Efficient Smooth GMM and Dimension Reduction

Pascal Lavergne
Simon Fraser University

Valentin Patilea
CREST-ENSAI

November 2007

Preliminary and incomplete

Abstract

We propose a new GMM criterion for models defined by conditional moment restrictions based on local averaging. It resembles a statistic based on smoothing techniques used in specification testing. Depending on whether the smoothing parameter is fixed or decreases to zero with the sample size, our approach defines a whole class of estimators. We show that consistency and asymptotic normality follows in both cases. However we show that at a first-order, letting the smoothing parameter tend to zero yields a semiparametric efficient estimator, and we provide a two-step efficient version. We also investigate a dimension-reduction device in the context of smooth GMM. While the resulting estimator does not attain the semiparametric efficiency bound, its higher-order properties may be preferable.

Keywords: Dimensionality, Moment conditions, Smoothing methods.

1 Introduction

Generalized Method of Moments (GMM) is an important and popular method that estimate parameters by matching at best population moment restrictions at the sample level. Beginning with Hansen (1982), a large theoretical literature has been devoted to derive its asymptotic properties, the form of the optimal instruments, see e.g. Hansen (1985), and how to estimate them, see e.g. Robinson (1987) and Newey (1993). However, since many econometric models involve conditional moment restrictions, or equivalently an infinite (countable or uncountable) number of moment restrictions, recent research has focused on estimation of such models. Carrasco and Florens (2000) proposed a generalization of GMM to a continuum of moment conditions. Donald, Imbens, and Newey (2003) simultaneously studied GMM and Empirical Likelihood (EL) with a number of moment conditions growing with the sample size. Antoine, Bonnal, and Renault (2007), Kitamura, Tripathi, and Ahn (2004), and Smith (2007) proposed different versions of Generalized EL (GEL). All these authors derived estimators that attain under some assumptions the semiparametric efficiency bound as established by Chamberlain (1987), and all rely on a user-chosen parameter, whether it is a regularization parameter, as in Carrasco and Florens (2000), a bandwidth parameter, as in Antoine, Bonnal, and Renault (2007), Kitamura, Tripathi, and Ahn (2004), and Smith (2007), or the number of series functions, as in Donald, Imbens, and Newey (2003). The latter parameter, or its inverse in the cases of a bandwidth or regularization parameter, can be roughly interpreted as the number of unconditional moment restrictions or instruments taken into account in estimation.

Recently, Dominguez and Lobato (2004) point out that in nonlinear models an arbitrary finite number of instruments, and even the optimal ones, may fail to globally identify the parameters of interest, so that the GMM estimator can be inconsistent, and provide some examples, see also Dominguez and Lobato (2007) for an example based on the CAPM. This is a crucial issue since efficiency first requires consistency, consistency involves the choice of unconditional moment restrictions to be considered, and in practice, one can never be sure that the chosen moment restrictions identify the parameter of interest. Dominguez and Lobato (2004) then propose the first consistent estimator that does not require a user-chosen parameter, but still exploits all conditional moment restrictions. To attain efficiency though, their first-step estimator should be modified using a Newton-Raphson algorithm in the direction of the efficient GMM estimator as proposed by Newey (1993).

Our first main contribution is to propose a new class of parametric estimators in models defined by conditional moment restrictions that bridges the gap between Dominguez and Lobato's approach and the previous proposals. We propose a new GMM criterion that is not based on empirical means over the sample as the usual one, but on local averaging, in a manner similar to the GEL estimators referred to above. Our criterion resembles the statistic based on smoothing techniques used in specification testing by Delgado, Dominguez and Lavergne (2005), and so we labeled it as the *smooth GMM* criterion. Our approach defines a whole class of estimators, some similar in spirit to Dominguez and Lobato's estimator when the smoothing parameter is fixed, while others are closer to previous proposals when the smoothing parameter decreases to zero as the sample size increases. We show that consistency and asymptotic normality follows in both cases. As argued above, this property has important empirical implications, since in practice we can never set the smoothing parameter arbitrarily close to zero. However we show that at a first-order, letting the smoothing parameter tend to zero yields a semiparametric efficient estimator. This provides an argument in favor of effectively smoothing the GMM criterion. We also provide a two-step version of our estimator that attains the semiparametric efficiency bound. By contrast to the two-step estimators of Robinson (1987) and Newey (1993), ours does not require estimation of conditional expectations of derivatives.

Our second main contribution is to investigate a dimension-reduction device, as proposed by Lavergne and Patilea (2006) for regression checks, in the context of GMM estimation. We consider restrictions conditional on single linear indices and then integrate over all possible indices. While the resulting estimator cannot attain the semiparametric efficiency bound, its higher-order properties may be preferable. We make this intuition precise by studying an higher-order expansion of our GMM estimator that resembles the one derived by Newey and Smith (2004) for GMM and GEL based on a fixed number of unconditional moment restrictions. This analysis sheds light on the implications of dimension-reduction.

The paper is organized as follows. In Section 2, we define the smooth GMM estimator, we study its asymptotic properties, and we propose a two-step efficient estimator. In Section 3, we introduce our dimension-reduction adaptation of smooth GMM estimators, we study its asymptotic properties and we propose a two-step estimator that asymptotically has minimum variance in our class of estimators. In Section 4, we study the higher-order properties of our smooth GMM estimators.

2 Smooth GMM

2.1 The criterion

Let $g(Z,\theta) = (g^{(1)}(Z,\theta), ..., g^{(r)}(Z,\theta))'$ be a $r-$vector valued function, $r \geq 1$, with $Z = (Y', X')' \in \mathbb{R}^{d+q}$, $d \geq 1$, $q \geq 1$, and $\theta \in \Theta \subset \mathbb{R}^p$, $p \geq 1$. Assume that for a unique $\theta_0 \in \Theta$

$$\mathbb{E}\left[g(Z,\theta_0)|X\right] = 0 \quad \text{a.s.} \tag{2.1}$$

Given an i.i.d. sample $\{Z_1, \dots Z_n\}$ from Z, the smooth GMM criterion in its simpler form writes

$$\frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} g'(Z_i,\theta) g(Z_j,\theta) K_{ij}\,,$$

$$\text{where} \quad K_{ij} = \frac{1}{h^q} K\left(\frac{X_i - X_j}{h}\right)\,, \qquad 1 \leq i \neq j \leq n\,,$$

