Journal of Econometrics 156 (2010) 284–303

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Simultaneous selection and weighting of moments in GMM using a trapezoidal kernel

Ivan A. Canay *

Department of Economics, Northwestern University, USA

ARTICLE INFO

Article history:
Received 11 March 2008
Received in revised form
23 June 2009
Accepted 27 October 2009
Available online 10 November 2009

JEL classification:
C13
C21

Keywords:
Optimal instruments
Flat-top kernels
Empirical likelihood bootstrap
GMM

ABSTRACT

This paper proposes a novel procedure to estimate linear models when the number of instruments is large. At the heart of such models is the need to balance the trade off between attaining asymptotic efficiency, which requires more instruments, and minimizing bias, which is adversely affected by the addition of instruments. Two questions are of central concern: (1) What is the optimal number of instruments to use? (2) Should the instruments receive different weights? This paper contains the following contributions toward resolving these issues. First, I propose a kernel weighted generalized method of moments (GMM) estimator that uses a trapezoidal kernel. This kernel turns out to be attractive to select and weight the number of moments. Second, I derive the higher order mean squared error of the kernel weighted GMM estimator and show that the trapezoidal kernel generates a lower asymptotic variance than regular kernels. Finally, Monte Carlo simulations show that in finite samples the kernel weighted GMM estimator performs on par with other estimators that choose optimal instruments and improves upon a GMM estimator that uses all instruments.

1. Introduction

Many applied economics settings involve instrumental variable (IV) models in which the number of instruments (or moment conditions) becomes large. A fundamental dilemma at the root of such estimation problems arises in the form of a trade off between asymptotic variance and finite sample bias. As discussed both in Morimune (1983) and Bekker (1994), asymptotic efficiency requires the use of all valid moments while the finite sample properties of the estimators are sensitive to the set of instruments actually used. In this framework two questions are of main concern: (1) what is the optimal number of instruments to use? and (2) should the instruments receive different weights? This paper proposes a generalized method of moments (GMM) estimator to deal with such settings. The main contributions of the paper are the following: (i) the use of a trapezoidal kernel to select and weight the number of moments, and (ii) the derivation of the higher order mean squared error of the kernel weighted GMM estimator. The paper also suggests the bootstrap as a feasible data-dependent bandwidth selection rule.

The proposed GMM estimator builds conceptually on the IV estimators of Donald and Newey (2001) and Okui (submitted for publication). Donald and Newey (2001) choose instruments to minimize the higher order MSE and show that this avoids cases where asymptotic inferences are poor due to large bias. In contrast, Okui (submitted for publication) divides the instruments into a set of main instruments and a set of less important instruments. A shrinkage factor is assigned to the instruments in the less important set which is chosen to minimize the higher order MSE. My estimator captures both ideas by using a kernel weighted GMM estimator as in Kuersteiner (2002). The novel feature is the introduction of a trapezoidal kernel to perform two tasks simultaneously. First, the kernel selects the optimal number of instruments; second, it partitions the selected instruments into two classes: one group of instruments that receive weights equal to one (full weight), and a second group of instruments that receive only partial weight.

Fig. 1 displays the trapezoidal kernel and illustrates why it offers an appealing tool for the selection of instruments. In the diagram, a number M of the instruments have nonzero weight and a fraction c of those have full weight. As the figure illustrates, when $c = 1$ the kernel reduces to a truncated kernel and thus represents the case of pure selection. In contrast, when $c = 0$ the trapezoidal kernel reduces to a Bartlett kernel which assigns a different weight to each of the M relevant instruments. In addition to conceptual simplicity, flat-top kernels are known to have favorable asymptotic properties in spectral density and probability density estimation (see Politis and Romano, 1995, 1999; Politis, 2007). This fact suggests that there might also exist possible asymptotic gains to using such kernels in the current setting. I explore this intuition formally and find the sense in which it is true both theoretically and in Monte Carlo simulations.

* Corresponding address: Department of Economics, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208, USA. Tel.: +1 847 491 2929.
E-mail address: iacanay@northwestern.edu.

0304-4076/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2009.10.036

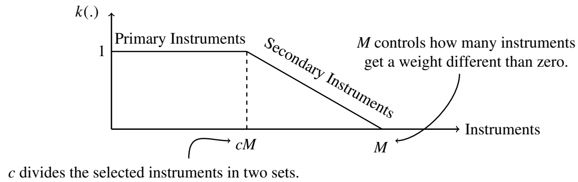

Fig. 1. Trapezoidal kernel.

A second contribution of this paper is to derive the higher order MSE of the kernel GMM estimator when a trapezoidal kernel is used. The derived expression shows a fundamental variance-bias trade off which can be balanced both in choosing the number of instruments M and in deciding the optimal degree of flatness, c. The approach in this part is similar in spirit to Kuersteiner (2002), though the derivation and the resulting expansion differ substantially due to particular features of the trapezoidal kernel.[1]The new approximation, however, causes a second complication: a plug-in bandwidth selection rule is no longer possible due to constants that do not have feasible sample counterparts.

To deal with the difficulty arising from not being able to use a plug-in rule, I use an empirical likelihood bootstrap procedure to select the bandwidth M and the constant c. The idea to use the bootstrap to select moment conditions has been also used by Inoue (2006), where moment conditions are chosen by minimizing approximate coverage error probabilities of confidence regions. A formal justification of the bootstrap properties is beyond the scope of this paper, but Inoue (2006) discusses the formal properties in a similar context. In the current work, the bootstrap allows for feasible implementation of the selection procedure. Evidence on its performance is provided by simulations in the last section.

In addition to the contributions discussed above, the proposed kernel weighted GMM estimator has several attractive features. First, selection of instruments to use along with a partition of these instruments into a full and partial weight subsets is all obtained simultaneously. Second, the estimators by Donald and Newey (2001) and Okui (submitted for publication) turn out to be particular cases of this more general approach. Finally, the estimator allows for the presence of heteroskedasticity.[2] On the other hand, a limitation of the method proposed here is that it requires the practitioner to order (or rank) the instruments according to their importance before implementing the selection procedure.

Another estimator related to the ideas presented here is the random effect quasi-maximum likelihood introduced by Chamberlain and Imbens (2004). In that paper the authors put a random coefficient structure on the relation between the single endogenous regressor and the many instruments. Since the first stage coefficients are thought as independent draws from the same normal distribution, the likelihood function depends on a small set of parameters that does not increase with the number of instruments. The estimator divides the instruments in two sets (strong and weak) and therefore can be categorized as a shrinkage method. The approach has many attractive features but it only covers situations in which there is homoskedasticity and only one endogenous variable. The method proposed here, on the other hand, is applicable when there are multiple endogenous variables and heteroskedasticity. Finally, the results of the present paper differ from those in Hansen et al. (2006), Han and Phillips (2006) and Hausman et al. (2009), who consider asymptotic theory that allows for an increasing number of instruments but do not actually select the number of instruments.

The outline of the paper is as follows. Section 2 presents the model and briefly discusses two of the current alternatives for the selection and weighting of moments. Section 3 introduces the kernel weighted GMM estimator paying particular attention to the role of the trapezoidal kernel and its effect on the asymptotic MSE. Section 4 describes the bootstrap procedure to determine the data-dependent bandwidth rule. Section 5 presents Monte Carlo simulations and Section 6 concludes.

2. The framework: Selection versus shrinkage

The model I focus on in this paper is the one considered by Donald and Newey (2001). Let $\{\varpi_i' = (y_i, Y_i', x_i') : i \le n\}$ be a random sample from the model,

$$y_i = \gamma_0' Y_i + \beta_0' z_i + \varepsilon_i = \delta_0' W_i + \varepsilon_i, \tag{2.1}$$

$$W_i = \pi' x_i + u_i, \quad E(\varepsilon_i | x_i) = 0 \tag{2.2}$$

where y_i is a scalar, W_i is $p \times 1$ vector that includes the vector of endogenous variables Y_i and the vector of exogenous regressors z_i, and x_i is a $\overline{K} \times 1$ vector of exogenous variables including z_i. Let g_i be $g_i = g(\varpi_i, \delta) = x_i(y_i - \delta' W_i)$ so that $E[g(\varpi_i, \delta_0)] = 0$ follows from $E(\varepsilon_i | x_i) = 0$. Finally, let w_i^k for $k = 1, \ldots, p$ denote a typical element of W_i and $M \in [p, \overline{K}]$ denote the number of instruments actually used for estimating δ.

In this framework two questions are of main concern: (1) what is the optimal number M of instruments to use? and (2) should all the elements in x_i receive the same weight?

Donald and Newey (2001) address the first question by choosing the optimal number of instruments of a 2SLS estimator. Instrument selection is based on minimizing a scalar version of the second order MSE that results after using the first M instruments in x_i. The behavior of this estimator in finite samples will depend on how well the second order MSE approximates the finite sample MSE. This, of course, will depend on the characteristics of the case analyzed. In particular, it has been recognized in the literature that situations where the instruments are weak lead to a break down of the reliability of the traditional higher order expansions (Hahn et al., 2004; Rothenberg, 1984).

Contrary to Donald and Newey (2001), Okui (submitted for publication) addresses the second question instead of the first one. The author assumes that the instrument vector x_i can be partitioned in two sets, $x_i' = (r_{1i}', r_{2i}')$, where r_{1i} are the "primary" instruments (including z_i) and r_{2i} are the "secondary" instruments, and proposes a shrinkage 2SLS. Here there is no selection of instruments. Instead, the idea is to shrink the effect of r_{2i} on the estimation of δ. To be precise, let s denote a shrinkage parameter and let $P^s = P_{r1} + sP_{r2}$ be the modified projection matrix. The shrinkage 2SLS is just a 2SLS estimator that uses P^s instead of the usual projection matrix. This is, $\hat{\delta}_s = (W'P^sW)^{-1}W'P^sy$. To make the method operational Okui (submitted for publication) derives the second order MSE of $\hat{\delta}_s$ as a function of s and chooses $s \in [0, 1]$ as the minimizer of this function. The same comments as before apply to the accuracy of the higher order expansion to get s^*. Note also that the treatment on the two sets of instruments is not smooth since all variables in r_{1i} will receive a weight of 1 while all the variables in r_{2i} will be weighted by s. When s is small this suggests that a reallocation of instruments from r_{1i} to r_{2i} could affect the results to a great extent.

Before proceeding any further, I mention two important remarks. First, each of these two estimators gives an answer to only one of the questions previously stated. Second, both estimators assume homoskedasticity and their behavior under the presence

1 Kuersteiner (2002) uses kernels that have curvature at zero and exploits this assumption to derive an asymptotic MSE that depends on k_q (the qth generalized derivative of the kernel at the origin). For kernels with no curvature (i.e. $k_q = 0$) this approximation does not give any guidance choosing the optimal M.

2 This is important since most applications with microdata do not satisfy a homoskedasticity assumption and thus two stages least squares (2SLS) and limited information maximum likelihood (LIML) are no longer the best IV estimators. Despite this concern, nearly all of the literature on instrument selection has focused solely on 2SLS and LIML, with the exception of Donald et al. (2008).

of heteroskedasticity is unknown. In particular, 2SLS is no longer the best (minimum asymptotic variance) IV estimator which leads to the possibility of finding a better estimator in this case. The kernel weighted GMM estimator I propose in this paper tackles both questions simultaneously, gives the optimal partition of the instruments, and is robust to heteroskedasticity of unknown form.

3. Kernel weighted GMM with a trapezoidal kernel

The estimator I propose is a modification of the kernel weighted GMM estimator in Kuersteiner (2002). The main differences are the kernel used and the data-dependent bandwidth rule. Essentially, the MSE derived by Kuersteiner (2002) exploits an assumption regarding the kernel (i.e., a nonzero higher order derivative at zero) that plays a key role in the form of the MSE formula. The trapezoidal kernel used here does not satisfy this assumption, so that a new MSE formula is necessary. Furthermore, Kuersteiner's bandwidth selection rule exploits the fact that all the constants in the MSE can be consistently estimated (in the same way as in Andrews, 1991), something that will not hold for the estimator introduced here.

Consider the model of the previous section given by Eqs. (2.1) and (2.2). The following notation and assumptions will be used through the paper.

Notation 1. Let $Q_M = E(x_{Mi}W_i')$ be $M \times p$, where $x_{Mi} \equiv (x_i^1, x_i^2, \ldots, x_i^M)'$ contains the first M instruments and W_i is $p \times 1$. The jth row of Q_M is denoted by Γ_j', a $1 \times p$ vector containing the covariance of instrument j with w_i^k, $k = 1, \ldots, p$. Also let Ω_M be $E(x_{Mi}x_{Mi}'\varepsilon_i^2)$ with typical element ω_{jl}. The inverse of this matrix is Ω_M^{-1} and has typical element ϑ_{lj}^M. Finally, define $D_M = Q_M'\Omega_M^{-1}Q_M$ and $d_M = Q_M'\Omega_M^{-1}V_M$ where $V_M = \frac{1}{\sqrt{n}}\sum x_{Mi}\varepsilon_i$.

Assumption 3.1. Let $x_i^j \in \{x_i^1, x_i^2, \ldots, x_i^M\}$ be an arbitrary instrument. (a) $E(x_i^j) = 0$ and $E((x_i^j)^4) < \infty$ for all j. (b) There is no $\tau \neq \{0, 0, \ldots\}$ such that $\sum_{j=1}^{\infty}\tau_j x_i^j = 0$ a.s. (c) $\sum_{j,s,l,m=1}^{\infty}|cum(x_i^j, x_i^s, x_i^l, x_i^m)| < \infty$ where $cum(\cdot)$ denotes the fourth order cumulant of x_i.

Assumption 3.2. (a) $\sum_{j=1}^{\infty}\sum_{l\neq j}^{\infty}|E(x_i^l x_i^j \varepsilon_i^2)| < \infty$. (b) $E((x_i^j)^2\varepsilon_i^2) > 0$ for all j. (c) There is a constant $\sigma^2 \in (0, \infty)$ such that $\sum_{j=1}^{\infty}|E((x_i^j)^2\varepsilon_i^2) - \sigma^2| < \infty$. (d) $E\left(\varepsilon_i^4\right) < \infty$.

Assumption 3.3. $E|x_i^j\varepsilon_i|^3$, $E|x_i^j x_i^l \varepsilon_i|^3$ and $E|x_i^j w_i^k|^3$ for $k = 1, \ldots, p$ are bounded uniformly in j and l.

Assumption 3.1(a) imposes a zero mean that simplifies the notation and calculations but it is not essential.[3] Assumption 3.1(b) says that the infinite sequence of instruments is not linearly dependent with probability one, while 3.1(c) limits the dependence of the sequence of instruments.[4] Assumption 3.2 limits the degree of heteroskedasticity and implies that there exists a matrix $S_M = \Omega_M - \sigma^2 I_M$ with typical element s_{jl} satisfying $\lim_{M\to\infty}\sum_{j=1}^{M}\sum_{l=1}^{M}\left|s_{jl}\right| < \infty$. Finally, Assumption 3.3 implies a uniform in j central limit theorem for such random variables.

These assumptions imply that $\Omega_M = O(M)$ and guarantee that the operator Ω defined below has a well-defined, bounded inverse on the space of l^2 sequences (i.e., $\Gamma \in l^2$ if $\sum_j^{\infty}\left\|\Gamma_j\right\| < \infty$ where $\|\cdot\|$ denotes the Euclidean norm). Following Kuersteiner (2002), I use the sequence space l^2 to characterize the limit of D_M as $M \to \infty$. Here the operator Ω is defined component-wise by its image for all $\Gamma = \left\{\Gamma_j\right\}_{j=1}^{\infty} \in l^2$ as $b_i = \lim_{M\to\infty}\sum_j^M \Gamma_j\omega_{ji}$ where ω_{ji} is the j, i element of Ω_M. Lemma C.1 in Appendix C shows that under Assumptions 3.1 and 3.2, Ω_M^{-1} exists for all M, $\Omega \in L(l^2, l^2)$, and Ω^{-1} exists and is bounded.[5] Thus, since each column of $Q = \lim_M Q_M$ is an element of l^2, it follows that the limiting operator $Q'\Omega^{-1}$ maps a matrix with l^2 columns into another matrix with l^2 columns. Finally, the assumptions are used in Lemma C.2 to show that $\|\Omega_M^{-1} - \Omega^{-1}\|_* \to 0$, where $\|\cdot\|_*$ denotes the sup-norm.[6]

Now I can define $D = \lim_M D_M$ and $d_0 = \lim_M d_M$ to note that $D^{-1}d_0 \rightsquigarrow N(0, D^{-1})$ as $n \to \infty$ under standard assumptions. The infeasible estimator based on a nonrandom matrix $D_M^{-1}Q_M'\Omega_M^{-1}$ is given by:

$$\delta_{n,M} = D_M^{-1}Q_M'\Omega_M^{-1}\frac{1}{n}X_M'y_i = \delta_0 + \frac{1}{\sqrt{n}}D_M^{-1}d_M. \tag{3.1}$$

This infeasible estimator is consistent and asymptotically normal since $\sqrt{n}(\delta_{n,M} - \delta_0) - D^{-1}d_0 = D_M^{-1}d_M - D^{-1}d_0 = o_p(1)$ as $M, n \to \infty$. A feasible version of $\delta_{n,M}$ is obtained by replacing the nonrandom matrices with random counterparts. However, the previous statement about asymptotic normality will hold only when M goes to infinity at an appropriate rate. The following assumption takes this into consideration.

Assumption 3.4. The number of instruments M is assumed to be an increasing sequence such that $M \to \infty$ as $n \to \infty$ and $M = o(n^{1/2})$.

