hat{p}_n}{\sqrt{\hat{q}_n}}=\frac{A-p_n}{\sqrt{q_n}}-\frac{1}{\sqrt{q_n}}\left(\hat{p}_n-p_n\right)-\frac{A-p_n}{2\sqrt{q_n}q_n}\left(\hat{q}_n-q_n\right)+o_p\left(1\right)$$

This leads to the result.

References

Andrews, D. (1991) "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation", *Econometrica*, **59**, No. 3, 817-858.

Bierens, H. (1990) "A Consistent Conditional Moment Test of Functional Form", *Econometrica*, Vol. 58, No. 6, 1443-1458.

Billingsley, P. (1968) *Convergence of Probability Measures*, John Wiley & Sons, New York.

Chamberlain, G. (1987) "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions", *Journal of Econometrics*, 34, 305-334.

Chen, X. and H. White (1998) "Central Limit and Functional Central Limit Theorems for Hilbert-Valued Dependent Heterogeneous Arrays with Applications", *Econometric Theory,* **14**, 260-284. See also working paper, 1992, University of California, San Diego.

Conley, T., L. P. Hansen, E. Luttmer, and J. Scheinkman (1997) "Short-Term Interest Rates as Subordinated Diffusions", *The Review of Financial Studies*, Vol. 10, No. 3, 525-577.

Dauxois, J., A. Pousse, and Y. Romain (1982) "Asymptotic Theory for the Principal Component Analysis of a Vector Random Function: Some Applications to Statistical Inference", *Journal of Multivariate Analysis*, 12, 136-154.

Davidson, J. (1994) *Stochastic Limit Theory*, Oxford University Press, Oxford.

De Jong, R.M. and H.J. Bierens (1994) "On the Limit Behavior of a Chi-square Type Test if the Number of Conditional Moments tested approaches Infinity", *Econometric Theory*, **9**, 70-90.

Dunford, N. and J. Schwartz (1963) *Linear Operators, part II,* Wiley & Sons, New York.

Golub, G.H., M. Health, and G. Wahba (1979) "Generalized Cross Validation as a Method for Choosing a Good Ridge Parameter", *Technometrics*, 21, 215-224.

Groetsch, C. (1993) *Inverse Problems in the Mathematical Sciences* , Vieweg, Wiesbaden.

Hall A. (1993) "Some Aspects of the Generalized Method of Moments Estimation". In: G.S. Maddala, C.R. Rao and H.D. Vinod, eds., *Handbook of Statistics*, Vol 11, 393-417. North Holland, Amsterdam.

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

Hansen, L. and J. Scheinkman (1995) "Back to the Future: Generating Moment Implications for Continuous-time Markov Processes", *Econometrica*, **63**, 767-804.

Hansen, P. C. (1992) "Numerical tools for analysis and solution of Fredholm integral equations of the first kind", *Inverse Problems*, **8**, 849-872.

Hochstadt, H. (1973) *Integral Equations.* Wiley and Sons.

Huber, P. J. (1973) "Robust Regression: Asymptotics, conjectures and Monte Carlo", *The Annals of Statistics*, Vol. 1, No. 5, 799-821.

Kailath, T. (1971) "RKHS Approach to Detection and Estimation Problems-Part I", *IEEE Trans. Inform. Theory* **IT-17**, 530-549.

Karr, A. (1986) *Point Processes and their Statistical Inference.* Marcel Dekker, New York.

Koenker, R. and J. Machado (1997) "GMM Inference when the Number of Moment Conditions is Large", mimeo.

Kutoyants, Y. (1984) *Parameter estimation for stochastic processes*, Heldermann Verlag, Berlin.

Nashed, M.Z. and G. Wahba (1974) "Generalized inverses in reproducing kernel spaces: An approach to regularization of linear operator equations", *SIAM Journal on Mathematical Analysis*, Vol. 5., No. 6, 974-987.

Newey, W. (1990) "Efficient Instrumental Variables Estimation of Nonlinear Models", *Econometrica*, Vol. 58, No. 4, 809-837.

Newey, W. and K. West (1987)"A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix", *Econometrica*, **55**, No. 3, 703-708.

Ogaki, M. (1993) "Generalized Method of Moments: Econometric applications". In: G.S. Maddala, C.R. Rao and H.D. Vinod, eds., *Handbook of Statistics*, Vol 11, 455-488. North Holland, Amsterdam.

Parzen, E. (1959) "Statistical Inference on time series by Hilbert Space Methods,I.", Technical Report No.23, Applied Mathematics and Statistics Laboratory, Stanford.

Parzen, E. (1970) "Statistical Inference on time series by RKHS methods", 12th Biennial Seminar Canadian Mathematical Congress Proc., R. Pyke, ed., Canadian Mathematical Society, Montreal.

Portnoy, S. (1985) "Asymptotic Behavior of M Estimators of p Regression Parameters when p^2n is large; II. Normal Approximation", *The Annals of Statistics*, Vol. 13, No.4, 1403-1417.

Serfling, R. J. (1980) *Approximation Theorems of Mathematical Statistics.* Wiley & Sons, New York.

Shorack,G. and J. Wellner (1986) *Empirical Processes with Applications to Statistics*, Wiley & Sons, New York.

Wahba, G. (1973) “Convergence Rates of Certain Approximate Solutions to Fredholm Integral Equations of the First Kind”, *Journal of Approximation Theory*, **7**, 167-185.

White, H (1994) *Estimation, Inference and Specification Analysis*, Cambridge University Press.