with a multivariate kernel $K(\cdot)$. This is the test statistic introduced by Dominguez, Delgado and Lavergne (2005), who themselves generalize the statistic of Zheng (1996) and Li and Zhang (1996) for testing regression models, used in more general contexts by Fan and Li (1996), Lavergne and Vuong (2000), Lavergne (2001). In the testing problem, h is a smoothing parameter tending to zero when the sample size increases and the criterion has limit

$$\mathbb{E}\left[g'(Z,\theta)\mathbb{E}\left[g(Z,\theta)|X\right]f(X)\right] = \mathbb{E}\left[\mathbb{E}\left[g'(Z,\theta)|X\right]\mathbb{E}\left[g(Z,\theta)|X\right]f(X)\right]\,,$$

where $f(\cdot)$ is the density of X. Hence, provided we have a consistent estimator for θ_0, the statistic can be used for testing the conditional moment restrictions (2.1). Here we use the statistic mainly for estimation purposes and we thus do not assume the existence of a preliminary consistent estimator.

When h is fixed, our criterion resembles the one proposed by Dominguez and Lobato (2004), which in the univariate case writes

$$\frac{1}{n^3}\sum_{k=1}^{n}\left[\sum_{i=1}^{n} g(Z_i,\theta)\mathbb{I}(X_i \leq X_k)\right]^2 = \frac{1}{n^2}\sum_{i,j=1}^{n} g(Z_i,\theta)g(Z_j,\theta)\left[\frac{1}{n}\sum_{k=1}^{n}\mathbb{I}(X_i \leq X_k)\mathbb{I}(X_j \leq X_k)\right].$$

The main difference with our criterion is that the weight in the double sum depends on all observations of X, which likely introduces more variability in estimation. A more general formulation of the smooth GMM criterion that includes the above one as a special case is analyzed in a companion work (in progress). Since we may want to weigh and combine the different components of $g(\cdot,\theta)$ in a different manner, we introduce a sequence of non-random matrices $W(\cdot)$ and we define a more general criterion as

$$M_n(\theta) = \frac{1}{n(n-1)} \sum_{1\leq i\neq j\leq n} g'(Z_i,\theta) W_n^{-1/2}(X_i) W_n^{-1/2}(X_j) g(Z_j,\theta) K_{ij} \,.$$

2.2 Consistency

Let us begin with the identifiability assumption of θ_0.

Assumption 1 *(i) The parameter space Θ is compact (ii) θ_0 is the unique value in Θ satisfying (2.1).*

Under this assumption, we first study the consistency of

$$\tilde{\theta}_n = \arg\max_{\Theta} M_n(\theta)\,,$$

and show that it is consistent whether h is fixed or tends to zero. To understand why, let us focus on $\mathbb{E}M_n(\theta)$ and assume $W_n(X)$ is the identity matrix for simplicity. Denoting by $\mathcal{F}[l](\cdot)$ the Fourier transform of the function $l(\cdot)$, we have

$$\begin{aligned}
\mathbb{E}M_n(\theta) &= \mathbb{E}\left[g'(Z_1,\theta)g(Z_2,\theta)h^{-q}K\left((X_i-X_j)/h\right)\right] \\
&= (2\pi)^{-q/2}\,\mathbb{E}\left[g'(Z_1,\theta)g(Z_2,\theta)\int_{\mathbb{R}^q}\exp\left(-it(X_1-X_2)\right)\mathcal{F}[K](ht)\,dt\right] \\
&= (2\pi)^{q/2}\sum_{k=1}^{r}\left\{\int_{\mathbb{R}^q}\left|\mathcal{F}\left[\mathbb{E}[g^{(k)}(Z,\theta)|X=\cdot]f(\cdot)\right](t)\right|^2\mathcal{F}[K](ht)\,dt\right\}\,, \qquad (2.2)
\end{aligned}$$

Hence, if $\mathcal{F}[K](\cdot)$ is strictly positive on $\mathbb{R}^q$, $\mathbb{E}M_n(\theta)$ is positive and equals zero iff $\theta=\theta_0$. Indeed, using the unicity of the Fourier transform and Assumption 1,

$$\begin{aligned}
\mathbb{E}M_n(\theta)=0 \;&\Leftrightarrow\; \mathcal{F}\left[\mathbb{E}[g^{(k)}(Z,\theta)|X=\cdot]f(\cdot)\right](t)=0 \quad \forall\, t\in\mathbb{R}^q,\ k=1,\ldots r \\
&\Leftrightarrow\; \mathbb{E}[g^{(k)}(Z,\theta)|X]f(X)=0 \quad \text{a.s.},\ k=1,\ldots r \Leftrightarrow \theta=\theta_0\,.
\end{aligned}$$

Equation (2.2) sheds light on the fundamental reason why a consistent estimator obtains whether h is fixed or tends to zero. The expectation of the GMM criterion indeed accounts for the Fourier transform of $\mathbb{E}[g(Z,\theta)|X]$ on its whole domain, that is at all frequencies. A sufficient condition for consistency is then the strict positivity of the Fourier transform of $K(\cdot)$. It is fulfilled for instance by products of the triangular, normal, Laplace or Cauchy densities, but also by more general kernels, including kernels taking negative values.

It is then clear that $\widehat{\theta}_n$ is consistent for θ_0 provided $\sup_{\theta\in\Theta}|M_n(\theta)-\mathbb{E}M_n(\theta)|=o_{\mathbb{P}}(1)$. This uniform convergence is easily obtained from a uniform law of large numbers for $U-$processes. Let us then make the following assumptions.

Assumption 2 *(i) For all n, $W_n(\cdot)$ is a sequence of $r \times r$ symmetric positive definite non-random matrix functions and there is a symmetric positive definite matrix function $W(\cdot)$ such that* $\sup_u \|W_n(u) - W(u)\|$. *(ii) The families $\mathcal{G}_k = \{W_n^{-1/2}(\cdot)g(\cdot,\theta) : \theta \in \Theta, n = 1, \ldots\}$, $1 \leq k \leq r$, are Euclidean for an envelope F with $\mathbb{E}\|F\|^2 < \infty$.*

Here, the matrix norm $\|\cdot\|$ is the usual extension of the Euclidean norm.