Notation 2. Let $k(\cdot)$ denote a kernel weight function and $\mathcal{K}_M = diag\,(k\,(1/M)\,, \ldots, k\,(1))$ a matrix having kernel weight $k(j/M)$ in the jth diagonal element and zeros otherwise. Thus, the number of instruments M plays the role of an integer bandwidth. Denote by $g_i^M(\delta) = (y_i - \delta' W_i)\left(x_i^1, x_i^2, \ldots, x_i^M\right)' = (y_i - \delta' W_i)x_{Mi}$ the moment function with M instruments and note that $E[g_i^M(\delta_0)] = 0$ for all M. Finally, let $\hat{\Omega}_M = \frac{1}{n}\sum_{i=1}^n x_{Mi}x_{Mi}'\hat{\varepsilon}_i^2$, $\hat{Q}_M = \frac{1}{n}\sum_{i=1}^n x_{Mi}W_i'$, $\hat{D}_M = \hat{Q}_M'\mathcal{K}_M\hat{\Omega}_M^{-1}\mathcal{K}_M\hat{Q}_M$, $\hat{d}_M = \hat{Q}_M'\mathcal{K}_M\hat{\Omega}_M^{-1}\mathcal{K}_M V_M$ and $\hat{\varepsilon}_i = y_i - \tilde{\delta}'W_i$ for a preliminary estimator $\tilde{\delta}$ satisfying $\sqrt{n}(\tilde{\delta} - \delta) = O_p(1)$.

Propositions A.1 and A.2 in Appendix A show that under the stated assumptions $\hat{D}_M$ and $\hat{d}_M$ are consistent estimators of $D = \lim_M D_M$ and $d_0 = \lim_M d_M$, in the sense that $\|\hat{D}_M - D\| = o_p(1)$ and $\|\hat{d}_M - d_0\| = o_p(1)$.

Now it is time to introduce the kernel weighted GMM estimator of δ_0. This is,

$$\hat{\delta}_{n,M} = \arg\min_{\delta}\left\{n\left(\frac{1}{n}\sum_{i=1}^n g_i^M(\delta)\right)'\left(\mathcal{K}_M\hat{\Omega}_M^{-1}\mathcal{K}_M\right) \times \left(\frac{1}{n}\sum_{i=1}^n g_i^M(\delta)\right)\right\},$$

and it can be written in closed form as

$$\hat{\delta}_{n,M} = \left(\hat{Q}_M'\mathcal{K}_M\hat{\Omega}_M^{-1}\mathcal{K}_M\hat{Q}_M\right)^{-1}\hat{Q}_M'\mathcal{K}_M\hat{\Omega}_M^{-1}\mathcal{K}_M\frac{X_M'y}{n} = \delta_0 + \frac{1}{\sqrt{n}}\hat{D}_M^{-1}\hat{d}_M. \tag{3.2}$$

Note that if $\mathcal{K}_M = I_M$ (i.e., truncated kernel) this estimator reduces to a standard GMM estimator that uses M instruments. Therefore, the role of the kernel is to affect the weight matrix $\hat{\Omega}_M^{-1}$

[3] The alternative approach is to drop this assumption and work with centered variables, i.e., $\tilde{x}_i = x_i - \bar{x}$.

[4] What is needed here are restrictions on the fourth-moment dependence between subset of instruments to guarantee that the higher order variance term is of a sufficiently simple form. The summability assumption for the cumulants limits this dependence due to the relationship between moments and cumulants, see Barndorff-Nielsen and Cox (1989, Page 144).

[5] $L(l^2, l^2)$ denotes the set of linear operators that map l^2 sequences into l^2 sequences.

[6] Most of the properties used in Lemmas C.1 and C.2 are in Gohberg and Goldberg (1981), while a similar proof in a time series context is in Kuersteiner (2001).

in some beneficial way. In this paper I use a trapezoidal kernel which is the simplest representative of the class of flat-top kernels. This is,

$$k_c(x) = 1\{|x| \leq c\} + 1\{c < |x| \leq 1\} \times \frac{1-|x|}{1-c}, \tag{3.3}$$

where $c \in (0, 1)$ and $1\{\cdot\}$ is the indicator function. Notice that $k_c(x)$ approaches the Bartlett kernel as $c \to 0$ while as $c \to 1$ it approaches the truncated kernel. Thus, the trapezoidal kernel is general enough to include in the limit a pure selection case as in Donald and Newey (2001) and the GMM estimator by Kuersteiner (2002) with a Bartlett kernel.

Politis and Romano (1995) documented the favorable properties of this kernel in spectral estimation and a closer look at its shape illustrates why it is particularly appealing to select instruments. The kernel assigns full weight to the first fraction c of instruments. Therefore, while the constant c controls how many instruments receive a weight equal to one, the bandwidth M controls how many instruments get a weight different than zero.[7] This has some similarities to the shrinkage 2SLS although with an improvement. There is now a smooth decay between the primary instruments and the last of the secondary instruments.[8]

As noted by Kuersteiner (2002), the use of the kernel weight matrix $\mathcal{K}_M$ introduces an inefficiency by employing $\mathcal{K}_M\hat{\Omega}_M^{-1}\mathcal{K}_M$ instead of the optimal $\hat{\Omega}_M^{-1}$ as weight matrix. He shows that this inefficiency is not related to the first order asymptotic properties of the estimator in the sense that $\hat{\delta}_{n,M}$ is first order asymptotically equivalent to $D^{-1}d_0$. In this paper I extend this result to the case of the trapezoidal kernel since a key assumption used by Kuersteiner (2002, see Assumption F) is that the kernel function has some curvature at zero. This assumption requires that for some $q \in (0,\infty)$ there exists a constant $k_q \in (0,\infty)$ such that $k_q = \lim_{x\to 0}(1-k(x))/|x|^q$. k_q is known as the qth generalized derivative of kernel $k(\cdot)$ at the origin and it is equal to zero for all q in the case of flat-top kernels.

Before proceeding with the MSE derivation of my estimator, I state more specifically why Kuersteiner's results are not directly applicable. The asymptotic MSE of Proposition 3.3 in Kuersteiner (2002) is given by

$$AMSE(M) = \frac{M^2}{n}\mathcal{A}\left(\int_{-\infty}^{\infty} k^2(x)\mathrm{d}x\right)^2 + \frac{1}{M^{2q}}k_q^2\mathcal{B}^{(q)}, \tag{3.4}$$

where $\mathcal{A}$ and $\mathcal{B}^{(q)}$ are parameters that can be consistently estimated. Throughout the paper the AMSE is defined as the sum of the (higher order) asymptotic bias square (first term) and the (higher order) asymptotic variance (second term). The AMSE in (3.4) displays a trade off between higher efficiency due to more instruments (through the second term) and bias introduced by estimating more parameters (first term)*as long as* the generalized derivative of kernel $k(\cdot)$ is nonzero. For kernels with $k_q^2 = 0$ there is no such a trade off and the MSE behaves asymptotically like M^2/n. This is actually the case for all flat-top kernels and thus asymptotic approximations such as those in (3.4) give no guidance on how to pick the optimal bandwidth. Intuitively, any kernel with $k_q^2 = 0$ will have a variance term of order $o(M^{-2q})$ instead of $O(M^{-2q})$ and this, in principle, is an appealing feature.

With a trapezoidal kernel the leading variance term in (3.4) is no longer present. In fact, many terms in the asymptotic expansion will have an order of magnitude determined by the rate at which the correlation between the instruments and the regressors decay to zero, and not by the curvature of the kernel. Thus, to derive a new asymptotic MSE that displays a bias-variance trade off it is sufficient to impose some conditions on the sequence $\{\Gamma_j\}_{j=1}^{\infty}$, the sequence that captures the correlation between instruments and regressors, and on the elements ω_{jl} of Ω. By being explicit about the behavior of $\{\Gamma_j\}_{j=1}^{\infty}$ and $\{\omega_{jl}\}_{j\neq l}$ it is possible to establish the rate of those terms even when the kernels are flat at the origin. Therefore, I introduce the following assumption.

Assumption 3.5. There exists $\alpha > 1$ and $\kappa \in (0,\infty)$ such that: (a) $|j^{\alpha+1}\|\Gamma_j\| - \kappa| = o(1)$, and (b) $\sum_{j=1}^{\infty} j^{\alpha}\sum_{l\neq j}^{\infty}|\omega_{jl}| < \infty$, where ω_{jl} is the j, l element of Ω.

Part (a) of this assumption says that the instruments decay in importance at a polynomial rate equal to $\alpha + 1$. This is, it implies that the researcher knows the order of the instruments. Similar conditions have been used by Donald and Newey (2001) and Kuersteiner (2002).[9] I will exploit the following three basic consequences of this assumption: (a) there exists $1 < s < \alpha$ such that $\sum_{j=1}^{\infty} j^s\|\Gamma_j\| < \infty$, (b) α is greater than q in (3.4),[10](c) the sequence $j^{\alpha}\|\Gamma_j\|$ behaves in the limit as an harmonic series. Part (b) is a stronger version of Assumption 3.2(a). It basically implies that the correlation between instruments also decays fast enough as the distance between instruments (given by the chosen ordering) increases. This will guarantee some useful properties of the convolution $\sum_{l\neq j}^{\infty}\|\Gamma_j\||\omega_{jl}|$.

To get an approximate MSE, I follow Kuersteiner (2002) and expand $\hat{\delta}_{n,M}$ around $\delta_{n,M}$ by using a second order Taylor approximation of $\hat{D}_M^{-1}$ around D^{-1}. This leads to

$$\sqrt{n}(\hat{\delta}_{n,M} - \delta_0) = D^{-1}\Big[I - (\hat{D}_M - D)D^{-1} + (\hat{D}_M - D)D^{-1}(\hat{D}_M - D)D^{-1}\Big]\hat{d}_M + o_p(M/n^{1/2}) \tag{3.5}$$

which is a valid expansion when $M = o(n^{1/2})$. Next, the scalar version of the approximate MSE only takes into account terms that are largest in probability in the expression,

$$MSE(M, c) = n\ell' D^{1/2}E[(\hat{\delta}_{n,M} - \delta_0)(\hat{\delta}_{n,M} - \delta_0)']D^{1/2}\ell - 1,$$

where $\ell \in \mathbb{R}^p$ is a vector of weights and $D^{1/2}$ is used to standardize the asymptotic variance. The leading terms are of order $O_p(M^2/n)$ and $O_p(M^{-(2\alpha+1)})$ and the remaining terms are of all of order $o_p(M^2/n)$. Theorems 3.1 and 3.2 summarize two of the main properties of $\hat{\delta}_{n,M}$. All proofs are in Appendix A.

Theorem 3.1. *Suppose Assumptions 3.1 to 3.5 hold and $k_c(x)$ is defined as in (3.3). Then,*

$$\sqrt{n}(\hat{\delta}_{n,M} - \delta_0) - D^{-1}d_0 = o_p(1).$$

Theorem 3.2. *Suppose Assumptions 3.1 to 3.5 hold and $k_c(x)$ is defined as in (3.3). Then, for any $\ell \in \mathbb{R}^p$ with $\ell'\ell = 1$,*

$$AMSE(M, c) = \frac{M^2}{n}\frac{4}{9}(1+2c)^2\mathcal{C}_A + \frac{1}{M^{2\alpha+1}}\left[F_{\alpha,\sigma,\kappa}(c)\mathcal{C}_B + \mathcal{C}_C\right] \tag{3.6}$$

where $\mathcal{C}_A$, $\mathcal{C}_B$ and $\mathcal{C}_C$ are nonnegative constants that depend on α, κ, ℓ and σ^2 and

[7] The constant c is less important in non-parametric density estimation. In fact, Politis and Romano (1995) recommend to take c in the neighborhood of $1/2$. This might not a good recommendation here since different values of c imply different combinations of primary and secondary instruments.

[8] The approach in this paper however requires the practitioner to specify an order of the instruments while the shrinkage 2SLS only requires to correctly divide the instruments in two groups.

[9] Although in time series models the order of the instruments is directly implied by the model.

[10] To derive (3.4) it is required that $q < s$ so it follows that $\alpha > q$.

$$F_{\alpha,\sigma,\kappa}(c) = \frac{\kappa^2 c^2}{\sigma^2(1-c)^2}\left[\frac{1}{c^{2\alpha+1}(4\alpha^2-1)\alpha} + \frac{c(2\alpha-1)-\alpha}{c^2\alpha(2\alpha-1)} - \frac{1}{2\alpha+1}\right] \quad (3.7)$$

satisfies $\lim_{c\to 1} F_{\alpha,\sigma,\kappa}(c) = 0$ *and* $\lim_{c\to 0} F_{\alpha,\sigma,\kappa}(c) = \infty$.

Theorem 3.1 shows that using the trapezoidal kernel does not affect the first order optimality of the GMM estimator, so that the distortions introduced on the weight matrix are of second order. Theorem 3.2 gives an expression for the asymptotic MSE using the largest in probability terms depending on M and n. The asymptotic MSE displays a trade off in both the number of instruments M and the constant c. The bias term (first term) gets smaller as $M \to 0$ and $c \to 0$, so the bias introduced by the trapezoidal kernel is bigger than that of the Bartlett kernel but smaller than that of the truncated kernel (i.e., GMM). On the contrary, the variance term (second term) gets smaller as $M \to \infty$ and $c \to 1$. Comparing (3.6) with (3.4) we see that $\alpha > q$ implies that the variance term in (3.6) is of smaller order.

To explore the properties of the optimal values for M and c is easier to first concentrate M out of the AMSE. For any given c, I can obtain the optimal number of instruments that minimizes (3.6) in closed form,

$$M^*(c) = \left[\frac{(\alpha+1/2)[F_{\alpha,\sigma,\kappa}(c)\mathcal{C}_B + \mathcal{C}_C]}{\frac{4}{9}(1+2c)^2\mathcal{C}_A}\right]^{\frac{1}{2\alpha+3}} n^{\frac{1}{2\alpha+3}}. \quad (3.8)$$

Plugging (3.8) into (3.6) results in the following profile AMSE, although a closed form solution for the optimal c is not available due to a non-linear first order condition.

$$AMSE(c) = n^{-\left(\frac{2\alpha+1}{2\alpha+3}\right)}\left\{2\left[\left(\frac{1}{2}+\alpha\right)(F_{\alpha,\sigma,\kappa}(c)\mathcal{C}_B + \mathcal{C}_C)\right]^{\frac{2}{2\alpha+3}} \times \left[\frac{4}{9}\mathcal{C}_A(1+2c)^2\right]^{\frac{2\alpha+1}{2\alpha+3}}\right\}. \quad (3.9)$$

Eq. (3.8) shows that the optimal rate for M depends on α which is unknown. Thus, the optimal rate of the bandwidth is $o(n^{1/5})$ but it could actually be much smaller than this. On the other hand, Eq. (3.9) shows that the behavior of $c^* = \operatorname{argmin} AMSE(c)$ is independent of n. The higher order approximation does not give explicit expressions for the constants $\mathcal{C}_A, \mathcal{C}_B, \mathcal{C}_C$ but it can be shown that $\mathcal{C}_A$ depends only on $\sigma_{u\varepsilon} = E(u_i\varepsilon_i)$ (with a positive sign), $\mathcal{C}_C$ depends positively on κ and negatively on σ and α, while $\mathcal{C}_B$ is just a constant not depending on any of the primitive parameters of the model. Therefore, Eqs. (3.8) and (3.9) imply that $M^*(c)$ increases with κ (i.e., with stronger instruments) and decreases with α and $\sigma_{u\varepsilon}$ (i.e., with instruments that decay importance faster or a higher degree of endogeneity). In addition, c^* decays with a higher α, and increases as κ increases. Both M^* and c^* depend on unknown constants so that using them directly in the estimation process does not result in a feasible procedure. The next section discusses how to make the kernel weighted GMM estimator fully feasible.

4. Bandwidth selection rule

As shown in the previous section, the trapezoidal kernel allows simultaneous selection and weighting of the number of instruments, and it results in an AMSE that has a variance term of lower order compared to kernels that have curvature at zero. The major flexibility however comes at a price. Despite the bandwidth M, now there is a new parameter to be estimated (the constant c). Additionally, the optimal values in Eqs. (3.8) and (3.9) depend on quantities such as α and κ that are hard to estimate. In fact, the crude nature of the approximation makes impossible to get sample counterparts of $\mathcal{C}_A$, $\mathcal{C}_B$ and $\mathcal{C}_C$, meaning that a plug-in rule to get estimates of M and c would be infeasible. Therefore, the method I propose is to approximate the MSE of $\hat{\delta}_{n,M}$ as a function of c and M by using the bootstrap. The procedure also has the advantage of being simple to compute and does not require the estimation of any complicated function of the data.

The bootstrap for this case works as follows. Assume that we have a consistent preliminary estimator $\tilde{\delta}$ of δ_0 and recall that $\varpi_i = (y_i, Y_i, x_i)$ is the vector containing all the variables of the model. Let $\tilde{g}_i$ be defined as $\tilde{g}_i = g(\varpi_i, \tilde{\delta}) = x_i(y_i - \tilde{\delta}'W_i)$. Since it is well known that i.i.d. sampling in GMM estimation does not reproduce the moment condition that holds for the data (i.e., $E[g(\varpi_i, \delta_0) = 0]$) when the model is overidentified, I follow Brown and Newey (2002) and use the empirical likelihood (EL) bootstrap. The EL probabilities are given by

$$\tilde{p}_i = \frac{1}{n(1-\tilde{\lambda}'\tilde{g}_i)}, \qquad \tilde{\lambda} = \operatorname{argmax}_{\lambda'\tilde{g}_i<1}\sum_{i=1}^{N}\ln(1-\lambda'\tilde{g}_i), \qquad i = 1,\ldots,n. \quad (4.1)$$

Using these probabilities I draw B samples of n i.i.d. observations $\varpi_1^{*b},\ldots,\varpi_n^{*b}$ with replacement from the distribution with $\Pr(\varpi^* = \varpi_i) = \tilde{p}_i$ and calculate the bootstrap version of the MSE. The optimal values $(\hat{M}, \hat{c})$ are defined as the minimizers of the bootstrap MSE.

To be more precise, I use the following algorithm.

Algorithm 4.1. Let $\tilde{\delta}$ be a preliminary estimator and $\tilde{p}_i$ the EL probabilities.

1. Generate a grid of values for M and c. Denote the grid by Θ^g.[11]
2. For each $b = 1,\ldots,B$ draw n i.i.d. observations $\varpi_1^{*b},\ldots,\varpi_n^{*b}$ with replacement from the distribution with $\Pr(\varpi^* = \varpi_i) = \tilde{p}_i$.
3. For each $b = 1,\ldots,B$ compute the difference $\{\hat{\delta}^{*b}_{M,c} - \tilde{\delta}\}$ for all $(M,c) \in \Theta^g$, where $\hat{\delta}^{*b}_{M,c}$ is the kernel weighted GMM estimator with constant c and bandwidth M using the bootstrap sample b.
4. Calculate the bootstrap MSE as
$$BsMSE(M,c) = \frac{1}{B}\sum_{b=1}^{B}\ell'\left\{\left(\hat{\delta}^{*b}_{M,c} - \tilde{\delta}\right)\left(\hat{\delta}^{*b}_{M,c} - \tilde{\delta}\right)'\right\}\ell, \quad (4.2)$$
where ℓ as in the previous section is some vector of linear combination coefficients.
5. Define $(\hat{M}, \hat{c})$ as the minimizers of $BsMSE(M,c)$.

Remark 4.1. To compute the EL probabilities x_i has to include "all" of the available instruments, regardless of the actual number of instruments used to compute $\tilde{\delta}$.[12]

Remark 4.2. Note that the EL bootstrap uses EL probabilities but there is no EL estimator of δ_0 involved in the procedure.

As was the case for the two estimators discussed in Section 2, a preliminary estimator $\tilde{\delta}$ of δ is required to make the method feasible. Asymptotically, it does not make any difference to pick among different consistent estimators but in finite samples the results might change.[13] The sensitivity of the estimator to this choice is studied in the next section.

---

[11] In principle the bandwidth could take any positive real number and c could take any value in $(0, 1)$. However, since the role of M here is to select instruments, only the integers in $\{1, 2, \ldots, \bar{K}\}$ are of interest. Similarly, for a given bandwidth M only the values of c in $\left[\frac{1}{M}, 1-\frac{1}{M}\right]$ are relevant. This is because any $c < \frac{1}{M}$ produces just a re-scaling of units while any $c > 1-\frac{1}{M}$ implies weights equal to 1.

[12] For example, in some of the simulations of the next section $\tilde{\delta}$ uses only one instrument while $\tilde{g}_i$ is 25×1.

[13] For example, $\tilde{\delta}$ could be either a GMM estimator that uses all instruments in x_i or an IV estimator that uses the minimum number of instruments that are required to get identification.

The idea to use the bootstrap to select moment conditions has been introduced by Inoue (2006). The selection criterion in Inoue (2006) is not based on MSE but rather on minimizing approximate coverage error probabilities of confidence regions. In the present paper the argument in favor of the bootstrap method over plug-in selection rules coming from estimates of the second order asymptotic expansion is purely based on the fact that the bootstrap is simple and feasible. Intuitively, the bootstrap estimates the full Edgeworth expansion up to a certain order and so one would expect that the Algorithm above provides a valid asymptotic approximation in the sense that $AMSE(\hat{M}, \hat{c})/AMSE(M^*, c^*) \to^p 1$. Verification of this conjecture is beyond the scope of this paper. For a discussion of these issues in a similar context see Inoue (2006).

5. Monte Carlo results

In this Section 1 carry out a simulation study[14] to compare the performance of the kernel weighted GMM estimator against the 2SLS estimators of Donald and Newey (2001) and Okui (submitted for publication). The four models to be simulated share the common structure:

$$y_i = \gamma Y_i + \varepsilon_i \tag{5.1}$$

$$Y_i = \pi' x_i + u_i \tag{5.2}$$

where $\gamma = 0.1$ in all cases and π changes according to the models. For model 1, 3 and 4 $(\varepsilon_i, u_i)' \sim N(0, \Sigma)$ with Σ having unit variances and covariance $\sigma_{u\varepsilon}$ and $x_i \sim N(0, I_{\overline{K}})$. The theoretical first stage R^2 is $R_f^2 = \pi'\pi/(1 + \pi'\pi)$ as shown in Hahn and Hausman (2002).

Model 1: (Smooth Decay, Donald and Newey, 2001). The kth element of π is:

$$\pi_k = \phi(\overline{K})\left(1 - \frac{k}{\overline{K}+1}\right)^4, \quad \text{for } k = 1, \ldots, \overline{K},$$

where $\phi(\overline{K})$ is chosen so that $\pi'\pi = R_f^2/(1 - R_f^2)$. Here the strength of instruments decreases moderately in k.

Model 2: (Strong and Weak – Homoskedastic/Heteroskedastic). The first 5 instruments are strong, the next instruments decay in importance up to $\overline{K}/2$ and the rest of the instruments are completely irrelevant. This is:

$$\pi_j = \phi, \quad \pi_k = \phi(\overline{K})\left(1 - \frac{k-5}{\overline{K}-5+1}\right)^4,$$

$$\text{for } j = 1, \ldots, 5 \quad \text{and} \quad k = 6, \ldots, \overline{K}/2,$$

where $\phi = 0.98 \times \sqrt{\frac{R_f^2}{5(1-R_f^2)}}$ and $\phi(\overline{K})$ is chosen so that $\pi'\pi = R_f^2/(1 - R_f^2)$. This model also differs in the way (ε_i, u_i) and x_i are distributed. The disturbances are generated by the equations $\tilde{\varepsilon}_i = \sigma_{u\varepsilon} v_{1i} + (1 - \sigma_{u\varepsilon}) v_{2i}$ and $u_i = \sigma_{u\varepsilon} v_{1i} + (1 - \sigma_{u\varepsilon}) v_{3i}$, where v_{1i}, v_{2i} and v_{3i} follow independent t-distributions with unit variance and 10 degrees of freedom. The instruments have a lognormal distribution re-scaled to have zero mean and unit variance. Finally, in the homoskedastic version $\varepsilon_i = \tilde{\varepsilon}_i$, while $\varepsilon_i = |x_{1i}|\tilde{\varepsilon}_i$, in the heteroskedastic version (i.e., heteroskedasticity is caused by the first instrument).

Model 3: (No Decay, Hahn and Hausman, 2002). The kth element of π is set to:

$$\pi_k = \sqrt{\frac{R_f^2}{\overline{K}(1 - R_f^2)}}, \quad \text{for } k = 1, \ldots, \overline{K}.$$

Here all instruments are equally important and Assumption 3.5 is violated.

Model 4: (First Strong, Okui, submitted for publication). The elements of π are:

$$\pi_1 = \phi, \quad \pi_k = \frac{\phi}{\sqrt{\overline{K}-1}}, \quad \text{for } k = 2, \ldots, \overline{K},$$

where ϕ is chosen to satisfy $\pi'\pi = R_f^2/(1 - R_f^2)$. This model should favor the shrinkage 2SLS estimator as it implies that the first instrument is strong but the others are poor, violating Assumption 3.5.

Each experiment depends on n, $\sigma_{u\varepsilon}$ and R_f^2. The degree of endogeneity is summarized by $\sigma_{u\varepsilon}$ while R_f^2 controls the strength of the instruments. The values considered are $n \in \{100, 500\}$, $\sigma_{u\varepsilon} \in \{0.1, 0.5, 0.9\}$ and $R_f^2 \in \{0.1, 0.01\}$.[15] The total number of instruments $\overline{K}$ is set to 20 when $n = 100$ and to 25 when $n = 500$. Finally, the number of Monte Carlo replications is equal to 1000.

I compute six estimators for γ. A GMM and EL estimators that use all of the available moments,[16] a 2SLS estimator with Donald and Newey's (2001) optimal selection of instruments (DN2SLS), the shrinkage 2SLS estimator (S2SLS) of Okui (submitted for publication),[17] and the kernel weighted GMM where M and c are chosen using the bootstrap (BsKGMM). I also compute the infeasible kernel weighted GMM estimator that uses the true optimal values of M and c (InKGMM).[18] Recall that DN2SLS, S2SLS and BsKGMM require a preliminary estimator. For DN2SLS and S2SLS, I follow Donald and Newey (2001) and use a 2SLS where the number of instruments is chosen by the first stage cross validation criterion. For BsKGMM, I employ a GMM estimator with all valid moments.

As in Donald and Newey (2001), a variety of robust measures of central tendency and dispersion are presented. For each estimator I compute the median bias (Med. Bias), the median absolute deviations (MAD), the difference between the 0.1 and 0.9 quantile (Dec. Rge) in the distribution of $\hat{\delta}$ as well as the coverage rate (Cov. Rte) of a 95% confidence interval.[19] In addition to these measures, the mean squared error (MSE) is also computed to visualize possible problems of non-existence of finite sample moments.

To compute the coverage probability I use the following estimate of the asymptotic variance, as in Donald and Newey (2001) and Okui (submitted for publication),

$$\hat{V} = n^{-1}\tilde{\epsilon}'\tilde{\epsilon}(W^{*\prime}W)^{-1}(W^{*\prime}W^*)(W'W^*)^{-1} \tag{5.3}$$

where $\tilde{\epsilon} = y - W\tilde{\delta}$, $\tilde{\delta} = (W^{*\prime}W)^{-1}W^{*\prime}y$ and $W^* = P^{s\prime}W$ for 2SLS and $W^* = P^{M\prime}W$ for DN2SLS where P^M is the projection matrix with M instruments. On the other hand, the asymptotic variance of EL and KGMM for the model $E(g_i(\delta_0)) = 0$ is $V = (Q'\Omega^{-1}Q)^{-1}$ where $Q = E(\partial g_i(\delta_0)/\partial\delta')$, and $\Omega = E(g_i(\delta_0)g_i(\delta_0)')$. I compute $\hat{V}$ by replacing the population moments with sample analogs evaluated at the point estimates using the EL probabilities in the case of EL and the empirical measure in the case of KGMM.

14 All the computations were carried out using Matlab 7.6 for linux in the Condor cluster at the University of Wisconsin Computer Science Department.

15 These values reflect situations in which the instruments are weak. The associated values of the concentration parameter $\mu^2 = n\pi'E(x_i x_i')\pi/E(u_i)$ for models 1, 3 and 4 are {11.11, 1.01} and {55.55, 5.05}, for $n = 100$ and $n = 500$, respectively. Similar values correspond to Model 2.

16 For EL I first concentrate out the Lagrange multiplier and use a modified Newton–Raphson algorithm that uses both first and second derivatives. Then, I use a Davidon–Fletcher–Powell algorithm to maximize over the parameter of interest, using analytic first derivatives. Given a direction, a line search algorithm based on repeated quadratic approximations is used.

17 The first element in x_i is considered the primary instrument.

18 The optimal values of M and c (as well as the optimal MSE and Bias) were numerically computed for each model using 10^5 replications.

19 The coverage rate is computed as the mean of the indicator that the confidence interval contains the true parameter value.

Table 1
Model 1 with $n = 100$ and $n = 500$.

Sigma	Estimator	Med. Bias	Med. AD	Mean SE	Dec. Rge.	Cov. Rate	Med. Bias	Med. AD	Mean SE	Dec. Rge.	Cov. Rate
		$R_f^2 = 0.1, n = 100$					$R_f^2 = 0.01, n = 100$				
0.1	GMM	0.067	0.145	0.052	0.537	0.830	0.091	0.185	0.081	0.671	0.822
	EL	0.060	0.334	38.35	1.585	0.440	0.138	0.595	172.2	2.931	0.399
	DN2SLS	0.045	0.159	0.792	0.606	0.975	0.098	0.312	464.9	1.974	0.991
	S2SLS	0.062	0.124	0.040	0.449	0.950	0.099	0.158	0.071	0.572	0.953
	BsKGMM	0.056	0.152	0.061	0.586	0.840	0.087	0.191	0.110	0.735	0.827
	InKGMM	0.067	0.144	0.052	0.538	0.833	0.091	0.185	0.081	0.671	0.822
0.5	GMM	0.317	0.317	0.142	0.478	0.384	0.478	0.478	0.286	0.574	0.236
	EL	0.163	0.357	29.85	1.449	0.444	0.451	0.651	144.8	2.442	0.375
	DN2SLS	0.215	0.273	11.55	0.682	0.780	0.464	0.538	277.1	1.853	0.757
	S2SLS	0.321	0.324	0.127	0.453	0.519	0.476	0.477	0.278	0.500	0.323
	BsKGMM	0.289	0.298	0.123	0.570	0.503	0.479	0.483	0.298	0.636	0.286
	InKGMM	0.173	0.212	0.085	0.591	0.722	0.478	0.478	0.285	0.568	0.234
0.9	GMM	0.573	0.573	0.352	0.342	0.009	0.860	0.860	0.758	0.318	0.000
	EL	0.085	0.309	9.82	1.443	0.664	0.798	0.836	28.82	1.817	0.239
	DN2SLS	0.274	0.313	13.77	0.686	0.649	0.774	0.802	48.16	1.317	0.371
	S2SLS	0.490	0.490	0.257	0.458	0.233	0.864	0.864	0.743	0.295	0.008
	BsKGMM	0.361	0.364	0.185	0.637	0.360	0.852	0.852	0.734	0.374	0.011
	InKGMM	0.242	0.254	0.102	0.534	0.573	0.730	0.730	0.649	0.715	0.088
		$R_f^2 = 0.1, n = 500$					$R_f^2 = 0.01, n = 500$				
0.1	GMM	0.038	0.080	0.015	0.297	0.920	0.084	0.140	0.045	0.483	0.916
	EL	0.011	0.118	0.039	0.478	0.775	0.044	0.547	228.8	3.052	0.479
	DN2SLS	0.027	0.088	0.016	0.316	0.958	0.068	0.185	2029	0.906	0.988
	S2SLS	0.035	0.080	0.014	0.291	0.951	0.086	0.138	0.044	0.463	0.954
	BsKGMM	0.034	0.085	0.016	0.307	0.915	0.086	0.137	0.049	0.486	0.916
	InKGMM	0.038	0.080	0.015	0.298	0.919	0.085	0.139	0.045	0.483	0.916
0.5	GMM	0.156	0.157	0.036	0.285	0.628	0.410	0.410	0.204	0.437	0.261
	EL	0.009	0.114	0.035	0.448	0.824	0.152	0.545	210.5	3.044	0.580
	DN2SLS	0.069	0.104	0.023	0.344	0.876	0.360	0.399	1.781	0.853	0.674
	S2SLS	0.136	0.143	0.033	0.315	0.727	0.410	0.410	0.197	0.417	0.304
	BsKGMM	0.092	0.114	0.028	0.366	0.748	0.411	0.411	0.202	0.482	0.271
	InKGMM	0.058	0.095	0.019	0.317	0.871	0.307	0.314	0.159	0.624	0.588
0.9	GMM	0.278	0.278	0.081	0.235	0.177	0.748	0.748	0.574	0.262	0.000
	EL	−0.002	0.094	0.027	0.392	0.914	0.135	0.439	105.2	2.580	0.704
	DN2SLS	0.093	0.121	0.026	0.339	0.834	0.487	0.525	5.116	0.897	0.451
	S2SLS	0.152	0.163	0.038	0.338	0.750	0.714	0.714	0.502	0.399	0.054
	BsKGMM	0.095	0.116	0.023	0.314	0.786	0.714	0.714	0.490	0.486	0.047
	InKGMM	0.079	0.110	0.021	0.313	0.825	0.405	0.412	0.236	0.626	0.417

DN2SLS: 2SLS with Donald and Newey's optimal instruments, S2SLS: shrinkage 2SLS; BsKGMM: Bootstrap Kernel Weighted GMM; InKGMM: Infeasible Kernel Weighted GMM. 200 Bootstrap replications.

Tables 1 to 7 summarize the results.[20] In terms of bias, the performance of EL is really hard to match. None of the estimators that choose instruments optimally improves upon the performance of EL, except for the BsKGMM estimator in Model 2, Table 2 . This is not surprising considering the results in Donald et al. (2008), where the order of the bias term in the asymptotic MSE of EL is $O(M/n)$ instead of $O(M^2/n)$. This is also related to the results in Donald and Newey (2001) where the bias for 2SLS increases rapidly with the number of instruments while it does not for LIML. However, the bias of BsKGMM is always smaller than that of GMM, meaning that the new estimator introduces an improvement by choosing M and c optimally. This finding is consistent with the bias reduction suggested by the asymptotic MSE derived in Section 3.

When looking at the MAD it seems that none of the estimators dominates the rest and that all of them have the smallest MAD for some specification. BsKGMM does a particular good job in model 2 (Table 2), suggesting that the presence of heteroskedasticity can significantly affect estimators that assume homoskedasticity, as DN2SLS and S2SLS. This is related to the findings in Hausman et al. (2009), where LIML has higher bias and dispersion than a heteroskedastic version of Fuller's estimator. In addition, BsKGMM has a smaller MAD than GMM except when $\sigma_{u\varepsilon} = 0.1$.