Assumption 3 *$K(\cdot)$ is a symmetric, squared-integrable, bounded function of bounded variation with strictly positive Fourier transform. The integral of $K(\cdot)$ equals one.*

The Euclidean property is a mild one for parametric families of functions. We refer to Nolan and Pollard (1987), Pakes and Pollard (1989) and Sherman (1994) for the definition and properties of Euclidean families. Symmetry of the kernel is not necessary here, but will lead to simpler proofs later on.

Theorem 2.1 *For an i.i.d. sample, under Assumptions 1–3, $\tilde{\theta}_n \xrightarrow{p} \theta_0$ if $nh^{2q} \to \infty$.*

Proof. The family of functions $\{g'(Z_1,\theta)W_n^{-1/2}(X_1)W_n^{-1/2}(X_2)g(Z_2,\theta)K\left((X_1 - X_2)'\beta/h\right) : \theta \in \Theta, \beta \in \mathbb{S}^q, h > 0\}$ is Euclidean for a square-integrable envelope by Assumptions 2 and 3, Lemma 22(ii) of Nolan and Pollard (1987) and Lemma 2.14(ii) of Pakes and Pollard (1989). Then by Corollary 7 of Sherman (1994), $\sup_{\theta\in\Theta} |h^q M_n(\theta) - \mathbb{E}h^q M_n(\theta)| = O_{\mathbb{P}}(n^{-1/2})$. Replacing $g(Z,\theta)$ by $g_n(Z,\theta) \doteq W_n^{-1/2}(X)g(Z,\theta)$ in (2.2) yields

$$
\begin{aligned}
\mathbb{E}M_n(\theta) = 0 \quad &\Leftrightarrow \quad \mathcal{F}\left[\mathbb{E}\left[g_n^{(k)}(Z,\theta)|X = \cdot\right] f(\cdot)\right](t\beta) = 0 \quad \forall\, t \in \mathbb{R},\ \beta \in \mathbb{S}^q,\ k = 1,\ldots r \\
&\Leftrightarrow \quad W_n^{-1/2}(X)\mathbb{E}\left[g(Z,\theta)|X\right] = 0 \quad \text{a.s.} \Leftrightarrow \theta = \theta_0\ ,
\end{aligned}
$$

as $W_n(X)$ is positive definite. Consistency of $\tilde{\theta}_n$ is thus guaranteed by standard arguments whenever $nh^{2q} \to \infty$. ■

2.3 Asymptotic normality

Although asymptotic normality could be shown using a general approach, see e.g. Sherman (1993), we follow a classical one based on the first-order condition, because it will prove useful for further developments.

Assumption 4 *(i) θ_0 belongs to the interior of Θ. (ii) $g(Z,\theta)$ is twice continuously differentiable with respect to θ in a neighborhood $\mathcal{N}$ of θ_0. (i) $W^{-1/2}(X)\mathbb{E}\left[g(Z,\theta)|X\right]$ and $W^{-1/2}(X)\mathbb{E}\left[\nabla'_\theta g(Z,\theta)|X\right]$ are bounded uniformly in norm over $\mathcal{N}$ by a function $l(X)$ such that $\mathbb{E}\left|l(X)\right|^{2+\nu} < \varphi$ for some $\nu > 0$. The elements of $\mathbb{E}\left[\mathrm{H}_\theta\left(W^{-1/2}(X)g(Z,\theta)\right)|X\right]$, $k = 1,\ldots r$, are bounded uniformly over $\mathcal{N}$ by a integrable function.*

Here $\nabla_\theta g(Z,\theta) = \frac{\partial g(Z,\theta)}{\partial\theta}$ and H_θ respectively denote the gradient and hessian operators with respect to θ.

Theorem 2.2 *For an i.i.d. sample, under Assumptions 1–4,* $\sqrt{n}\left(\tilde{\theta}_n - \theta_0\right) \xrightarrow{d} N\left(0, \tilde{\Sigma}\right)$ *whether (a)* $h = 1$ *and*

$$V = \mathbb{E}\left[\mathbb{E}\left[\nabla_\theta g(Z_1,\theta_0)|X_1\right] W^{-1/2}(X_1) W^{-1/2}(X_2) \mathbb{E}\left[\nabla'_\theta g(Z_2,\theta_0)|X_2\right] K\left(X_1 - X_2\right)\right]$$

is non-singular, in which case $\tilde{\Sigma} = \tilde{\Sigma}_a = V^{-1}\Delta V^{-1}$ *with*

$$\begin{aligned} \Delta &= \mathbb{E}\left[\mathbb{E}\left[\nabla_\theta g(Z_1,\theta_0)|X_1\right] W^{-1/2}(X_1) W^{-1/2}(X_2) \mathrm{Var}\left[g(Z_2,\theta_0)|X_2\right] W^{-1/2}(X_2)\right. \\ &\qquad \left. W^{-1/2}(X_3) \mathbb{E}\left[\nabla'_\theta g(Z_3,\theta_0)|X_3\right] K\left(X_1 - X_2\right) K\left(X_2 - X_3\right)\right] ; \end{aligned}$$

or (b) $h \to 0$, $nh^{2p} \to \infty$, *and*

$$V = \mathbb{E}\left[\mathbb{E}\left[\nabla_\theta g(Z,\theta_0)|X\right] W^{-1}(X) \mathbb{E}\left[\nabla'_{\theta'} g(Z,\theta_0)|X\right] f(X)\right]$$

is non-singular, in which case $\tilde{\Sigma} = \tilde{\Sigma}_b = V^{-1}\Delta V^{-1}$ *with*

$$\Delta = \mathbb{E}\left[\mathbb{E}\left[\nabla_\theta g(Z,\theta_0)|X\right] W^{-1}(X) \mathrm{Var}\left[g(Z,\theta_0)|X\right] W^{-1}(X) \mathbb{E}\left[\nabla'_\theta g(Z,\theta_0)|X\right] f^2(X)\right] .$$

Proof. Denote $g_n(Z,\theta) \doteq W_n^{-1/2}(X) g(Z,\theta)$, $G_n(X,\theta) \doteq \mathbb{E}\left[\nabla'_\theta g_n(Z,\theta)|X\right]$, $\Omega_n(X,\theta) \doteq \mathrm{Var}\left[g_n(Z,\theta)|X\right]$, $G(X,\theta)$ and $\Omega(X,\theta)$ their respective limits. From the first-order condition, $\nabla_\theta M_n(\theta_0) = -\mathrm{H}_\theta M_n\left(\bar{\theta}\right)\left(\tilde{\theta}_n - \theta_0\right)$, where $\|\bar{\theta} - \theta_0\| \leq \|\tilde{\theta}_n - \theta_0\|$. Hoeffding's decomposition of U-statistics and the central limit theorem yield $\sqrt{n}\nabla_\theta M_n(\theta_0) \xrightarrow{d} N(0, 4\Delta)$, where $\Delta = \lim_{n\to\infty} \Delta_n$ with

$$\begin{aligned} \Delta_n &= \frac{n}{4}\mathbb{E}\left[\nabla_\theta M_n(\theta_0) \nabla'_\theta M_n(\theta_0)\right] = \frac{\Delta_{1n}}{(n-1)} + \frac{(n-2)}{(n-1)}\Delta_{2n} \\ &= \frac{1}{(n-1)}\mathbb{E}\left[G'_n(X_1,\theta_0)\Omega_n(X_2,\theta_0)G_n(X_1,\theta_0)K_{12}K_{21}\right] \\ &\quad + \frac{(n-2)}{(n-1)}\mathbb{E}\left[G'_n(X_1,\theta_0)\Omega_n(X_2,\theta_0)G_n(X_3,\theta_0)K_{12}K_{23}\right] . \end{aligned}$$

Recall that

$$\mathrm{H}_\theta M_n(\theta) = \frac{2}{n(n-1)}\sum_{i\neq j}\left(\nabla'_\theta g_n(Z_i,\theta)\nabla_\theta g'_n(Z_j,\theta) + \sum_{k=1}^{r}\mathrm{H}_\theta g_n^{(k)}(Z_i,\theta) g_n(Z_j,\theta)\right) K_{ij} .$$

(a) If $h = 1$, then Δ_{1n} tends to some finite limit, so that Δ is the one given in the Theorem. Moreover, by the uniform law of large numbers and the consistency of $\tilde{\theta}_n$, $\mathrm{H}_\theta M_n\left(\bar{\theta}\right) \xrightarrow{p} 2V$.

(b) If h tends to zero, then using Fourier transforms and the convolution law,

$$\begin{aligned}
\Delta_{2n} &= \mathbb{E}\left[G_n'(X_1,\theta_0)\Omega_n(X_2,\theta_0)G_n(X_3,\theta_0)K_{12}K_{23}\right] \\
&= (2\pi)^{-q}\,\mathbb{E}\Big[G_n'(X_1,\theta_0)\Omega_n(X_2,\theta_0)G_n(X_3,\theta_0) \\
&\qquad \int\!\!\int \exp\left(-it\left(X_1-X_2\right)\right)\exp\left(-iu\left(X_2-X_3\right)\right)\mathcal{F}\left[K\right](ht)\mathcal{F}\left[K\right](hu)\,dt\,du\Big] \\
&= (2\pi)^{q/2}\int\!\!\int \mathcal{F}\left[G_n'f(\cdot,\theta_0)\right](-t)\mathcal{F}\left[\Omega_n f(\cdot,\theta_0)\right](t-u) \\
&\qquad \mathcal{F}\left[G_n f(\cdot,\theta_0)\right](u)\mathcal{F}\left[K\right](ht)\mathcal{F}\left[K\right](hu)\,dt\,du \\
&\to (2\pi)^{-q/2}\int\!\!\int \mathcal{F}\left[G'f(\cdot,\theta_0)\right](-t)\mathcal{F}\left[\Omega f(\cdot,\theta_0)\right](t-u)\mathcal{F}\left[Gf(\cdot,\theta_0)\right](u)\,dt\,du \\
&= (2\pi)^{q/2}\,\mathcal{F}\left[G'f\Omega fGf(\cdot,\theta_0)\right](0) = \mathbb{E}\left[G(X,\theta_0)\Omega(X,\theta_0)G'(X,\theta_0)f^2(X)\right] = \Delta\,.
\end{aligned}$$

Similarly $\mathrm{H}_\theta Q_n(\theta) \xrightarrow{p} 2V$. Last,

$$h^q\Delta_{1n} = \mathbb{E}\left[\nabla_\theta' g_n(Z_1,\theta_0)g_n(Z_2,\theta)g_n'(Z_2,\theta)\nabla_\theta' g_n(Z_1,\theta_0)h^{-q}K^2\left((X_1-X_2)/h\right)\right]\,,$$

which tends to a finite limit using the same arguments as for Δ_{2n}. ■

2.4 Efficiency

We are now ready to study the variances of the different estimators.

Theorem 2.3 *Under the assumptions of Theorem 2.1*
(i) The semiparametric efficiency bound

$$\left\{\mathbb{E}\left[\mathbb{E}[\nabla_\theta g(Z,\theta_0)|X]\mathrm{Var}^{-1}[g(Z,\theta_0)|X]\mathbb{E}[\nabla_\theta' g(Z,\theta_0)|X]\right]\right\}^{-1}$$

is attained by $\tilde\theta_n$ in Case (b) when $W^{-1}(X) = \mathrm{Var}\left[g(Z,\theta_0)|X\right]f(X)$.
(ii) The semiparametric efficiency bound is attained by $\tilde\theta_n$ in Case (a) only if it exists a matrix $W(\cdot)$ such that

$$\begin{aligned}
&\mathbb{E}[\nabla_\theta' g(Z_2,\theta_0)|X_2] \\
&\quad = \mathrm{Var}\left[g(Z_2,\theta_0)|X_2\right]W^{-1/2}(X_2)\mathbb{E}\left[W^{-1/2}(X_1)\nabla_\theta' g(Z_1,\theta_0)K\left(X_1-X_2\right)|X_2\right]\Delta^{-1}V\,.
\end{aligned}$$

Proof. Part (i) is immediate. The proof of (ii) relies on the following lemma, which is a generalization of Tripathi (1999).