The MSE shows a different picture. Here, S2SLS and BsKGMM have the lowest values in most cases. In particular, model 3 is dominated by S2SLS (as expected) and model 2 is dominated by BsKGMM. BsKGMM improves in all cases over GMM when the endogeneity is high ($\sigma_{u\varepsilon} = 0.9$), and in half of the cases when $\sigma_{u\varepsilon} = 0.5$.[21]

Regarding the coverage rate, S2SLS and DN2SLS are always the estimators with coverage rate closest to the nominal size, followed by EL. DN2SLS performs reasonably well in model 2 as well, where the errors are heteroskedastic. BsKGMM performs better when $n = 500$ and $R_f^2 = 0.1$ (with an associated concentration parameter of $\mu^2 = 55$), but the empirical sizes are still below the nominal size. In some cases BsKGMM shows a coverage rate that is much better than that of GMM (e.g., 78% versus 18% in Table 1).[22]

It is of interest to compare BsKGMM with InKGMM in an attempt to see how much is lost by using a data-dependent rule

[20] One should be careful interpreting the numbers in view of the fact that 1000 Monte Carlo simulations are probably not enough to be accurate to more than two decimals points.

[21] Across all models, DN2SLS and EL have values of MSE that are extremely high. This is caused by the presence of fat tails, casting doubts about the existence of finite sample moments for these two estimators.

[22] The coverage rate is highly sensitive to low values of R_f^2 (i.e., low values of the concentration parameter) reflecting the poor approximation of standard asymptotics under the presence of weak instruments.

Table 2
Model 2 (Heteroskedastic) with $n = 100$ and $n = 500$.

Sigma	Estimator	Med. Bias	Med. AD	Mean SE	Dec. Rge.	Cov. Rate	Med. Bias	Med. AD	Mean SE	Dec. Rge.	Cov. Rate
		$R_f^2 = 0.1, n = 100$					$R_f^2 = 0.01, n = 100$				
0.1	GMM	0.022	0.065	0.015	0.277	0.863	0.035	0.078	0.018	0.305	0.863
	EL	0.024	0.217	6.434	1.225	0.302	0.059	0.355	79.56	2.367	0.315
	DN2SLS	0.033	0.245	25.02	1.592	0.814	0.057	0.582	990.5	7.524	0.949
	S2SLS	0.032	0.145	0.514	0.831	0.772	0.055	0.161	0.748	0.823	0.836
	BsKGMM	0.021	0.072	0.028	0.317	0.880	0.037	0.084	0.028	0.347	0.880
	InKGMM	0.022	0.065	0.015	0.277	0.863	0.035	0.078	0.018	0.305	0.863
0.5	GMM	0.137	0.137	0.040	0.280	0.546	0.201	0.202	0.066	0.320	0.440
	EL	0.115	0.240	77.37	1.167	0.304	0.266	0.389	69.37	2.491	0.282
	DN2SLS	0.139	0.281	3920	1.601	0.713	0.354	0.740	9250	6.775	0.819
	S2SLS	0.190	0.238	0.600	0.861	0.564	0.274	0.296	1.193	0.972	0.544
	BsKGMM	0.123	0.134	0.053	0.310	0.664	0.194	0.196	0.112	0.397	0.535
	InKGMM	0.136	0.136	0.040	0.279	0.545	0.201	0.202	0.066	0.320	0.440
0.9	GMM	0.274	0.274	0.097	0.286	0.115	0.398	0.398	0.185	0.294	0.051
	EL	0.191	0.284	14.19	1.100	0.323	0.429	0.502	95.60	1.707	0.173
	DN2SLS	0.259	0.362	18.31	1.585	0.555	0.699	0.967	182.2	6.111	0.576
	S2SLS	0.353	0.378	0.881	0.852	0.265	0.525	0.526	1.781	1.118	0.097
	BsKGMM	0.248	0.249	0.100	0.346	0.292	0.391	0.391	0.229	0.356	0.107
	InKGMM	0.225	0.227	0.084	0.336	0.373	0.398	0.398	0.185	0.294	0.051
		$R_f^2 = 0.1, n = 500$					$R_f^2 = 0.01, n = 500$				
0.1	GMM	0.020	0.053	0.007	0.201	0.930	0.045	0.087	0.018	0.312	0.921
	EL	−0.002	0.119	0.095	0.523	0.496	0.064	0.690	445.2	4.930	0.380
	DN2SLS	0.008	0.163	0.163	0.777	0.637	0.037	0.364	409.2	3.629	0.853
	S2SLS	0.020	0.156	0.325	0.887	0.608	0.049	0.185	1.396	1.373	0.702
	BsKGMM	0.019	0.052	0.008	0.209	0.934	0.043	0.084	0.025	0.338	0.914
	InKGMM	0.020	0.053	0.007	0.201	0.930	0.045	0.087	0.018	0.312	0.921
0.5	GMM	0.084	0.088	0.015	0.197	0.749	0.204	0.205	0.060	0.288	0.512
	EL	−0.026	0.113	0.150	0.497	0.572	0.210	0.730	609.4	5.377	0.445
	DN2SLS	0.043	0.190	0.195	0.809	0.571	0.259	0.441	74.72	3.940	0.724
	S2SLS	0.096	0.199	0.380	0.965	0.496	0.263	0.310	1.704	1.575	0.507
	BsKGMM	0.075	0.088	0.015	0.236	0.773	0.204	0.205	0.063	0.307	0.540
	InKGMM	0.058	0.073	0.012	0.220	0.883	0.205	0.205	0.060	0.288	0.510
0.9	GMM	0.150	0.151	0.030	0.194	0.339	0.368	0.368	0.154	0.284	0.059
	EL	−0.052	0.099	0.233	0.420	0.764	0.252	0.694	169.3	4.647	0.484
	DN2SLS	0.077	0.191	0.230	0.822	0.576	0.472	0.562	838.9	3.552	0.496
	S2SLS	0.158	0.292	0.503	0.999	0.378	0.481	0.490	2.179	1.551	0.172
	BsKGMM	0.075	0.096	0.023	0.314	0.671	0.361	0.362	0.157	0.307	0.121
	InKGMM	0.047	0.073	0.014	0.250	0.876	0.333	0.333	0.152	0.401	0.147

DN2SLS: 2SLS with Donald and Newey's optimal instruments, S2SLS: shrinkage 2SLS; BsKGMM: Bootstrap Kernel Weighted GMM; InKGMM: Infeasible Kernel Weighted GMM. 200 Bootstrap replications.

rather than the infeasible true values of M and c. Across the different specifications both estimators show similar measures, suggesting that,

$$(AMSE(\hat{M}, \hat{c}) - AMSE(M^*, c^*))/AMSE(M^*, c^*) \quad (5.4)$$

is small, which is the usual justification given for the data-dependent selection rules (see Donald and Newey, 2001; Okui, submitted for publication). There are a few exceptions. In model 1 for $n = 500$, $R_f^2 = 0.01$ and $\sigma_{u\varepsilon} = 0.9$, BsKGMM performs worse than InKGMM. Note that exactly the same configuration of parameters cause problems for the data-dependent shrinkage rule in Okui (submitted for publication, Table 6). This relates to the fact that the Theorem presented in Okui (submitted for publication), which essentially shows that differences like those in Eq. (5.4) are small, does not imply that the estimator that uses the feasible data-dependent shrinkage parameter attains the minimum MSE. Therefore, cases where BsKGMM and InKGMM perform differently may capture estimation error and do not imply that the bootstrap is not working. In fact, in most of the cases the performance between the two is very similar. Table 5 provides further evidence. The estimated M and c are never below the true M and c, so there might be a positive bias in estimating these quantities. There are two phenomenon that are worth noting. First, in most cases $c = \frac{1}{M}$ is not an optimal value suggesting that the optimal kernel does not usually have a Bartlett shape. Second, both the feasible and infeasible values of M decay with $\sigma_{u\varepsilon}$ which is consistent with the behavior implied by Eq. (3.8). The table also reports the bootstrap versions of the bias and MSE (where, for example, BsMSE is computed as in Eq. (4.2) evaluated at the feasible optimal (M, c)). Both of these quantities are shifted to the left relative to the true bias and MSE with a median always below and, in some cases, with the 75th quantile below the true bias or MSE. Note however that these differences have no direct implications on (5.4) being large or small.

Finally, Tables 6 and 7 address the robustness of the results to changes in some of the key elements that characterize the BsKGMM estimator.[23] Table 6 illustrates how DN2SLS, S2SLS and BsKGMM perform when EL is used as preliminary estimator. Not surprisingly the bias of the three estimators goes down when EL is used, and the Dec. Rge. goes up. The MAD goes up or down depending on the specification. Note that the bias of BsKGMM seems to be more sensitive to the preliminary estimator, especially when $\sigma_{u\varepsilon}$ is high, and that its MSE explodes by using EL as preliminary estimator when $R_f^2 = 0.01$.[24]

23 Additional simulations are included in a supplementary appendix (see Canay, 2009).

24 In all the simulations performed (including those not presented here) BsKGMM never showed a high MSE when GMM with all instruments was used as preliminary

Table 3
Model 2 (Homoskedastic) with $n = 100$ and $n = 500$.

Sigma	Estimator	Med. Bias	Med. AD	Mean SE	Dec. Rge.	Cov. Rate	Med. Bias	Med. AD	Mean SE	Dec. Rge.	Cov. Rate
		$R_f^2 = 0.1, n = 100$					$R_f^2 = 0.01, n = 100$				
0.1	GMM	0.053	0.137	0.048	0.530	0.768	0.084	0.172	0.072	0.646	0.775
	EL	0.027	0.324	24.35	1.401	0.388	0.075	0.529	106.3	2.586	0.355
	DN2SLS	0.041	0.178	0.445	0.797	0.964	0.072	0.364	53.59	2.165	0.985
	S2SLS	0.057	0.130	0.040	0.465	0.954	0.093	0.158	0.060	0.560	0.950
	BsKGMM	0.051	0.147	0.056	0.556	0.787	0.095	0.187	0.095	0.704	0.796
	InKGMM	0.052	0.136	0.048	0.528	0.768	0.084	0.174	0.072	0.646	0.776
0.5	GMM	0.314	0.315	0.142	0.504	0.367	0.464	0.464	0.273	0.574	0.226
	EL	0.172	0.380	10.36	1.522	0.361	0.474	0.609	153.1	2.374	0.333
	DN2SLS	0.188	0.250	83.69	0.763	0.783	0.433	0.556	3947	1.993	0.788
	S2SLS	0.319	0.319	0.129	0.450	0.505	0.472	0.472	0.261	0.474	0.336
	BsKGMM	0.291	0.302	0.137	0.598	0.449	0.468	0.469	0.284	0.626	0.267
	InKGMM	0.186	0.220	0.100	0.647	0.667	0.466	0.466	0.274	0.573	0.228
0.9	GMM	0.596	0.596	0.381	0.388	0.013	0.857	0.857	0.748	0.321	0.000
	EL	0.173	0.349	21.41	1.395	0.505	0.798	0.836	36.33	1.520	0.193
	DN2SLS	0.322	0.363	12.78	0.773	0.598	0.796	0.827	3.183	1.241	0.393
	S2SLS	0.553	0.553	0.316	0.441	0.146	0.856	0.856	0.741	0.280	0.007
	BsKGMM	0.424	0.429	0.240	0.663	0.284	0.852	0.852	0.733	0.370	0.005
	InKGMM	0.275	0.294	0.127	0.574	0.542	0.743	0.743	0.642	0.685	0.142
		$R_f^2 = 0.1, n = 500$					$R_f^2 = 0.01, n = 500$				
0.1	GMM	0.034	0.083	0.015	0.301	0.895	0.094	0.147	0.049	0.496	0.885
	EL	−0.005	0.128	0.048	0.519	0.705	0.055	0.548	227.0	3.069	0.437
	DN2SLS	0.020	0.082	0.017	0.326	0.955	0.078	0.195	4.3	0.961	0.987
	S2SLS	0.026	0.081	0.015	0.289	0.942	0.093	0.136	0.049	0.465	0.943
	BsKGMM	0.031	0.084	0.016	0.304	0.894	0.092	0.152	0.055	0.519	0.878
	InKGMM	0.034	0.082	0.015	0.302	0.892	0.094	0.147	0.049	0.496	0.885
0.5	GMM	0.153	0.154	0.037	0.288	0.599	0.423	0.423	0.213	0.472	0.230
	EL	−0.001	0.116	0.051	0.482	0.747	0.216	0.556	157.3	2.741	0.507
	DN2SLS	0.047	0.098	0.021	0.348	0.886	0.344	0.383	2.8	0.879	0.701
	S2SLS	0.143	0.148	0.034	0.303	0.700	0.425	0.426	0.210	0.431	0.287
	BsKGMM	0.098	0.118	0.031	0.365	0.717	0.421	0.421	0.213	0.493	0.255
	InKGMM	0.046	0.092	0.019	0.330	0.868	0.303	0.309	0.169	0.628	0.588
0.9	GMM	0.279	0.279	0.087	0.259	0.152	0.764	0.764	0.591	0.290	0.001
	EL	−0.002	0.099	0.032	0.412	0.858	0.183	0.459	74.6	2.682	0.625
	DN2SLS	0.081	0.105	0.022	0.321	0.846	0.502	0.534	28.4	0.839	0.453
	S2SLS	0.197	0.199	0.052	0.330	0.617	0.741	0.741	0.538	0.375	0.040
	BsKGMM	0.083	0.107	0.022	0.317	0.782	0.742	0.742	0.521	0.500	0.062
	InKGMM	0.072	0.103	0.020	0.322	0.826	0.423	0.429	0.254	0.620	0.469

DN2SLS: 2SLS with Donald and Newey's optimal instruments, S2SLS: shrinkage 2SLS; BsKGMM: Bootstrap Kernel Weighted GMM; InKGMM: Infeasible Kernel Weighted GMM. 200 Bootstrap replications.

Table 7 has two parts. The top part quantifies the gains in choosing c optimally versus fixing its value to some standard magnitude. I consider three values: $c = 0$ (i.e., Bartlett kernel), $c = 1/2$, (i.e., the rule of thumb recommended by Politis and Romano (1995) for spectral density estimation), and $c = 1$ (i.e., truncated kernel). The top part also shows the performance of BsKGMM when an IV estimator that includes the first element in x_i is used as preliminary estimator. From the table nothing seems to suggest that choosing c optimally introduces significant noise. Typically the estimator with $c = 0$ has lower bias and higher Dec. Rge., which illustrates the bias-variance trade off in c. A similar conclusion applies when the preliminary estimator uses only one instrument. Such a preliminary estimator reduces the bias of the BsKGMM but increases the dispersion. The bottom part of Table 7 shows a situation in which the instruments are not correctly ordered (the first half of the instruments was swapped with the second half). Even in this case BsKGMM performs equal or better than the standard GMM. Its bias and MAD are better than those of DN2SLS and S2SLS. It is interesting to observe that DN2SLS seems to be more sensitive than BsKGMM to the wrong ordering. For example, using the unordered instruments increases the bias of DN2SLS from 0.093 to 0.775 in some cases, while the same numbers for BsKGMM are 0.095 and 0.242. Finally, using the correct ordering seems to be less relevant when the instruments are weak, although DN2SLS is still more sensitive than BsKGMM.

estimator. A similar phenomenon arises in Table 7 when using IV as preliminary estimator.

6. Discussion

The paper derives the properties of a kernel weighted GMM estimator that uses a trapezoidal kernel as weight function. It establishes first order optimality of the estimator and derives an expression for the second order MSE. The asymptotic MSE displays a trade off in the bandwidth parameter M and in the constant c that determines the flatness of the kernel. The paper also shows that the trapezoidal kernel generates an asymptotic variance that is of lower order than that of regular kernels.

One aspect that is important for future research is the data-dependent bandwidth selection rule based on the calculation of a bootstrap MSE. This paper conjectures the validity of this approach and all the evidence provided in favor of this conjecture is based on simulations where the feasible bootstrap procedure is compared with the infeasible kernel weighted GMM estimator. Moreover, a supplementary appendix contains two further checks. First, I compute a 2SLS estimator that uses the EL bootstrap to pick the number of instruments and compare its performance to a plug-in rule. Second, I employ subsampling as an alternative to the

Table 4
Models 3 and 4 with $n = 500$.