Lemma 2.1 *Let $A \in \mathbb{R}^{n\times p}$ and $B \in \mathbb{R}^{n\times q}$ be random matrices such that $\mathbb{E}\|A\|^2 < \infty$, $\mathbb{E}\|B\|^2 < \infty$, and $\mathbb{E}\left(A'A\right)$ is non-singular. Then $\mathbb{E}\left(B'B\right) - \mathbb{E}\left(B'A\right)\mathbb{E}^{-1}\left(A'A\right)\mathbb{E}\left(A'B\right)$ is positive semidefinite, with equality iff $B = A\mathbb{E}^{-1}\left(A'A\right)\mathbb{E}\left(A'B\right)$.*

To prove the lemma, consider $\Lambda = \mathbb{E}^{-1}(A'A)\,\mathbb{E}(A'B) \in \mathbb{R}^{p\times q}$. Then

$$\mathbb{E}\left[(B - A\Lambda)'(B - A\Lambda)\right] = \mathbb{E}(B'B) - \mathbb{E}(B'A)\,\mathbb{E}^{-1}(A'A)\,\mathbb{E}(A'B)$$

is positive semidefinite by definition, and is zero only when $B = A\Lambda$.
To prove (ii), take $B = \text{Var}^{-1/2}[g(Z,\theta_0)|X]\mathbb{E}[\nabla'_\theta g(Z,\theta_0)|X]$ and

$$A = \text{Var}^{1/2}\left[g(Z,\theta_0)|X\right]W^{-1/2}(X)\mathbb{E}\left[W^{-1/2}(X_1)\nabla'_\theta g(Z_1,\theta_0)K\left(X_1 - X\right)|X\right] .$$

This choice gives the semiparametric efficiency bound as $\mathbb{E}^{-1}(B'B)$, and the bound is attained for $B = A\Lambda$ where $\Lambda = \Delta^{-1}V$. ■

We now propose a two-step estimator that attains the efficiency bound.

3 Dimension reduction

3.1 The principle

Since the dimension of X may adversely affect the properties of the estimator, we propose here another GMM criterion that aims to address this issue. To reduce the dimensionality of the problem, we use the following result, which states that constancy of a conditional expectation on X is equivalent to constancy of all expectations conditional on linear indices of X. A proof is found in Lavergne and Patilea (2007), but the result can also be deduced from Theorem 1 of Bierens (1982).

Lemma 3.1 *Let $X \in \mathbb{R}^q$ and $U \in \mathbb{R}^c$ be random vectors, with $\mathbb{E}\|U\| < \infty$.*
$\mathbb{E}(U \mid X) = \mathbb{E}(U)$ a.s. $\iff \mathbb{E}(U \mid X'\beta) = \mathbb{E}(U)$ a.s. $\forall \beta \in \mathbb{R}^q : \|\beta\| = 1$.

This result is used among others by Bierens (1982), Bierens and Ploberger (1997), Escanciano (2006), and Lavergne and Patilea (2006, 2007) to build consistent tests of parametric regression functions. As a straightforward consequence, we obtain the following corollary.

Corollary 3.2 *Consider random vectors $X \in \mathbb{R}^q$ and $U(\theta) \in \mathbb{R}^c$ depending on a parameter $\theta \in \Theta$, such that $\mathbb{E}\|U(\theta)\| < \infty$ for all θ. Suppose that $f_\beta(X'\beta)$, the density of $X'\beta$, exists for all β of norm unity and that $\mathbb{E}\left[U(\theta)|X'\beta\right]f_\beta(X'\beta)$ exists and is squared-integrable. Then $\mathbb{E}(U(\theta) \mid X) = 0$ a.s. is equivalent to*

$$\int_{\mathbb{S}^q} \mathbb{E}\left[U'(\theta)\mathbb{E}\left(U(\theta)|X'\beta\right)f_\beta(X'\beta)\right]d\beta = 0\,, \tag{3.3}$$

where $\mathbb{S}^q$ is the hypersphere of radius one and $d\beta$ is the uniform measure on $\mathbb{S}^q$.

The statement remains true if $U(\theta)$ is replaced by $W^{-1/2}(X'\beta)U(\theta)$ for any matrix function $W(\cdot)$ defined on the support of $X'\beta$ such that $W(X'\beta)$ is symmetric definite positive for all $X'\beta$.

For a moment, let us choose a single direction β and look for θ_0 that fulfills

$$\mathbb{E}\left[g(Z,\theta_0)|X'\beta\right] = 0, \qquad a.s.$$

An immediate modification of $M_n(\theta)$ yields

$$\frac{1}{n(n-1)} \sum_{1\leq i\neq j\leq n} g'(Z_i,\theta)W_n^{-1/2}(X_i'\beta)W_n^{-1/2}(X_j'\beta)g(Z_j,\theta)\frac{1}{h}K\left(\frac{(X_i-X_j)'\beta}{h}\right)$$

as a criterion to minimize, where $K(\cdot)$ is now a univariate kernel. Now integrate on the hypersphere to get

$$Q_n(\theta) = \frac{1}{n(n-1)} \sum_{1\leq i\neq j\leq n} \mathbb{E}_{\mathbb{S}^q}\left[g'(Z_i,\theta)W_n^{-1/2}(X_i'\beta)W_n^{-1/2}(X_j'\beta)g(Z_j,\theta)K_{ij\beta}\right], \tag{3.4}$$

$$\text{where} \quad K_{ij\beta} = \frac{1}{h}K\left(\frac{(X_i-X_j)'\beta}{h}\right), \qquad 1\leq i\neq j\leq n,$$

and $\mathbb{E}_{\mathbb{S}^q}h(\beta) = \int_{\mathbb{S}^q} h(\beta)\,d\beta$.

3.2 Consistency

The analysis of the consistency of $\widehat{\theta}_n = \arg\max_{\Theta} Q_n(\theta)$ is similar the one of $\tilde{\theta}_n$. That consistency requires weaker the weaker restriction $nh^2 \to \infty$ is a first illustration of the consequences of dimension-reduction.

Theorem 3.3 *For an i.i.d. sample, under Assumptions 1–3, $\widehat{\theta}_n \xrightarrow{p} \theta_0$ if $nh^2 \to \infty$.*

Proof. The proof follows the same lines as before, modifying (2.2) by letting $g_{n\beta}(X'\beta,\theta) \doteq \mathbb{E}\left[g(Z,\theta)|X'\beta\right] W_n^{-1/2}(X'\beta)$ to obtain

$$\begin{aligned}
\mathbb{E}Q_n(\theta) &= \mathbb{E}\left\{\mathbb{E}_{\mathbb{S}^q}\left[g_{n\beta}'(X_i'\beta,\theta)g_{n\beta}(X_j'\beta,\theta)K_{ij\beta}\right]\right\} \\
&= (2\pi)^{-1/2}\,\mathbb{E}_{\mathbb{S}^q}\left\{\mathbb{E}\left[g_{n\beta}'(X_1'\beta,\theta)g_{n\beta}(X_2'\beta,\theta)h^{-1}\int \exp\left(-it(X_1-X_2)'\beta\right)\mathcal{F}\left[K\right](ht)\,dt\right]\right\} \\
&= (2\pi)^{1/2}\,\mathbb{E}_{\mathbb{S}^q}\sum_{k=1}^{r}\left\{\int\left|\mathcal{F}\left[g_{n\beta}^{(k)}(X'\beta,\theta)f_\beta(\cdot)\right](t)\right|^2\mathcal{F}\left[K\right](ht)\,dt\right\}. \blacksquare
\end{aligned}$$

3.3 Asymptotic normality

The following assumption replaces Assumption 4 in the forthcoming result.