Sigma	Estimator	Med. Bias	Med. AD	Mean SE	Dec. Rge.	Cov. Rate	Med. Bias	Med. AD	Mean SE	Dec. Rge.	Cov. Rate
		Model 3: $R_f^2 = 0.1$					Model 3: $R_f^2 = 0.01$				
0.1	GMM	0.031	0.076	0.014	0.294	0.926	0.080	0.139	0.043	0.473	0.918
	EL	−0.002	0.111	0.037	0.454	0.795	0.024	0.537	400.6	3.229	0.498
	DN2SLS	0.035	0.076	0.016	0.295	0.955	0.073	0.196	0.780	1.032	0.988
	S2SLS	0.028	0.076	0.014	0.286	0.948	0.077	0.135	0.043	0.456	0.961
	BsKGMM	0.030	0.082	0.015	0.300	0.928	0.080	0.140	0.048	0.485	0.911
	InKGMM	0.032	0.079	0.014	0.295	0.927	0.079	0.138	0.044	0.474	0.917
0.5	GMM	0.151	0.151	0.035	0.286	0.641	0.410	0.410	0.202	0.446	0.269
	EL	−0.004	0.107	0.033	0.423	0.834	0.137	0.567	174.8	2.844	0.593
	DN2SLS	0.096	0.142	0.039	0.449	0.808	0.369	0.430	6808	1.078	0.638
	S2SLS	0.116	0.129	0.029	0.332	0.767	0.415	0.415	0.192	0.437	0.319
	BsKGMM	0.136	0.150	0.035	0.347	0.658	0.410	0.410	0.201	0.476	0.281
	InKGMM	0.089	0.114	0.026	0.342	0.782	0.351	0.351	0.184	0.580	0.418
0.9	GMM	0.274	0.274	0.081	0.231	0.179	0.747	0.747	0.572	0.263	0.000
	EL	−0.004	0.094	0.025	0.372	0.927	0.114	0.436	88.53	2.401	0.704
	DN2SLS	0.057	0.137	0.042	0.481	0.876	0.469	0.576	39.84	1.317	0.519
	S2SLS	0.102	0.133	0.028	0.374	0.838	0.673	0.673	0.458	0.500	0.099
	BsKGMM	0.128	0.143	0.033	0.339	0.647	0.729	0.729	0.526	0.384	0.022
	InKGMM	0.096	0.127	0.028	0.360	0.782	0.429	0.432	0.296	0.691	0.398
		Model 4: $R_f^2 = 0.1$					Model 4: $R_f^2 = 0.01$				
0.1	GMM	0.031	0.079	0.014	0.299	0.931	0.079	0.136	0.043	0.479	0.919
	EL	−0.001	0.116	0.036	0.443	0.789	−0.006	0.557	188.8	3.288	0.495
	DN2SLS	0.036	0.077	0.484	0.285	0.958	0.092	0.217	744.0	1.262	0.990
	S2SLS	0.032	0.077	0.013	0.286	0.952	0.083	0.129	0.042	0.451	0.955
	BsKGMM	0.031	0.083	0.015	0.305	0.926	0.079	0.136	0.048	0.490	0.915
	InKGMM	0.031	0.082	0.015	0.301	0.926	0.080	0.135	0.044	0.475	0.918
0.5	GMM	0.149	0.150	0.035	0.286	0.634	0.413	0.413	0.202	0.444	0.253
	EL	−0.006	0.111	0.033	0.440	0.831	0.150	0.541	198.0	3.062	0.604
	DN2SLS	0.199	0.217	1.255	0.479	0.690	0.459	0.492	133.2	1.179	0.600
	S2SLS	0.149	0.150	0.034	0.280	0.672	0.419	0.419	0.203	0.398	0.268
	BsKGMM	0.148	0.152	0.035	0.309	0.649	0.415	0.415	0.203	0.456	0.263
	InKGMM	0.150	0.151	0.035	0.289	0.633	0.415	0.415	0.202	0.443	0.259
0.9	GMM	0.275	0.275	0.080	0.231	0.186	0.748	0.748	0.575	0.273	0.000
	EL	−0.005	0.095	0.027	0.403	0.912	0.079	0.435	63.97	2.872	0.709
	DN2SLS	0.299	0.359	11.87	1.009	0.695	0.785	0.812	67.15	1.341	0.382
	S2SLS	0.273	0.273	0.081	0.256	0.335	0.759	0.759	0.580	0.281	0.010
	BsKGMM	0.256	0.256	0.072	0.273	0.327	0.740	0.740	0.558	0.302	0.006
	InKGMM	0.269	0.269	0.080	0.265	0.310	0.749	0.749	0.574	0.271	0.000

DN2SLS: 2SLS with Donald and Newey's optimal instruments, S2SLS: shrinkage 2SLS; BsKGMM: Bootstrap Kernel Weighted GMM; InKGMM: Infeasible Kernel Weighted GMM. 200 Bootstrap replications.

bootstrap. The results are similar in all cases and nothing suggests that the bootstrap is having problems in picking the optimal number of instruments.

Finally, the simulations show an excellent performance of EL in terms of bias. This is not surprising since the higher order bias of EL is $O(M/n)$, as opposed to $O(M^2/n)$ for GMM. The price of this lower bias is a thicker-tailed sampling distribution in finite samples. Extending the current approach to an EL setting could be a promising direction for future research. Preliminary simulation results that compare a standard EL estimator to an EL estimator that uses the bootstrap introduced in Section 4 to select instruments, indicate that in most cases the latter estimator has a lower bias and MAD than the former.

Acknowledgements

The author was deeply grateful to Jack Porter and Bruce Hansen for their guidance and time. I am specially indebted to Jack Porter for thoughtful discussions and constant encouragement. The present version of the paper has been improved thanks to valuable suggestions made by the associate editor and two anonymous referees. I have also benefited from conversations with Ken West, Dennis Kristensen, Terry Iverson and participants at the UW-Madison econometrics seminar. Access to the Condor Cluster provided by the University of Wisconsin Computer Science Department is gratefully acknowledged.

Appendix A. Proofs of the propositions and theorems

Proposition A.1. *Suppose Assumptions 3.1 to 3.5 hold. Then,* $\|\widehat{D}_M - D\| = O_p(M/n^{1/2}) = o_p(1)$ *where* $\widehat{D}_M = \widehat{Q}_M'\mathcal{K}_M\widehat{\Omega}_M^{-1}\mathcal{K}_M\widehat{Q}_M$ *and* $D = Q'\Omega Q$.

Proof. The first step is to split the error in four parts:

$$\widehat{D}_M - D = H_1 + H_2 + H_3 + H_4$$

where $H_1 = Q_M'\mathcal{K}_M\Omega_M^{-1}\mathcal{K}_M Q_M - Q'\Omega^{-1}Q$, $H_2 = \widehat{Q}_M'\mathcal{K}_M\Omega_M^{-1}\mathcal{K}_M\widehat{Q}_M - Q_M'\mathcal{K}_M\Omega_M^{-1}\mathcal{K}_M Q_M$, $H_3 = -\widehat{Q}_M'\mathcal{K}_M\Omega_M^{-1}(\widehat{\Omega}_M - \Omega_M)\Omega_M^{-1}\mathcal{K}_M\widehat{Q}_M$ and $H_4 = \widehat{Q}_M'\mathcal{K}_M R_M\mathcal{K}_M\widehat{Q}_M$. The terms H_3 and H_4 contain a Taylor series expansion of $\widehat{\Omega}_M^{-1}$ around Ω_M^{-1} given by:

$$\widehat{\Omega}_M^{-1} = \Omega_M^{-1} - \Omega_M^{-1}(\widehat{\Omega}_M - \Omega_M)\Omega_M^{-1} + R_M \quad \text{(A.1)}$$

where $R_M = \Omega_M^{-1}(\widehat{\Omega}_M - \Omega_M)\Omega_M^{-1}(\widehat{\Omega}_M - \Omega_M)\Omega_M^{-1} = o_p(\|\widehat{\Omega}_M - \Omega_M\|)$.

(H1) Lemma B.1 shows that H_1 can be written as $H_1 = H_{11} + H_{12} + H_{13} + H_{14}$, where

$$H_{11} \equiv Q_M'\Omega_M^{-1}Q_M - Q'\Omega^{-1}Q = O(M^{-(2\alpha+1)}) \quad \text{(A.2)}$$

$$H_{12} \equiv Q_M'(I - \mathcal{K}_M)\Omega_M^{-1}(I - \mathcal{K}_M)Q_M = O(M^{-(2\alpha+1)}) \quad \text{(A.3)}$$

$$H_{13} \equiv -Q_M'\Omega_M^{-1}(I - \mathcal{K}_M)Q_M = O(M^{-(\alpha+1)}) \quad \text{(A.4)}$$

Table 5
Feasible versus infeasible optimal instruments. $R_f^2 = 0.1$ and $n = 500$.

Model	Param	$\sigma_{u\epsilon} = 0.1$			$\sigma_{u\epsilon} = 0.5$			$\sigma_{u\epsilon} = 0.9$		
		25th Quan.	50th Quan.	75th Quan.	25th Quan.	50th Quan.	75th Quan.	25th Quan.	50th Quan.	75th Quan.
1	M-DN	7	10	15	5	6	8	3	4	5
	s-Sk	0.825	0.928	0.99	0.37	0.47	0.59	0.13	0.15	0.18
	M-BG	23	25	25	15	21	25	8	10	13
	c-BG	0.900	0.950	0.950	0.150	0.500	0.900	0.150	0.350	0.750
	BsMSE	0.009	0.011	0.013	0.009	0.011	0.014	0.005	0.006	0.008
	BsBias	−0.005	0.014	0.030	0.030	0.039	0.050	0.025	0.031	0.048
	M-Inf	–	25	–	–	15	–	–	11	–
	c-Inf	–	0.95	–	–	0.25	–	–	0.20	–
	MSE-Inf	–	0.015	–	–	0.019	–	–	0.021	–
	Bias-Inf	–	0.030	–	–	0.050	–	–	0.064	–
2 Het	M-DN	5	6	8	5	6	8	5	5	6
	s-Sk	0.616	0.828	0.95	0.471	0.680	0.889	0.304	0.424	0.64
	M-BG	21	24	25	17	23	25	8	13	20
	c-BG	0.875	0.950	0.950	0.750	0.950	0.950	0.500	0.800	0.900
	BsMSE	0.003	0.004	0.006	0.003	0.004	0.006	0.003	0.003	0.005
	BsBias	−0.015	0.004	0.022	0.006	0.020	0.032	0.012	0.019	0.028
	M-Inf	–	25	–	–	14	–	–	7	–
	c-Inf	–	0.950	–	–	0.900	–	–	0.850	–
	MSE-Inf	–	0.007	–	–	0.012	–	–	0.013	–
	Bias-Inf	–	0.016	–	–	0.052	–	–	0.038	–
2 Hom	M-DN	6	7	11	5	5	6	4	5	5
	s-Sk	0.822	0.929	0.99	0.385	0.492	0.623	0.139	0.168	0.2
	M-BG	23	25	25	13	20	24	6	7	9
	c-BG	0.850	0.950	0.950	0.200	0.700	0.900	0.500	0.750	0.800
	BsMSE	0.008	0.010	0.013	0.009	0.011	0.014	0.005	0.007	0.009
	BsBias	−0.007	0.012	0.028	0.027	0.038	0.051	0.024	0.030	0.038
	M-Inf	–	25	–	–	10	–	–	6	–
	c-Inf	–	0.950	–	–	0.450	–	–	0.750	–
	MSE-Inf	–	0.015	–	–	0.020	–	–	0.006	–
	Bias-Inf	–	0.030	–	–	0.043	–	–	0.021	–
3	M-DN	15	22	24	2	4	8	1	1	2
	s-Sk	0.801	0.910	0.98	0.326	0.421	0.55	0.107	0.128	0.16
	M-BG	24	25	25	21	24	25	13	16	19
	c-BG	0.950	0.950	0.95	0.500	0.950	0.95	0.050	0.050	0.100
	BsMSE	0.009	0.011	0.01	0.010	0.012	0.02	0.007	0.009	0.012
	BsBias	−0.004	0.014	0.032	0.043	0.055	0.071	0.036	0.043	0.052
	M-Inf	–	25	–	–	23	–	–	12	–
	c-Inf	–	0.950	–	–	0.050	–	–	0.050	–
	MSE-Inf	–	0.015	–	–	0.025	–	–	0.027	–
	Bias-Inf	–	0.031	–	–	0.079	–	–	0.075	–
4	M-DN	22	24	25	4	8	15	1	2	3
	s-Sk	0.847	0.940	0.99	0.415	0.515	0.63	0.151	0.179	0.21
	M-BG	25	25	25	22	25	25	15	19	23
	c-BG	0.950	0.950	0.95	0.950	0.950	0.95	0.900	0.950	0.95
	BsMSE	0.009	0.011	0.01	0.010	0.013	0.02	0.011	0.015	0.02
	BsBias	−0.004	0.014	0.034	0.052	0.067	0.089	0.080	0.100	0.120
	M-Inf	–	25	–	–	25	–	–	18	–
	c-Inf	–	0.950	–	–	0.950	–	–	0.950	–
	MSE-Inf	–	0.015	–	–	0.034	–	–	0.079	–
	Bias-Inf	–	0.032	–	–	0.150	–	–	0.263	–

True values are denoted with "Inf". M-DN: instruments selected by Donald and Newey's, M-BG: instruments selected by Bootstrap Kernel, c-BG: trapezoidal constant selected by Bootstrap Kernel, s-SK: 2SLS shrinkage parameter. BsMSE and BsBias denote the bootstrap versions of the MSE and Bias, respectively.

$$H_{14} \equiv -Q_M'(I - \mathcal{K}_M)\Omega_M^{-1}Q_M = O(M^{-(\alpha+1)}). \tag{A.5}$$

It follows that $H_1 = O(M^{-(\alpha+1)})$.

(H2) Lemma B.2 shows that $H_2 = H_{21} + H_{22} + H_{23}$, where

$$H_{21} \equiv (\widehat{Q}_M - Q_M)'\mathcal{K}_M\Omega_M^{-1}\mathcal{K}_M(\widehat{Q}_M - Q_M) = O_p(M/n) \tag{A.6}$$

$$H_{22} \equiv Q_M'\mathcal{K}_M\Omega_M^{-1}\mathcal{K}_M(\widehat{Q}_M - Q_M) = O_p(n^{-1/2}) \tag{A.7}$$

$$H_{23} \equiv (\widehat{Q}_M - Q_M)'\mathcal{K}_M\Omega_M^{-1}\mathcal{K}_M Q_M = O_p(n^{-1/2}) \tag{A.8}$$

which means that $H_2 = O_p(n^{-1/2})$ since $M = o(n^{1/2})$.

(H3) Lemma B.3 shows that $H_3 = H_{31} + H_{32} + H_{33} + H_{34}$, where

$$\begin{aligned} H_{31} &\equiv -(\widehat{Q}_M - Q_M)'\mathcal{K}_M\Omega_M^{-1}(\widehat{\Omega}_M - \Omega_M)\Omega_M^{-1}\mathcal{K}_M(\widehat{Q}_M - Q_M) \\ &= o_p(M/n) \end{aligned} \tag{A.9}$$

$$\begin{aligned} H_{32} &\equiv -Q_M'\mathcal{K}_M\Omega_M^{-1}(\widehat{\Omega}_M - \Omega_M)\Omega_M^{-1}\mathcal{K}_M(\widehat{Q}_M - Q_M) \\ &= O_p(M/n) \end{aligned} \tag{A.10}$$

$$\begin{aligned} H_{33} &\equiv -(\widehat{Q}_M - Q_M)'\mathcal{K}_M\Omega_M^{-1}(\widehat{\Omega}_M - \Omega_M)\Omega_M^{-1}\mathcal{K}_M Q_M \\ &= O_p(M/n) \end{aligned} \tag{A.11}$$

$$\begin{aligned} H_{34} &\equiv Q_M'\mathcal{K}_M\Omega_M^{-1}(\widehat{\Omega}_M - \Omega_M)\Omega_M^{-1}\mathcal{K}_M Q_M \\ &= O_p(n^{-1/2}) \end{aligned} \tag{A.12}$$

which means that $H_3 = O_p(n^{-1/2})$ since $M = o(n^{1/2})$.

(H4) Finally, $H_4 = o_p(n^{-1/2})$ because this term depends on $R_M = o_p(\|\widehat{\Omega}_M - \Omega_M\|)$ meaning that $H_4 = o_p(\|H_3\|)$.

Therefore, $\|\widehat{D}_M - D\| = O_p(\max\{n^{-1/2}, M^{-(\alpha+1)}\}) = O_p(M/n^{1/2}) = o_p(1)$. ■

Table 6
Model 1: EL as the preliminary estimator.

Sigma	Estimator	Med. Bias	Med. AD	Mean SE	Dec. Rge.	Cov. Rate	Med. Bias	Med. AD	Mean SE	Dec. Rge.	Cov. Rate
		$R_f^2 = 0.1, n = 100$					$R_f^2 = 0.01, n = 100$				
0.1	DN2SLS	0.045	0.159	0.792	0.606	0.975	0.098	0.312	465.0	1.974	0.991
	S2SLS	0.062	0.124	0.040	0.449	0.950	0.099	0.158	0.071	0.572	0.953
	BsKGMM	0.056	0.152	0.061	0.586	0.840	0.087	0.191	0.110	0.735	0.827
	DN2SLS*	0.036	0.173	0.903	0.720	0.982	0.085	0.443	499.5	3.269	0.994
	S2SLS*	0.056	0.131	0.057	0.513	0.960	0.095	0.176	0.130	0.736	0.964
	BsKGMM*	0.046	0.157	0.094	0.628	0.875	0.085	0.231	0.377	0.949	0.878
0.5	DN2SLS	0.215	0.273	11.56	0.682	0.780	0.464	0.538	277.1	1.853	0.757
	S2SLS	0.321	0.324	0.127	0.453	0.519	0.476	0.477	0.278	0.500	0.323
	BsKGMM	0.289	0.298	0.129	0.570	0.503	0.479	0.483	0.298	0.636	0.286
	DN2SLS*	0.182	0.262	12.11	0.801	0.824	0.444	0.610	314.1	2.958	0.848
	S2SLS*	0.303	0.313	0.130	0.542	0.596	0.473	0.482	0.318	0.637	0.452
	BsKGMM*	0.243	0.280	0.146	0.686	0.591	0.470	0.485	0.706	0.832	0.388
0.9	DN2SLS	0.274	0.313	13.78	0.686	0.649	0.774	0.802	48.17	1.317	0.371
	S2SLS	0.490	0.490	0.257	0.458	0.233	0.864	0.864	0.743	0.295	0.008
	BsKGMM	0.361	0.364	0.185	0.637	0.360	0.852	0.852	0.743	0.374	0.011
	DN2SLS*	0.243	0.311	20.07	0.941	0.736	0.733	0.788	59.92	2.260	0.527
	S2SLS*	0.403	0.404	0.209	0.539	0.418	0.841	0.841	0.732	0.442	0.062
	BsKGMM*	0.187	0.288	1.182	0.939	0.719	0.833	0.835	1.123	0.566	0.114
		$R_f^2 = 0.1, n = 500$					$R_f^2 = 0.01, n = 500$				
0.1	DN2SLS	0.027	0.088	0.016	0.316	0.958	0.068	0.185	2029	0.906	0.988
	S2SLS	0.035	0.080	0.014	0.291	0.951	0.086	0.137	0.044	0.463	0.954
	BsKGMM	0.034	0.085	0.016	0.307	0.915	0.086	0.138	0.049	0.486	0.916
	DN2SLS*	0.028	0.088	0.017	0.318	0.958	0.054	0.275	2089	1.520	0.993
	S2SLS*	0.037	0.081	0.015	0.293	0.949	0.073	0.162	0.096	0.657	0.975
	BsKGMM*	0.036	0.089	0.017	0.308	0.926	0.075	0.151	15.89	0.613	0.922
0.5	DN2SLS	0.069	0.104	0.023	0.344	0.876	0.360	0.399	1.781	0.853	0.674
	S2SLS	0.136	0.143	0.033	0.315	0.727	0.410	0.410	0.197	0.417	0.304
	BsKGMM	0.092	0.114	0.028	0.366	0.748	0.411	0.411	0.202	0.482	0.271
	DN2SLS*	0.062	0.102	0.023	0.346	0.894	0.289	0.416	466.4	1.583	0.825
	S2SLS*	0.124	0.140	0.033	0.355	0.769	0.382	0.397	0.216	0.655	0.512
	BsKGMM*	0.054	0.111	0.028	0.392	0.849	0.380	0.392	5.941	0.702	0.872
0.9	DN2SLS	0.093	0.121	0.026	0.339	0.834	0.487	0.525	5.116	0.897	0.451
	S2SLS	0.152	0.163	0.038	0.338	0.750	0.714	0.714	0.502	0.399	0.054
	BsKGMM	0.095	0.116	0.023	0.314	0.786	0.714	0.714	0.490	0.486	0.047
	DN2SLS*	0.090	0.121	0.026	0.342	0.847	0.366	0.484	166.2	1.908	0.712
	S2SLS*	0.135	0.151	0.036	0.357	0.798	0.542	0.542	0.404	0.649	0.332
	BsKGMM*	0.041	0.134	0.054	0.530	0.899	0.452	0.472	0.443	0.944	0.496

DN2SLS: 2SLS with Donald and Newey's optimal instruments, S2SLS: shrinkage 2SLS; BsKGMM: Bootstrap Kernel Weighted GMM (all with usual preliminary estimators). * means that EL is used as the preliminary estimator.