Assumption 4' *(i)* θ_0 *belongs to the interior of* Θ. *(ii)* $g(Z,\theta)$ *is twice continuously differentiable with respect to* θ *in a neighborhood* $\mathcal{N}$ *of* θ_0. *(i)* $W^{-1/2}(X'\beta)\mathbb{E}\left[g(Z,\theta)|X'\beta\right]$ *and* $W^{-1/2}(X'\beta)\mathbb{E}\left[\nabla'_\theta g(Z,\theta)|X'\beta\right]$ *are bounded uniformly in norm over* $\mathcal{N}$ *and* $\mathbb{S}^q$ *by a function* $l(X'\beta)$ *such that* $\mathbb{E}\left|l(X'\beta)\right|^{2+\nu} < \varphi$ *for some* $\nu > 0$ *and for all* $\beta \in \mathbb{S}^q$. *The elements of* $\mathbb{E}\left[\mathrm{H}_\theta\left(W^{-1/2}(X'\beta)g(Z,\theta)\right)|X'\beta\right]$, $k = 1,\ldots r$, *are bounded uniformly over* $\mathcal{N}$ *and* $\mathbb{S}^q$ *by a integrable function.*

Theorem 3.4 *For an i.i.d. sample, under Assumptions 2–4',* $\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \xrightarrow{d} N\left(0,\Sigma\right)$ *whether (a)* $h = 1$ *and*

$$\begin{aligned} V &= \mathbb{E}_{\mathbb{S}^q}\left\{\mathbb{E}\left[\mathbb{E}\left[\nabla'_\theta g(Z_1,\theta_0)|X_1'\beta\right]W^{-1/2}(X_1'\beta)\right.\right. \\ &\qquad \left.\left. W^{-1/2}(X_2'\beta)\mathbb{E}\left[\nabla_\theta g(Z_2,\theta_0)|X_2'\beta\right]K\left((X_1-X_2)'\beta\right)\right]\right\} \end{aligned}$$

is non-singular, in which case $\Sigma = \Sigma_a = V^{-1}\Delta V^{-1}$, *where*

$$\begin{aligned} \Delta &= \mathbb{E}_{\mathbb{S}^q}\mathbb{E}_{\mathbb{S}^q}\left\{\mathbb{E}\left[\mathbb{E}\left[\nabla'_\theta g(Z_1,\theta_0)|X_1'\beta\right]W^{-1/2}(X_1'\beta)W^{-1/2}(X_2'\beta)\mathrm{Var}\left[g(Z_2,\theta_0)|X_2'\beta, X_2'\alpha\right]\right.\right. \\ &\qquad \left.\left. W^{-1/2}(X_2'\alpha)W^{-1/2}(X_3'\alpha)\mathbb{E}\left[\nabla_\theta g(Z_3,\theta_0)|X_3'\alpha\right]K\left((X_1-X_2)'\beta\right)K\left((X_2-X_3)'\alpha\right)\right]\right\}\ ; \end{aligned}$$

or (ii) $h \to 0$, $nh^2 \to \infty$, *and*

$$V = \mathbb{E}_{\mathbb{S}^q}\left\{\mathbb{E}\left[\mathbb{E}\left[\nabla'_\theta g(Z,\theta_0)|X'\beta\right]W^{-1}(X'\beta)\mathbb{E}\left[\nabla_\theta g(Z,\theta_0)|X'\beta\right]f_\beta(X'\beta)\right]\right\}$$

is non-singular, in which case $\Sigma = \Sigma_b = V^{-1}\Delta V^{-1}$, *where*

$$\begin{aligned} \Delta &= \mathbb{E}_{\mathbb{S}^q}\left\{\mathbb{E}\left[\mathbb{E}\left[\nabla'_\theta g(Z,\theta_0)|X'\beta\right]W^{-1}(X'\beta)\mathrm{Var}\left[g(Z,\theta_0)|X'\beta\right]\right.\right. \\ &\qquad \left.\left. W^{-1}(X'\beta)\mathbb{E}\left[\nabla_\theta g(Z,\theta_0)|X'\beta\right]f_\beta^2(X'\beta)\right]\right\}\ . \end{aligned}$$

Proof. The proof follows Theorem 2.2's lines with the following differences. Denote $G_{n\beta}(X'\beta,\theta) \doteq \nabla'_\theta g_{n\beta}(X'\beta,\theta)$, $\Omega_{n\beta\alpha}(X'\beta,X'\alpha,\theta) \doteq \mathbb{E}\left[g_{n\beta}(X'\beta,\theta)g'_{n\alpha}(X'\alpha,\theta)|X'\beta,X'\alpha\right]$, $G_\beta(X'\beta,\theta)$ and $\Omega_{\beta,\alpha}(X'\beta,X'\alpha,\theta)$ their respective limits, and $f_{\beta\alpha}(X'\beta,X'\alpha)$ the joint density of $(X'\beta,X'\alpha)$. We have

$$\begin{aligned} \mathrm{H}_\theta Q_n\left(\theta\right) &= \frac{2}{n(n-1)}\sum_{i\neq j}\mathbb{E}_{\mathbb{S}^q}\left\{\left(\nabla_\theta g'_{n\beta}(Z_i,\theta)\nabla_\theta g'_{n\beta}(Z_j,\theta)\right.\right. \\ &\qquad \left.\left. + \sum_{k=1}^{r}\mathrm{H}_\theta g_{n\beta}^{(k)}(Z_i,\theta)g_{n\beta}(Z_j,\theta)\right)K_{ij\beta}\right\} \end{aligned}$$

and $\Delta_n = \frac{n}{4}\mathbb{E}\left[\nabla_\theta Q_n(\theta_0)\nabla_{\theta'}Q_n(\theta_0)\right] = \frac{\Delta_{1n}}{(n-1)} + \frac{(n-2)}{(n-1)}\Delta_{2n}$

$$= \frac{1}{(n-1)}\mathbb{E}_{\mathbb{S}^q}\mathbb{E}_{\mathbb{S}^q}\left\{\mathbb{E}\left[\nabla_\theta g'_{n\beta}(Z_1,\theta_0)g_{n\beta}(Z_2,\theta)g'_{n\alpha}(Z_2,\theta)\nabla'_\theta g_{n\alpha}(Z_1,\theta_0)K_{12\beta}K_{21\alpha}\right]\right\}$$
$$+\frac{(n-2)}{(n-1)}\mathbb{E}_{\mathbb{S}^q}\mathbb{E}_{\mathbb{S}^q}\left\{\mathbb{E}\left[\nabla_\theta g'_{n\beta}(Z_1,\theta_0)g_{n\beta}(Z_2,\theta)g'_{n\alpha}(Z_2,\theta)\nabla'_\theta g_{n\alpha}(Z_3,\theta_0)K_{12\beta}K_{23\alpha}\right]\right\}.$$

(a) If $h=1$, we obtain the desired result.

(b) If $h\to 0$, then using Fourier transforms and the convolution law,

$$\begin{aligned}
\Delta_{2n} &= \mathbb{E}_{\mathbb{S}^q}\mathbb{E}_{\mathbb{S}^q}\left\{\mathbb{E}\left[G_{n\beta}(X'_1\beta,\theta_0)\Omega_{n\beta\alpha}(X'_2\beta,X'_2\alpha,\theta_0)G'_{n\alpha}(X'_3\alpha,\theta_0)K_{12\beta}K_{23\alpha}\right]\right\}\\
&= (2\pi)\,\mathbb{E}_{\mathbb{S}^q}\mathbb{E}_{\mathbb{S}^q}\left\{\int\int\left[\mathcal{F}\left[G_{n\beta}f_\beta(\cdot,\theta_0)\right](-t)\mathcal{F}\left[\Omega_{n\beta\alpha}f_{\beta\alpha}(\cdot,\cdot,\theta_0)\right](t,-u)\right.\right.\\
&\qquad \left.\left.\mathcal{F}\left[G'_{n\alpha}f_\alpha(\cdot,\theta_0)\right](u)\mathcal{F}[K](ht)\mathcal{F}[K](hu)\,dt\,du\right]\right\}\\
&\to \mathbb{E}_{\mathbb{S}^q}\mathbb{E}_{\mathbb{S}^q}\left\{\int\int\left[\mathcal{F}\left[G_\beta f_\beta(\cdot,\theta_0)\right](-t)\mathcal{F}\left[\Omega_{\beta\alpha}f_{\beta\alpha}(\cdot,\cdot,\theta_0)\right](t,-u)\right.\right.\\
&\qquad \left.\left.\mathcal{F}\left[G'_\alpha f_\alpha(\cdot,\theta_0)\right](u)\,dt\,du\right]\right\}\\
&= \mathbb{E}_{\mathbb{S}^q}\mathbb{E}_{\mathbb{S}^q}\left\{\mathcal{F}\left[G_\beta f_\beta\Omega_{\beta\alpha}f_{\beta\alpha}G'_\alpha f_\alpha(\cdot,\cdot,\theta_0)\right](0,0)\right\}\\
&= \mathbb{E}_{\mathbb{S}^q}\mathbb{E}_{\mathbb{S}^q}\left\{\mathbb{E}\left[G_\beta(X'\beta,\theta_0)f_\beta(X'\beta)\Omega_{\beta,\alpha}(X'\beta,X'\alpha,\theta_0)G'_\alpha(X'\alpha,\theta_0)f_\alpha(X'\alpha)\right]\right\}\\
&= \mathbb{E}_{\mathbb{S}^q}\left\{\mathbb{E}\left[G_\beta(X'\beta,\theta_0)f_\beta(X'\beta)\mathbb{E}_{\mathbb{S}^q}\left[\Omega_{\beta,\alpha}(X'\beta,X'\alpha,\theta_0)G'_\alpha(X'\alpha,\theta_0)f_\alpha(X'\alpha)\right]\right]\right\}\\
&= \mathbb{E}_{\mathbb{S}^q}\left\{\mathbb{E}\left[G_\beta(X'\beta,\theta_0)f_\beta(X'\beta)\Omega_\beta(X'\beta,\theta_0)G'_\beta(X'\beta,\theta_0)f_\beta(X'\beta)\right]\right\} = \Delta\,,
\end{aligned}$$

where $\Omega_\beta(X'\beta,\theta) = \mathrm{Var}\left[W^{-1/2}(X'\beta)g(Z,\theta)|X'\beta\right]$. Similarly

$$\mathrm{H}_\theta Q_n(\theta) \xrightarrow{p} 2\mathbb{E}_{\mathbb{S}^q}\left\{\mathbb{E}\left[G_\beta(X'\beta,\theta_0)G'_\beta(X'\beta,\theta_0)f_\beta(X'\beta)\right]\right\} = 2V\,.$$

Last, by Cauchy-Schwarz,

$$\begin{aligned}
0\le h\Delta_{1n} &\le \mathbb{E}_{\mathbb{S}^q}\mathbb{E}_{\mathbb{S}^q}\left\{\mathbb{E}\left[\nabla'_\theta g(Z_1,\theta_0)W_n^{-1/2}(X'_1\beta)W_n^{-1/2}(X'_2\beta)g(Z_2,\theta)\right.\right.\\
&\qquad \left.\left.g'(Z_2,\theta)W_n^{-1/2}(X'_2\beta)W_n^{-1/2}(X'_1\beta)\nabla_\theta g(Z_1,\theta_0)h^{-1}K^2\left((X_1-X_2)'\beta/h\right)\right]\right\},
\end{aligned}$$

which tends to a finite limit using the same arguments as for Δ_{2n}. ■

3.4 Efficiency

Theorem 3.5 *Under the assumptions of Theorem 3.3,*
(i) $W^{-1}(X'\beta) = \mathrm{Var}\left[g(Z,\theta_0)|X'\beta\right]f_\beta(X'\beta)$ *yields the minimum possible asymptotic variance for* $\widehat{\theta}_n$*, which writes*

$$\left\{\mathbb{E}_{\mathbb{S}^q}\left(\mathbb{E}\left[\mathbb{E}[\nabla_\theta g'(Z,\theta_0)|X'\beta]\mathrm{Var}^{-1}[g(Z,\theta_0)|X'\beta]\mathbb{E}[\nabla'_\theta g(Z,\theta_0)|X'\beta]\right]\right)\right\}^{-1}.$$