Proposition A.2. *Suppose Assumptions 3.1 to 3.5 hold. Then,* $\|\widehat{d}_M - d_0\| = O_p(\max\{M/n^{1/2}, M^{-\alpha}\}) = o_p(1)$ *where* $\widehat{d}_M = \widehat{Q}_M'\mathcal{K}_M\widehat{\Omega}_M^{-1}\mathcal{K}_M V_M$ *and* $d_0 = Q'\Omega^{-1}V$.

Proof. Lemma B.4 shows that $\widehat{d}_M$ can be decomposed in eight terms with the following order of magnitudes:

$$d_0 = Q'\Omega^{-1}V = O_p(1) \quad \text{and} \quad Ed_0d_0' = D \tag{A.13}$$

$$d_1 = Q_M'\Omega_M^{-1}V_M - Q'\Omega^{-1}V = O_p(M^{-\alpha}) \tag{A.14}$$

$$d_2 = Q_M'(I - \mathcal{K}_M)\Omega_M^{-1}(I - \mathcal{K}_M)V_M = O(M^{-(2\alpha+1)}) \tag{A.15}$$

$$d_3 = -Q_M'(I - \mathcal{K}_M)\Omega_M^{-1}V_M - Q_M'\Omega_M^{-1}(I - \mathcal{K}_M)V_M = O_p(M^{-\alpha}) \tag{A.16}$$

$$d_4 = -Q_M'\mathcal{K}_M\Omega_M^{-1}(\widehat{\Omega}_M - \Omega_M)\Omega_M^{-1}\mathcal{K}_M V_M = O_p(n^{-1/2}) \tag{A.17}$$

$$d_5 = (\widehat{Q}_M - Q_M)'\mathcal{K}_M\Omega_M^{-1}\mathcal{K}_M V_M = O_p(M/n^{1/2}) \tag{A.18}$$

$$d_6 = -(\widehat{Q}_M - Q_M)'\mathcal{K}_M\Omega_M^{-1}(\widehat{\Omega}_M - \Omega_M)\Omega_M^{-1}\mathcal{K}_M V_M = O_p(M/n) \tag{A.19}$$

$$d_7 = \widehat{Q}_M'\mathcal{K}_M R_M \mathcal{K}_M V_M = o_p(M/\sqrt{n}), \tag{A.20}$$

meaning that $\|\widehat{d}_M - d_0\| = \max\{O_p(M/n^{1/2}), O(M^{-\alpha})\} = o_p(1)$. ■

Proof of Theorem 3.1. Note that $\sqrt{n}(\widehat{\delta}_{n,M} - \delta) = \widehat{D}_M^{-1}\widehat{d}_M = [D^{-1} - D^{-1}(\widehat{D}_M - D)D^{-1} + o_p(M/n^{1/2})](\widehat{d}_M - d_0 + d_0)$. This means that $\sqrt{n}(\widehat{\delta}_{n,M} - \delta) - D^{-1}d_0 = D^{-1}(\widehat{d}_M - d_0) - D^{-1}(\widehat{D}_M - D)D^{-1}d_0 + o_p(M/n^{1/2}) = O_p(\max\{M/n^{1/2}, M^{-\alpha}\}) = o_p(1)$ by Propositions A.1 and A.2. ■

Proof of Theorem 3.2. The asymptotic MSE is given by $AMSE(M, c, \ell) = n\ell' D^{1/2}E[(\widehat{\delta}_{n,M} - \delta)(\widehat{\delta}_{n,M} - \delta)']D^{1/2}\ell - 1$ where

$$\sqrt{n}(\widehat{\delta}_{n,M} - \delta) = D^{-1}[I - (H_1 + H_2 + H_3 + H_4)D^{-1}] \times \sum_{k=0}^{7} d_k + o_p(M/n^{1/2}).$$

I consider terms in $n'D^{1/2}(\widehat{\delta}_{n,M} - \delta)(\widehat{\delta}_{n,M} - \delta)'D^{1/2}$ that depend on M, n and c that are largest in probability. These have the form $E(d_jd_k')$, $E(d_kd_0')D^{-1}H_j$ and $H_kD^{-1}E(d_0d_0')D^{-1}H_j'$.

Terms vanishing at rate $M^{-(\alpha+1)}$: these are $E(d_3d_0')$ and $E(d_0d_0')D^{-1}(H_{13} + H_{14})$. By Lemma B.7 these two terms cancel out because they are of opposite sign.

Terms vanishing at rate $M^{-(2\alpha+1)}$: these are $E(d_2d_0')$, $-E(d_0d_0')D^{-1}H_{12}'$, $E(d_1d_0')$, $-E(d_0d_0')D^{-1}H_{11}$, (Ed_3d_3') and $E(d_1d_1')$. The results in Lemmas B.8 and B.9 show that the first four terms cancel out, while by Lemma B.10 $M^{(2\alpha+1)}E(d_1d_1') = \mathbf{C}_3$ for p.d. matrix with finite norm. Thus, we can write $M^{(2\alpha+1)}\ell' D^{-1/2}$

Table 7
Model 1: (top) Fixed c versus optimal c. (bottom) Unordered instruments.

Sigma	Estimator	Med. Bias	Med. AD	Mean SE	Dec. Rge.	Cov. Rate	Med. Bias	Med. AD	Mean SE	Dec. Rge.	Cov. Rate
		$R_f^2 = 0.1, n = 500$					$R_f^2 = 0.01, n = 500$				
0.5	GMM	0.156	0.157	0.036	0.285	0.628	0.410	0.410	0.204	0.437	0.261
	DN2SLS	0.069	0.104	0.023	0.344	0.876	0.360	0.399	1.781	0.853	0.674
	BsKGMM	0.092	0.115	0.029	0.372	0.747	0.409	0.410	0.202	0.479	0.264
	BsKGMM:c=1	0.098	0.121	0.030	0.381	0.754	0.406	0.407	0.203	0.473	0.267
	BsKGMM:c=0.5	0.074	0.106	0.025	0.356	0.799	0.379	0.381	0.184	0.499	0.350
	BsKGMM:c=0	0.077	0.102	0.021	0.314	0.805	0.336	0.337	0.163	0.546	0.430
	BsKGMM*	0.070	0.105	0.024	0.353	0.808	0.334	0.360	145.2	0.814	0.514
	BsKGMM*:c=1	0.092	0.132	0.035	0.402	0.822	0.326	0.402	171.3	1.074	0.640
	BsKGMM*:c=0.5	0.070	0.118	0.032	0.403	0.854	0.293	0.374	171.3	1.160	0.673
	BsKGMM*:c=0	0.005	0.162	0.087	0.650	0.953	0.139	0.547	487.5	2.881	0.975
0.9	GMM	0.278	0.278	0.081	0.235	0.177	0.748	0.748	0.574	0.262	0.000
	DN2SLS	0.093	0.121	0.026	0.339	0.834	0.487	0.525	5.116	0.897	0.451
	BsKGMM	0.094	0.115	0.023	0.314	0.779	0.714	0.714	0.490	0.473	0.051
	BsKGMM:c=1	0.095	0.119	0.024	0.323	0.815	0.718	0.718	0.504	0.464	0.066
	BsKGMM:c=0.5	0.068	0.109	0.023	0.343	0.848	0.661	0.661	0.434	0.508	0.092
	BsKGMM:c=0	0.094	0.116	0.023	0.313	0.751	0.606	0.606	0.387	0.395	0.042
	BsKGMM*	0.075	0.109	0.023	0.339	0.817	0.430	0.443	0.316	0.774	0.396
	BsKGMM*:c=1	0.130	0.163	0.054	0.452	0.696	0.495	0.569	17.05	1.248	0.468
	BsKGMM*:c=0.5	0.098	0.140	0.046	0.436	0.756	0.446	0.518	17.02	1.274	0.488
	BsKGMM*:c=0	0.003	0.161	0.091	0.638	0.932	0.183	0.473	171.1	2.747	0.854
		$R_f^2 = 0.1, n = 500$					$R_f^2 = 0.01, n = 500$				
0.1	GMM	0.038	0.080	0.015	0.297	0.920	0.084	0.140	0.045	0.483	0.916
	EL	0.011	0.118	0.039	0.478	0.775	0.044	0.547	228.9	3.052	0.479
	DN2SLS	0.035	0.080	0.042	0.295	0.955	0.100	0.214	904.3	1.477	0.989
	S2SLS	0.035	0.080	0.014	0.294	0.953	0.086	0.137	0.044	0.467	0.953
	BsKGMM	0.036	0.081	0.016	0.300	0.912	0.084	0.141	0.049	0.494	0.917
0.5	GMM	0.156	0.157	0.036	0.285	0.628	0.410	0.410	0.204	0.437	0.261
	EL	0.009	0.114	0.035	0.448	0.824	0.152	0.545	210.5	3.044	0.580
	DN2SLS	0.215	0.248	1116	1.023	0.702	0.473	0.521	5408	1.710	0.593
	S2SLS	0.160	0.161	0.036	0.276	0.664	0.417	0.417	0.205	0.410	0.279
	BsKGMM	0.142	0.145	0.033	0.297	0.675	0.411	0.411	0.204	0.463	0.273
0.9	GMM	0.278	0.278	0.081	0.235	0.177	0.748	0.748	0.574	0.262	0.000
	EL	−0.002	0.094	0.027	0.392	0.914	0.135	0.439	105.2	2.580	0.704
	DN2SLS	0.775	0.915	4342	3.149	0.740	0.883	0.900	37.81	1.386	0.421
	S2SLS	0.303	0.303	0.099	0.240	0.155	0.765	0.765	0.590	0.245	0.004
	BsKGMM	0.242	0.242	0.064	0.255	0.316	0.731	0.731	0.550	0.298	0.003

DN2SLS: 2SLS with Donald and Newey's optimal instruments, S2SLS: shrinkage 2SLS; BsKGMM: Bootstrap Kernel Weighted GMM ($c = 1$, $c = 0.5$ and $c = 0$ denote cases where c is fixed at the corresponding values). * means that an IV estimator with the first instrument is used as the preliminary estimator.

$E(d_1 d_1')D^{-1/2}\ell = \mathcal{C}_C$. Finally, it follows from Lemmas B.5 and C.8 that,

$$\ell' D^{-1/2} E(d_3 d_3') D^{-1/2}\ell = \frac{\mathcal{C}_B F_{\alpha,\sigma,\kappa}(c)}{M^{2\alpha+1}} + o(M^{-(2\alpha+1)})$$

for some nonnegative constant $\mathcal{C}_B$ and $F_{\alpha,\sigma,\kappa}(c) = \frac{\kappa^2 c^2}{\sigma^2(1-c)^2}\left[\frac{1}{c^{2\alpha+1}(4\alpha^2-1)\alpha} + \frac{c(2\alpha-1)-\alpha}{c^2\alpha(2\alpha-1)} - \frac{1}{2\alpha+1}\right]$. This constitute the variance term.

Terms growing with M at rate M^2/n : the largest of these terms is $E(d_5 d_5')$. By Lemma B.6,

$$\ell' D^{-1/2} E(d_5 d_5') D^{-1/2}\ell = \frac{M^2}{n}\left(\int k_c^2(x)\mathrm{d}x\right)^2 \mathcal{C}_A + o(M^2/n).$$

for some constant $\mathcal{C}_A > 0$ and $\int k_c^2(x)\mathrm{d}x = \frac{2}{3}(1+2c)$. This constitutes the bias term in the AMSE. ■

Appendix B. Primary lemmas

Lemma B.1. *Suppose Assumptions 3.1 to 3.5 hold and let H_1 be $Q_M' \mathcal{K}_M \Omega_M^{-1} \mathcal{K}_M Q_M - Q'\Omega^{-1}Q$. This term can be decomposed in the four matrices given by Eq. (A.2) to (A.5).*

Proof. I analyze each of the four terms separately starting with H_{11}.

(H_{11}) Let $H_{11} = H_{111} + H_{112}$ with $H_{111} = Q'(\Omega_M^{*-1} - \Omega^{-1})Q$ and $H_{112} = Q_M'\Omega_M^{-1}Q_M - Q'\Omega_M^{*-1}Q$, where Ω_M^* is an infinite dimensional matrix defined by

$$\Omega_M^* = \begin{bmatrix} \Omega_M & 0 \\ 0 & \sigma^2 I_\infty \end{bmatrix} \tag{B.1}$$

where σ^2 is defined in Assumption 3.2. By Lemma C.7,

$$\|H_{112}\| = \left\| -\sigma^{-2}\sum_{j=M+1}^{\infty}\Gamma_j\Gamma_j'\right\| \le \sigma^{-2}\sum_{j=M+1}^{\infty}\|\Gamma_j\Gamma_j'\|$$
$$= \sigma^{-2}\sum_{j=M+1}^{\infty}\|\Gamma_j\|^2 = M^{-(2\alpha+1)}\frac{1}{\sigma^2(2\alpha+1)} + o(M^{-(2\alpha+1)}).$$

On the other hand, note that $(\Omega_M^{*-1} - \Omega^{-1}) = \Omega_M^{*-1}(\Omega - \Omega_M^*)\Omega^{-1}$ and define $a_k = \sum_{j=1}^{M}\Gamma_j\vartheta_{jk}$. By Lemma C.9 this term decays as $\|\Gamma_k\|$, so that

$$\|H_{111}\| \le \sum_{j_1,j_2,j_3,j_4=1}^{\infty}\|\Gamma_{j_1}\|\,|\vartheta_{j_1j_2}^*|\,|\omega_{j_2j_3} - \omega_{j_2j_3}^*|\,|\vartheta_{j_3j_4}|\,\|\Gamma_{j_4}\|$$
$$\le \sum_{j_3=M+1}^{\infty}\|a_{j_3}^*\|\,|\omega_{j_3j_3} - \sigma^2|\,\|a_{j_3}\|$$

$$+2\sum_{j_3=M+1}^{\infty}\sum_{j_2=1}^{\infty}\|a_{j_2}^*\|\,|\omega_{j_2j_3}|\,\|a_{j_3}\| \leq M^{-(2\alpha+1)}\times C_1$$

since the sums go to zero at rate $M^{-(2\alpha+1)}$ by Lemma C.7. Thus, $\|H_{11}\| = O(M^{-(2\alpha+1)})$.