(ii) The bound in (i) is attained by $\widehat{\theta}_n$ in Case (a) only if it exists a matrix $W(\cdot)$ such that

$$\mathbb{E}[\nabla'_\theta g(Z_2,\theta_0)|X'_2\alpha] = \mathrm{Var}^{1/2}\left[g(Z_2,\theta_0)|X'_2\beta, X'_2\alpha\right] W^{-1/2}(X'_2\alpha)$$
$$\mathbb{E}\left[W^{-1/2}(X'_1\alpha)\nabla'_\theta g(Z_1,\theta_0)K\left((X_1-X_2)'\alpha\right)|X'_2\beta, X'_2\alpha\right]\Delta^{-1}V\,.$$

Proof. Part (i) follows from Lemma 2.1 by choosing $B = \mathrm{Var}^{-1/2}[g(Z,\theta_0)|X'\beta]\mathbb{E}[\nabla'_\theta g(Z,\theta_0)|X'\beta]$ and

$$A = \mathrm{Var}^{1/2}\left[g(Z,\theta_0)|X'\beta\right] W^{-1}(X'\beta)\mathbb{E}\left[\nabla'_\theta g(Z,\theta_0)|X'\beta\right] f_\beta(X'\beta)\,.$$

Part (ii) follows similarly with $B = \mathrm{Var}^{-1/2}[g(Z_2,\theta_0)|X'_2\beta, X'_2\alpha]\mathbb{E}[\nabla'_\theta g(Z_2,\theta_0)|X'_2\alpha]$ and A is

$$\mathrm{Var}^{1/2}\left[g(Z_2,\theta_0)|X'_2\beta, X'_2\alpha\right] W^{-1/2}(X'_2\alpha)\mathbb{E}\left[W^{-1/2}(X'_1\alpha)\nabla'_\theta g(Z_1,\theta_0)K\left((X_1-X_2)'\alpha\right)|X'_2\beta, X'_2\alpha\right]\,.$$

The bound is attained for $B = A\Lambda$ where $\Lambda = \Delta^{-1}V$. ■

REFERENCES

Antoine, B., Bonnal, H., and Renault, E. (2007). On the efficient use of the informational content of estimating equations: Implied probabilities and Euclidean empirical likelihood. *J. Econometrics* **138(2)**, 461-487.

Bierens, H.J. (1982). Consistent model specification tests. *J. Econometrics* **20**, 105-134.

Bierens, H.J. and Ploberger W. (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica* **65**, 1129-1151.

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

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *J. Econometrics* **34(3)**, 305-334.

Delgado, M.A., Dominguez, M.A., and Lavergne, P. (2006). Consistent tests of conditional moment restrictions. *Ann. Econom. Statist.* **81 (1)**, 33–67.

Dominguez, M.A., and Lobato, I.N. (2004). Consistent estimation of models defined by conditional moment restrictions. *Econometrica* **72(5)**, 1601-1615.

Dominguez, M.A., and Lobato, I.N. (2007). A consistent specification test for models defined by conditional moment restrictions. Working paper, ITAM.

Donald, S.G., Imbens, G.W., Newey, W.K. (2003). Empirical likelihood estimation and consistent tests with conditional moment restrictions. *J. Econometrics* **117(1)**, 55-93.

Escanciano, J.C. (2006). A consistent diagnostic test for regression models using projections. *Econometric Theory* **22(6)**, 1030-1051.

Fan, Y., and Li, Q. (1996). Consistent model specification tests: omitted variables and semiparametric functional forms. *Econometrica* **64**, 865-890.

Hansen, L.P. (1982). Large sample properties of Generalized-Method of Moments estimators. *Econometrica* **50(4)**, 1029-1054.

Hansen, L.P. (1985). A method for calculating bounds on the asymptotic covariance matrices of Generalized-Method of Moments estimators. *J. Econometrics* **30(1-2)**, 203-238.

Hansen L.P., and Singleton K.J. (1982). Generalized instrumental variables estimation of non-linear rational-expectations models. *Econometrica* **50(5)**, 1269-1286.

Kitamura, Y., Tripathi, G., and Ahn, H. (2004). Empirical likelihood-based inference in conditional moment restriction models. *Econometrica* **72(6)**, 1667-1714.

Lavergne, P., and V. Patilea (2006). Breaking the curse of dimensionality in nonparametric testing. *J. Econometrics* forthcoming.

Lavergne, P. (2001). An equality test across nonparametric regressions. *J. Econometrics* **103**, 307-344.

Lavergne, P., and Vuong, Q. (2000). Nonparametric significance testing. *Econometric Theory* **16**, 576-601.

Lavergne, P., and V. Patilea (2007). One for all and all for one: Dimension reduction for regression checks. Working paper, Simon Fraser University.

Li, Q., and S. Wang (1998). A simple consistent bootstrap test for a parametric regression function. *J. Econometrics* **87 (1)**, 145–165.

Newey, W.K. (1990). Efficient instrumental variables estimation of nonlinear models *Econometrica* **58(4)**, 809-837.

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

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

Nolan, D., and Pollard, D. (1987). $U-$processes : Rates of convergence. *Ann. Statist.* **15**, 780-799.

Pakes, A., and Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica* **57**, 1027-1057.

Robinson, P.M. (1987). Asymptotically efficient estimation in the presence of heteroscedasticity of unknown form. *Econometrica* **55**, 875-891.

Smith, R.J. (2007). Efficient information theoretic inference for conditional moment restrictions. *J. Econometrics* **138(2)**, 430-460.

Sherman, R.P. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica* **61**, 123-137.

Sherman, R.P. (1994a). Maximal inequalities for degenerate $U-$processes with applications to optimization estimators. *Ann. Statist.* **22**, 439-459.

Sherman, R.P. (1994b). U-processes in the analysis of a generalized semiparametric regression estimator. *Econometric Theory* **10**, 372-395.

Tripathi, G. (1999). A matrix extension of the Cauchy-Schwarz inequality. *Economics Letters* **63**, 1-3.

Zheng, J.X. (1996). A consistent test of functional form via nonparametric estimation techniques. *J. Econometrics* **75 (2)**, 263–289.