(H_12) First I note that the trapezoidal kernel satisfies $1-k(j/M)\equiv 0$ for all $j \leq cM$ where $c \in (0,1)$. Thus, I can write $H_{12} \equiv Q_M'(I-\mathcal{K}_M)\Omega_M^{-1}(I-\mathcal{K}_M)Q_M$ as

$$H_{12} \equiv \sum_{j=1}^{M}\sum_{l=1}^{M}\Gamma_j(1-k(j/M))\vartheta_{lj}^M(1-k(l/M))\Gamma_l' \equiv \sum_{j=cM+1}^{M}\sum_{l=cM+1}^{M}\Gamma_j(1-k(j/M))\vartheta_{lj}^M(1-k(l/M))\Gamma_l'.$$

Then note that $1-k(j/M)=\frac{c}{1-c}\left(\frac{j}{cM}-1\right)$ for $c<j/M\leq 1$. To simplify notation define

$$\bar{k}_\alpha(j/M)=\frac{1-k(j/M)}{(j/M)^\alpha}=\frac{c}{1-c}\left(\frac{1/c}{(j/M)^{\alpha-1}}-\frac{1}{(j/M)^\alpha}\right) \quad (B.2)$$

for $c<j/M\leq 1$ (being zero otherwise) and notice that $\bar{k}_\alpha$ goes to 0 for $j=cM+1$ as $M\to\infty$, while $\bar{k}_\alpha\to 1$ for $j\to M$ as $M\to\infty$. Now I write H_{12} as

$$H_{12} \equiv M^{-2\alpha}\sum_{j=cM+1}^{M}\sum_{l=cM+1}^{M} j^\alpha\Gamma_j\bar{k}_\alpha(j/M)\vartheta_{lj}^M\bar{k}_\alpha(l/M)l^\alpha\Gamma_l'$$
$$= M^{-2\alpha-1}M\sum_{j=cM+1}^{M}\sum_{l=cM+1}^{M} j^\alpha\Gamma_j\bar{k}_\alpha(j/M)\vartheta_{lj}^M\bar{k}_\alpha(l/M)l^\alpha\Gamma_l' = O(M^{-(2\alpha+1)}).$$

Since $\lim M\sum_{j=cM+1}^{M}\sum_{l=cM+1}^{M} j^\alpha\Gamma_j\bar{k}_\alpha(j/M)\vartheta_{lj}^M\bar{k}_\alpha(l/M)l^\alpha\Gamma_l' < \infty$ by Lemma C.8.

(H_{13}) Write H_{13} as $-\sum_{j=1}^{M}\sum_{l=1}^{M}\Gamma_j\vartheta_{lj}^M(1-k(l/M))\Gamma_l'$. Using $\bar{k}_\alpha(j/M)$ as defined above,

$$M^{\alpha+1}H_{13} = -M^{\alpha+1}\sum_{j=1}^{M}\sum_{l=cM+1}^{M}\Gamma_j\vartheta_{lj}^M(1-k(l/M))\Gamma_l'$$
$$= -M\sum_{j=1}^{cM}\sum_{l=cM+1}^{M}\Gamma_j\vartheta_{lj}^M\bar{k}_\alpha(l/M)l^\alpha\Gamma_l' - M\sum_{j=cM+1}^{M}\sum_{l=cM+1}^{M}\Gamma_j\vartheta_{lj}^M\bar{k}_\alpha(l/M)l^\alpha\Gamma_l'$$
$$= (A)+(B) = O(1)$$

where the last equality follows after noticing that $(A)=o(1)$ by the same arguments as those in display (C.2), and by Lemma C.8

$$|(B)|\leq M\sum_{j=cM+1}^{M}\sum_{l=cM+1}^{M} j^\alpha\|\Gamma_j\||\vartheta_{lj}^M|\bar{k}_\alpha(l/M)l^\alpha\|\Gamma_l'\| = O(1).$$

(H_{14}) This term can be analyzed in the same way as H_{13}. ■

Lemma B.2. *Let Assumptions 3.1 to 3.5 hold and let H_2 be defined as $H_2=\widehat{Q}_M'\mathcal{K}_M\Omega_M^{-1}\mathcal{K}_M\widehat{Q}_M - Q_M'\mathcal{K}_M\Omega_M^{-1}\mathcal{K}_MQ_M$. This term can be decomposed as in Eq. (A.6) to (A.8).*

Proof. I look again at each term separately, starting with H_{21}.

(H_{21}) Use the fact that $\|\widehat{\Gamma}_j-\Gamma_j\|=O_p(n^{-1/2})$ and apply the CSI. Then,

$$E\|H_{21}\| \leq \sum_{j=1}^{M}\sum_{l=1}^{M}E\left(\|(\widehat{\Gamma}_j-\Gamma_j)k(j/M)\vartheta_{lj}^Mk(l/M)(\widehat{\Gamma}_l-\Gamma_l)'\|\right)$$
$$\leq \sum_{j=1}^{M}\sum_{l=1}^{M}E\left(\|\widehat{\Gamma}_j-\Gamma_j\|^2\right)^{1/2}|\vartheta_{lj}^M|E\left(\|\widehat{\Gamma}_l-\Gamma_l\|^2\right)^{1/2}$$
$$\leq C_1n^{-1}\times\sum_{j=1}^{M}\sum_{l=1}^{M}\left|\vartheta_{lj}^M\right| = O(M/n)$$

where C_1 is some constant and $\sum_{j=1}^{M}\sum_{l=1}^{M}|\vartheta_{lj}^M| = O(M)$. The result follows from the MI.

(H_{22}) Use MI and $E\left(\|\widehat{\Gamma}_j-\Gamma_j\|^2\right)^{1/2}=C_1n^{-1/2}$ uniformly in j by Assumption 3.3. Then,

$$E\|H_{22}\| \leq \sum_{j=1}^{M}\sum_{l=1}^{M}E\left(\|\Gamma_jk(j/M)\vartheta_{lj}^Mk(l/M)(\widehat{\Gamma}_l-\Gamma_l)'\|\right)$$
$$\leq \sum_{j=1}^{M}\sum_{l=1}^{M}\left\|\Gamma_j\vartheta_{lj}^M\right\|E\left\|\widehat{\Gamma}_l-\Gamma_l\right\|$$
$$\leq C_1n^{-1/2}\times\underbrace{\sum_{j=1}^{M}\sum_{l=1}^{M}\left\|\Gamma_j\vartheta_{lj}^M\right\|}_{<\infty} = O(n^{-1/2}).$$

(H_{23}) This term is similar to H_{22}. ■

Lemma B.3. *Let Assumptions 3.1 to 3.5 hold and let H_3 be defined as $H_3=-\widehat{Q}_M'\mathcal{K}_M\Omega_M^{-1}(\widehat{\Omega}_M-\Omega_M)\Omega_M^{-1}\mathcal{K}_M\widehat{Q}_M$. This term can be decomposed as shown in Eq. (A.9) to (A.12).*

Proof. Start with H_{31}.

$$H_{31} = \sum_{j_1,j_2,j_3,j_4}^{M}(\widehat{\Gamma}_{j_1}-\Gamma_{j_1})k(j_1/M)\vartheta_{j_1j_2}^M \times(\widehat{\omega}_{j_2j_3}-\omega_{j_2j_3})\vartheta_{j_3j_4}^Mk(j_4/M)(\widehat{\Gamma}_{j_4}'-\Gamma_{j_4}')$$

$$E\|H_{31}\| \leq \sum_{j_1,j_2,j_3,j_4}^{M}E\left\|(\widehat{\Gamma}_{j_1}-\Gamma_{j_1})k(j_1/M)\vartheta_{j_1j_2}^M \times(\widehat{\omega}_{j_2j_3}-\omega_{j_2j_3})\vartheta_{j_3j_4}^Mk(j_4/M)(\widehat{\Gamma}_{j_4}'-\Gamma_{j_4}')\right\|$$
$$\leq \sum_{j_1,j_2,j_3,j_4}^{M}\underbrace{E\left(\left\|\widehat{\Gamma}_{j_1}-\Gamma_{j_1}\right\|^2\right)^{1/2}}_{O(n^{-1/2})}\left|\vartheta_{j_1j_2}^M\right|\left|\vartheta_{j_3j_4}^M\right| \times\left(E\left\|(\widehat{\omega}_{j_2j_3}-\omega_{j_2j_3})(\widehat{\Gamma}_{j_4}'-\Gamma_{j_4}')\right\|^2\right)^{1/2}.$$

Now use $E\left|\widehat{\omega}_{j_2j_3}-\omega_{j_2j_3}\right|^2=O(n^{-1})$ and the CSI twice. The result then follows from the MI as follows,

$$E\|H_{31}\| \leq C_1n^{-1/2}\sum_{j_1,j_2,j_3,j_4}^{M}\left|\vartheta_{j_1j_2}^M\right|\left|\vartheta_{j_3j_4}^M\right| \times\left(\underbrace{\left(E\left|\widehat{\omega}_{j_2j_3}-\omega_{j_2j_3}\right|^4\right)^{1/2}}_{O(n^{-1})}\underbrace{\left(E\left\|(\widehat{\Gamma}_{j_4}'-\Gamma_{j_4}')\right\|^4\right)^{1/2}}_{O(n^{-1})}\right)^{1/2}$$
$$\leq C_1n^{-3/2}\sum_{j_1,j_2}^{M}\left|\vartheta_{j_1j_2}^M\right| \times\sum_{j_3,j_4}^{M}\left|\vartheta_{j_3j_4}^M\right| = O(M^2/n^{3/2}) = o(M/n).$$

(H_{32}) Since $E\left|\widehat{\omega}_{j_2j_3}-\omega_{j_2j_3}\right|^2=O(n^{-1})$,

$$H_{32} = \sum_{j_1,j_2,j_3,j_4}^{M}\Gamma_{j_1}k(j_1/M)\vartheta^M_{j_1j_2} \times(\widehat{\omega}_{j_2j_3}-\omega_{j_2j_3})\vartheta^M_{j_3j_4}k(j_4/M)(\widehat{\Gamma}'_{j_4}-\Gamma'_{j_4})$$

$$E\|H_{32}\| \le \sum_{j_1,j_2,j_3,j_4}^{M}\left\|\Gamma_{j_1}\vartheta^M_{j_1j_2}\right\|\left(E\left|\widehat{\omega}_{j_2j_3}-\omega_{j_2j_3}\right|^2\right)^{1/2} \times\left|\vartheta^M_{j_3j_4}\right|\left(E\left\|\widehat{\Gamma}'_{j_4}-\Gamma'_{j_4}\right\|^2\right)^{1/2} \le C_1 n^{-1}\underbrace{\sum_{j_1,j_2}^{M}\left\|\Gamma_{j_1}\vartheta^M_{j_1j_2}\right\|}_{<\infty}\underbrace{\sum_{j_3,j_4}^{M}\left|\vartheta^M_{j_3j_4}\right|}_{O(M)}=O(M/n)$$

and therefore $\|H_{32}\|=O_p(M/n)$ by the MI.

(H_{33}) This term is similar to H_{32}.

(H_{34}) Note that $H_{34}\equiv\sum_{j_2,j_3}^{M}a^M_{j_2}(\widehat{\omega}_{j_2j_3}-\omega_{j_2j_3})a^M_{j_3}$ where $a^M_j=\sum_l^M\Gamma_l k(l/M)\vartheta^M_{lj}$ is summable since $\sum_{j=1}^{\infty}\|a^M_j\|\le\sum_{j=1}^{\infty}\sum_{l=1}^{\infty}\left\|\Gamma_l\vartheta^M_{lj}\right\|<\infty$. Then,

$$E\|H_{34}\|\le\sum_{j_2,j_3}^{M}\left\|a^M_{j_2}\right\|\left(E\left|\widehat{\omega}_{j_2j_3}-\omega_{j_2j_3}\right|^2\right)^{1/2}\left\|a^M_{j_3}\right\|=O(n^{-1/2}).\quad\blacksquare$$

Lemma B.4. *Suppose Assumptions 3.1 to 3.5 hold and let $\widehat{d}_M$ be $\widehat{Q}'_M\mathcal{K}_M\widehat{\Omega}^{-1}_M\mathcal{K}_M V_M$. This term can be decomposed in the eight vectors given by Eq. (A.13) to (A.20).*

Proof. As in the other lemmas, I analyze term by term starting with d_0.

(d_0) Recall that $V_M=\frac{1}{\sqrt{n}}\sum x_{Mi}\varepsilon_i$, so that $EV_M=0$ and $V_M=O_p(1)$. This directly implies that $Ed_0=0$ and $d_0=O_p(1)$. In addition, notice that $E(d_0d_0')=Q'\Omega^{-1}E(VV')\Omega^{-1}Q$, where $VV'=\frac{1}{n}\sum_i^n\sum_s^n x_i\varepsilon_i\varepsilon_s x_s'$. Since the data is *i.i.d.*, we have $E(x_i\varepsilon_i\varepsilon_s x_s')=0$ for all $i\ne s$. Hence, $E(VV')=\frac{1}{n}\sum_i^n E(x_ix_i'\varepsilon_i^2)=\Omega$ and $Ed_0d_0'=Q'\Omega^{-1}Q=D$.

(d_1) Let d_1 be defined as in (A.14) and write $d_{11}=Q'(\Omega^{*-1}_M-\Omega^{-1})V$ and $d_{12}=Q'_M\Omega^{-1}_M V_M-Q'\Omega^{*-1}_M V$ where Ω^*_M is defined in (B.1). Then,

$$E\|d_{12}\| \le \sigma^{-2}\sum_{j=M+1}^{\infty}\|\Gamma_j\|\,E\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_i^l\varepsilon_i\right| \le \sigma^{-2}\sum_{j=M+1}^{\infty}\|\Gamma_j\|\underbrace{\left(E\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_i^l\varepsilon_i\right|^2\right)^{1/2}}_{O(1)} \le \sigma^{-2}\sum_{j=M+1}^{\infty}\|\Gamma_j\|=M^{-\alpha}\frac{\kappa}{\sigma^2\alpha}+o(M^{-\alpha})$$

by Lemma C.7. Next use that $(\Omega^{*-1}_M-\Omega^{-1})=\Omega^{*-1}_M(\Omega-\Omega^*_M)\Omega^{-1}$ and define $a_k=\sum_{j=1}^{M}\Gamma_j\vartheta_{jk}$. By Lemma C.9 this term decays as $\|\Gamma_k\|$. Then,

$$E\|d_{11}\| \le \sum_{j_1,j_2,j_3,j_4=1}^{\infty}\|\Gamma_{j_1}\|\left|\vartheta^*_{j_1j_2}\right|\left|\omega_{j_2j_3}-\omega^*_{j_2j_3}\right|\left|\vartheta_{j_3j_4}\right| \times E\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_i^l\varepsilon_i\right|$$

$$\le \sum_{j_3=M+1}^{\infty}\|a^*_{j_3}\|\left|\omega_{j_3j_3}-\sigma^2\right|\sum_{j_4=1}^{\infty}\left|\vartheta_{j_3j_4}\right| +2\sum_{j_3=M+1}^{\infty}\sum_{j_2=1}^{\infty}\|a^*_{j_2}\|\left|\omega_{j_2j_3}\right|\sum_{j_4=1}^{\infty}\left|\vartheta_{j_3j_4}\right| \le M^{-\alpha}\times C_1$$

since the sums go to zero at rate $M^{-\alpha}$ by Lemma C.7. Thus, $\|d_1\|=O_p(M^{-\alpha})$.

(d_2) Let d_2 be defined as in (A.15). I use $\bar{k}_\alpha$ as defined in (B.2) and the MI.

$$d_2=\sum_{j=1}^{M}\sum_{l=1}^{M}\Gamma_j\,(1-k(j/M))\,\vartheta^M_{jl}\,(1-k(l/M))\,\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_i^l\varepsilon_i$$

$$E\|d_2\|\le M^{-2\alpha}\sum_{j=cM+1}^{M}\sum_{l=cM+1}^{M}j^{2\alpha}\,\|\Gamma_j\|\left|\bar{k}_\alpha(j/M)\right|\left|\vartheta^M_{jl}\right|\left|\bar{k}_\alpha(l/M)\right| \times\underbrace{\left(E\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_i^l\varepsilon_i\right|^2\right)^{1/2}}_{O(1)}\le C_1\times M^{-2\alpha-1}M \times\sum_{l=cM+1}^{M}\sum_{j=cM+1}^{M}j^{2\alpha}\,\|\Gamma_j\|\left|\bar{k}_\alpha(j/M)\right|\left|\vartheta^M_{jl}\right|\left|\bar{k}_\alpha(l/M)\right| =O(M^{-(2\alpha+1)})$$

since $\lim M\sum_{l=cM+1}^{M}\sum_{j=cM+1}^{M}j^{2\alpha}\|\Gamma_j\|\left|\bar{k}_\alpha(j/M)\right|\left|\vartheta^M_{jl}\right|\left|\bar{k}_\alpha(l/M)\right|<\infty$ by Lemma C.8.

(d_3) Let d_3 be defined as in (A.16). Since d_3 has two terms we can write $d_3=d_{31}+d_{32}$ and look at d_{31}. Then,

$$d_{31}=-\sum_{j=cM+1}^{M}\sum_{l=1}^{M}\Gamma_j(1-k(j/M))\vartheta^M_{jl}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_i^l\varepsilon_i$$

$$E\|d_{31}\|\le M^{-\alpha}\sum_{l=1}^{M}\sum_{j=cM+1}^{M}j^{\alpha}\,\|\Gamma_j\|\left|\bar{k}_\alpha(j/M)\right|\left|\vartheta^M_{jl}\right| \times\underbrace{\left(E\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_i^l\varepsilon_i\right|^2\right)^{1/2}}_{O(1)} \le C_1\times M^{-\alpha}\sum_{l=1}^{M}\sum_{j=cM+1}^{M}j^{\alpha}\,\|\Gamma_j\|\left|\bar{k}_\alpha(j/M)\right|\left|\vartheta^M_{jl}\right|=O(M^{-\alpha})$$

since $\sum_{l=1}^{M}\sum_{j=cM+1}^{M}j^{\alpha}\,\|\Gamma_j\|\left|\bar{k}_\alpha(j/M)\right|\left|\vartheta^M_{jl}\right|<\infty$ by Lemma C.6. The case of d_{32} is identical.

(d_4) Let d_4 be defined as in (A.17). Using Lemma C.3, the L^r inequality and Jensen's inequality,

$$\sum_{j=1}^{M}\sum_{l=1}^{M}E(\sqrt{n}\left|\widehat{\omega}_{jl}-\omega_{jl}\right|)\le E\sqrt{n}\left\|\widehat{\Omega}_M-\Omega_M\right\| \le\left(\sum_{j=1}^{M}\sum_{l=1}^{M}nE\left|\widehat{\omega}_{jl}-\omega_{jl}\right|^2\right)^{1/2}=O(M), \tag{B.3}$$

meaning that $E(\sqrt{n}\left|\widehat{\omega}_{jl}-\omega_{jl}\right|)$ is summable in one of the indices. Define $a^M_j=\sum_l^M\Gamma_l k(l/M)\vartheta^M_{lj}$ as in Lemma B.3 and recall that a^M_j is absolutely summable. Then, write $\|d_4\|$ as,

$$\|d_4\|\le\sum_{j_1,j_2,j_3,j_4}^{M}\left\|\Gamma_{j_1}k(j_1/M)\vartheta^M_{j_1j_2}(\widehat{\omega}_{j_2j_3}-\omega_{j_2j_3})\vartheta^M_{j_3j_4}k(j_4/M)\right.$$

$$\times \frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i^{j_4}\varepsilon_i \Bigg\|$$

$$\leq \sum_{j_2,j_3,j_4}^{M} \|a_{j_2}\| \left|\widehat{\omega}_{j_2j_3}-\omega_{j_2j_3}\right| \left|\vartheta^M_{j_3j_4}\right| \underbrace{\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i^{j_4}\varepsilon_i\right|}_{O_p(1)}$$

$$\leq \frac{O_p(1)}{\sqrt{n}} \times \sum_{j_2}^{M} \|a_{j_2}\| \sum_{j_3,j_4}^{M} \sqrt{n}\left|\widehat{\omega}_{j_2j_3}-\omega_{j_2j_3}\right| \left|\vartheta^M_{j_3j_4}\right|$$

$$\overset{(1)}{=} O_p(1/\sqrt{n})$$

were $\overset{(1)}{=}$ follows from $\sum_{j_3,j_4}^{M} \sqrt{n}\left|\widehat{\omega}_{j_2j_3}-\omega_{j_2j_3}\right| \left|\vartheta^M_{j_3j_4}\right| = O(1)$ uniformly in j_2 by (B.3) and the fact that $\|a_{j_2}\|$ is summable.

(d_5) Let d_5 be defined as in (A.18). I use again that $E|\frac{1}{\sqrt{n}}\sum x_i^l\varepsilon_i|^2 = O(1)$. The result follows after using the CSI and the MI.

$$E\|d_5\| \leq \sum_{j=1}^{M}\sum_{l=1}^{M} E\left\|(\widehat{\Gamma}_j-\Gamma_j)k(j/M)\vartheta^M_{jl}k(l/M)\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i^l\varepsilon_i\right\|$$

$$\leq \sum_{j=1}^{M}\sum_{l=1}^{M}\left(E\left\|\widehat{\Gamma}_j-\Gamma_j\right\|^2\right)^{1/2}\left|\vartheta^M_{jl}\right|\left(E\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i^l\varepsilon_i\right|^2\right)^{1/2}$$

$$\leq C_1 n^{-1/2}\sum_{j=1}^{M}\sum_{l=1}^{M}\left|\vartheta^M_{jl}\right| = O(M/n^{1/2}).$$

(d_6) The fact that $\|d_6\| = O_p(M/n)$ as in (A.19) follows directly from (B.3), $\|\widehat{\Gamma}_{j_1}-\Gamma_{j_1}\| = O_p(n^{-1/2})$ and $\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i^{j_4}\varepsilon_i\right| = O_p(1)$ uniformly in j_4.

$$\|d_6\| \leq \sum_{j_1,j_2,j_3,j_4}^{M} E\Bigg\|(\widehat{\Gamma}_{j_1}-\Gamma_{j_1})k(j_1/M)\vartheta^M_{j_1j_2}$$

$$\times(\widehat{\omega}_{j_2j_3}-\omega_{j_2j_3})\vartheta^M_{j_3j_4}k(j_4/M)\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i^{j_4}\varepsilon_i\Bigg\|$$

$$\leq \sum_{j_1,j_2,j_3,j_4}^{M}\left\|\widehat{\Gamma}_{j_1}-\Gamma_{j_1}\right\|\left|\vartheta^M_{j_1j_2}\right|\left|\widehat{\omega}_{j_2j_3}-\omega_{j_2j_3}\right|\left|\vartheta^M_{j_3j_4}\right|$$

$$\times\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i^{j_4}\varepsilon_i\right|$$

$$\leq \frac{O_p(n^{-1/2})}{\sqrt{n}}\times\sum_{j_1,j_2}^{M}\left|\vartheta^M_{j_1j_2}\right|\sum_{j_3,j_4}^{M}\sqrt{n}\left|\widehat{\omega}_{j_2j_3}-\omega_{j_2j_3}\right|\left|\vartheta^M_{j_3j_4}\right|$$

$$\overset{(1)}{=} O_p(M/n)$$

were $\overset{(1)}{=}$ follows from $\sum_{j_3,j_4}^{M} \sqrt{n}\left|\widehat{\omega}_{j_2j_3}-\omega_{j_2j_3}\right| \left|\vartheta^M_{j_3j_4}\right| = O(1)$ uniformly in j_2 by (B.3) and the fact that $\sum_{j_1,j_2}^{M}\left|\vartheta^M_{j_1j_2}\right| = O(M)$.

(d_7) Define d_7 as in (A.20). Then, $d_7 = o_p(M/\sqrt{n})$ since $B = o_p(\|\widehat{\Omega}_M-\Omega_M\|) = o_p(M/\sqrt{n})$. ■

Lemma B.5. *Suppose Assumptions 3.1 to 3.5 hold and let d_3 be defined as in (A.16). Then, $M^{2\alpha+1}E(d_3d_3')$ can be approximated by $F_{\alpha,\sigma,\kappa}(c)\mathbf{C}_2$ where $\mathbf{C}_2$ is some matrix with finite norm and $F_{\alpha,\sigma,\kappa}(c) = \frac{\kappa^2c^2}{\sigma^2(1-c)^2}\left[\frac{1}{c^{2\alpha+1}(4\alpha^2-1)\alpha}+\frac{c(2\alpha-1)-\alpha}{c^2\alpha(2\alpha-1)}-\frac{1}{2\alpha+1}\right]$.*

Proof. Write $d_3 = d_{31}+d_{32}$ so that $d_3d_3' = d_{31}d_{31}'+d_{32}d_{32}'+d_{31}d_{32}'+d_{32}d_{31}'$. The first term is

$$E(d_{31}d_{31}') = Q_M'(I-\mathcal{K}_M)\Omega_M^{-1}E(V_MV_M')\Omega_M^{-1}(I-\mathcal{K}_M)Q_M$$
$$= Q_M'(I-\mathcal{K}_M)\Omega_M^{-1}(I-\mathcal{K}_M)Q_M = H_{12} = O(M^{-(2\alpha+1)}).$$

The second term includes $a_k = \sum_{j=1}^{M}\Gamma_j\vartheta_{jk}$. By Lemma C.9 $\|a_k\|$ decays as $\|\Gamma_k\|$ so that by Assumption 3.5 we have:

$$E(d_{32}d_{32}') = Q_M'\Omega_M^{-1}(I-\mathcal{K}_M)E(V_MV_M')(I-\mathcal{K}_M)\Omega_M^{-1}Q_M$$
$$= Q_M'\Omega_M^{-1}(I-\mathcal{K}_M)\Omega_M(I-\mathcal{K}_M)\Omega_M^{-1}Q_M$$
$$= M^{-2\alpha}\sum_{j_2,j_3=cM+1}^{M}\sum_{j_1,j_4}^{M} j_2^{\alpha}j_3^{\alpha}\Gamma_{j_1}\vartheta^M_{j_1,j_2}\bar{k}_\alpha\left(\frac{j_2}{M}\right)\omega_{j_2j_3}\bar{k}_\alpha$$
$$\times\left(\frac{j_3}{M}\right)\vartheta^M_{j_3j_4}\Gamma_{j_4}'$$
$$= M^{-2\alpha}\sum_{j_2,j_3=cM+1}^{M} a_{j_2}j_2^{\alpha}\bar{k}_\alpha\left(\frac{j_2}{M}\right)\omega_{j_2j_3}\bar{k}_\alpha\left(\frac{j_3}{M}\right)j_3^{\alpha}a_{j_3}'$$
$$= M^{-2\alpha-1}M\sum_{j_2,j_3=cM+1}^{M} a_{j_2}j_2^{\alpha}\bar{k}_\alpha\left(\frac{j_2}{M}\right)\omega_{j_2j_3}\bar{k}_\alpha\left(\frac{j_3}{M}\right)j_3^{\alpha}a_{j_3}'$$
$$= O(M^{-(2\alpha+1)})$$

since $\lim M\sum_{j_2,j_3=cM+1}^{M} a_{j_2}j_2^{\alpha}\bar{k}_\alpha\left(\frac{j_2}{M}\right)\omega_{j_2j_3}\bar{k}_\alpha\left(\frac{j_3}{M}\right)j_3^{\alpha}a_{j_3}' < \infty$. The next term is

$$E(d_{31}d_{32}') = Q_M'(I-\mathcal{K}_M)\Omega_M^{-1}E(V_MV_M')(I-\mathcal{K}_M)\Omega_M^{-1}Q_M$$
$$= Q_M'(I-\mathcal{K}_M)^2\Omega_M^{-1}Q_M$$
$$= \sum_{j=cM+1}^{M}\Gamma_j\left(\frac{j}{cM}-1\right)\left(\frac{j}{cM}-1\right)\sum_{l=1}^{M}\vartheta^M_{jl}\Gamma_l'$$
$$= M^{-2\alpha}\sum_{j=cM+1}^{M} j^{\alpha}\Gamma_j\bar{k}_\alpha(j/M)\bar{k}_\alpha(j/M)j^{\alpha}a_j'$$
$$= M^{-2\alpha-1}M\sum_{j=cM+1}^{M} j^{\alpha}\Gamma_j\bar{k}_\alpha(j/M)\bar{k}_\alpha(j/M)j^{\alpha}a_j'$$
$$= O(M^{-(2\alpha+1)})$$

since $\lim M\sum_{j=cM+1}^{M} j^{\alpha}\Gamma_j\bar{k}_\alpha(\frac{j}{M})\bar{k}_\alpha(\frac{j}{M})j^{\alpha}a_j < \infty$ by Lemmas C.8 and C.9.

Finally, Lemma C.8 shows that each of these sums is proportional to $F_{\alpha,\sigma,\kappa}(c)$ and have finite norm. Thus, I can write these sums as $F_{\alpha,\sigma,\kappa}(c)$ times a matrix with finite norm that does not depend on c. ■

Lemma B.6. *Suppose Assumptions 3.1 to 3.5 hold and let d_5 be defined as in (A.18). Then, $\ell'E(d_5d_5')\ell = \frac{M^2}{n}\left(\int k^2(x)dx\right)^2C_1 + o(M^2/n)$.*

Proof. Note that $d_5d_5' = (\widehat{Q}_M-Q_M)'\mathcal{K}_M\Omega_M^{-1}\mathcal{K}_MV_MV_M'\mathcal{K}_M\Omega_M^{-1}\mathcal{K}_M(\widehat{Q}_M-Q_M)$. Also define γ_i^j as $\gamma_i^j = x_i^jW_i'-E(x_i^jW_i')$ and $v_i^j = x_i^j\varepsilon_i$. Then I can write $\widehat{\Gamma}_j-\Gamma_j = n^{-1}\sum_{i=1}^{n}\gamma_i^j$. Notice that γ_i^j is a $p\times 1$ vector with typical element $\gamma_i^{j,m} = x_i^jw_i^m - E(x_i^jw_i^m)$. Now write $\ell'd_5d_5'\ell$ as:

$$\ell'd_5d_5'\ell = \frac{1}{n^3}\sum_{i_1,i_2,i_3,i_4}^{n}\sum_{j_1,j_2,j_3,j_4}^{M}\ell'\gamma_{i_1}^{j_1}k\left(\frac{j_1}{M}\right)\vartheta^M_{j_1j_2}k$$
$$\times\left(\frac{j_2}{M}\right)v_{i_2}^{j_2}v_{i_3}^{j_3}k\left(\frac{j_3}{M}\right)\vartheta^M_{j_3j_4}k\left(\frac{j_4}{M}\right)\gamma_{i_4}^{j_4\prime}\ell$$
$$= \frac{1}{n^3}\sum_{i_1,i_2,i_3,i_4}^{n}\sum_{j_1,j_2,j_3,j_4}^{M}\prod_{l=1}^{4}k\left(\frac{j_l}{M}\right)\vartheta^M_{j_1j_2}\vartheta^M_{j_3j_4}\ell'\gamma_{i_1}^{j_1}v_{i_2}^{j_2}v_{i_3}^{j_3}\gamma_{i_4}^{j_4\prime}\ell.$$

Since I have four zero mean scalar random variables I can use the relationship between moments and cumulants Barndorff-Nielsen and Cox (1989, Page 144) to write,

$$E(\ell'\gamma_{i_1}^{j_1}\upsilon_{i_2}^{j_2}\upsilon_{i_3}^{j_3}\gamma_{i_4}^{j_4\prime}\ell) = E(\ell'\gamma_{i_1}^{j_1}\upsilon_{i_2}^{j_2})E(\upsilon_{i_3}^{j_3},\gamma_{i_4}^{j_4\prime}\ell) + E(\ell'\gamma_{i_1}^{j_1}\upsilon_{i_3}^{j_3}) \times E(\upsilon_{i_2}^{j_2\prime}\gamma_{i_4}^{j_4\prime}\ell) + E(\ell'\gamma_{i_1}^{j_1}\gamma_{i_4}^{j_4\prime}\ell)E(\upsilon_{i_2}^{j_2}\upsilon_{i_3}^{j_3}) + cum(\ell'\gamma_{i_1}^{j_1},\upsilon_{i_2}^{j_2},\upsilon_{i_3}^{j_3},\gamma_{i_4}^{j_4\prime}\ell).$$

So that

$$E(\ell' d_5 d_5'\ell) = \frac{1}{n^3}\sum_{i_1,i_2,i_3,i_4}^{n}\sum_{j_1,j_2,j_3,j_4}^{M}\prod_{l=1}^{4}k\left(\frac{j_l}{M}\right)\vartheta_{j_1j_2}^{M}\vartheta_{j_3j_4}^{M} \times\Big\{E(\ell'\gamma_{i_1}^{j_1}\upsilon_{i_2}^{j_2})E(\upsilon_{i_3}^{j_3},\gamma_{i_4}^{j_4\prime}\ell) + E(\ell'\gamma_{i_1}^{j_1}\upsilon_{i_3}^{j_3})E(\upsilon_{i_2}^{j_2\prime}\gamma_{i_4}^{j_4\prime}\ell) + E(\ell'\gamma_{i_1}^{j_1}\gamma_{i_4}^{j_4\prime}\ell)E(\upsilon_{i_2}^{j_2}\upsilon_{i_3}^{j_3}) + cum(\ell'\gamma_{i_1}^{j_1},\upsilon_{i_2}^{j_2},\upsilon_{i_3}^{j_3},\gamma_{i_4}^{j_4\prime}\ell)\Big\} = (A)+(B)+(D)+(G).$$

Notice that $E(\ell'\gamma_{i_1}^{j_1}\upsilon_{i_2}^{j_2})$ will be zero for $i_1\neq i_2$ and then

$$E(\ell'\gamma_i^{j_1}\upsilon_i^{j_2}) = \sum_{m=1}^{p}E\left[\ell_m\left(x_i^{j_1}w_i^m - E(x_i^{j_1}w_i^m)\right)x_i^{j_2}\varepsilon_i\right] = \sum_{m=1}^{p}\ell_m E\left(x_i^{j_1}x_i^{j_2}w_i^m\varepsilon_i\right) = \sum_{m=1}^{p}\ell_m E\left(x_i^{j_1}x_i^{j_2}\underbrace{E\left(w_i^m\varepsilon_i|x_i\right)}_{\sigma_{\varepsilon u}^m}\right) = \ell'\sigma_{u\varepsilon}E\left(x_i^{j_1}x_i^{j_2}\right)$$

where for simplicity I use that $E\left(w_i^m\varepsilon_i|x_i\right)=\sigma_{u\varepsilon}^m$. Now focus on the first term.

$$\frac{n}{M^2}(A) = \frac{1}{M^2n^2}\sum_{i_1,i_2,i_3,i_4}^{n}\sum_{j_1,j_2,j_3,j_4}^{M}\prod_{l=1}^{4}k\left(\frac{j_l}{M}\right)\vartheta_{j_1j_2}^{M}\vartheta_{j_3j_4}^{M} \times E(\ell'\gamma_{i_1}^{j_1}\upsilon_{i_2}^{j_2})E(\upsilon_{i_3}^{j_3},\gamma_{i_4}^{j_4\prime}\ell)$$
$$= \frac{1}{M^2}\sum_{j_1,j_2,j_3,j_4}^{M}\prod_{l=1}^{4}k\left(\frac{j_l}{M}\right)\vartheta_{j_1j_2}^{M}\vartheta_{j_3j_4}^{M}\left(\ell'\sigma_{u\varepsilon}\right)^2E\left(x_i^{j_1}x_i^{j_2}\right)E\left(x_i^{j_3}x_i^{j_4}\right)$$
$$= \left(\ell'\sigma_{u\varepsilon}\right)^2\frac{1}{M}\sum_{j_1,j_2}^{M}k\left(\frac{j_1}{M}\right)k\left(\frac{j_2}{M}\right)\vartheta_{j_1j_2}^{M}E\left(x_i^{j_1}x_i^{j_2}\right) \times\frac{1}{M}\sum_{j_3,j_4}^{M}k\left(\frac{j_3}{M}\right)k\left(\frac{j_4}{M}\right)\vartheta_{j_3j_4}^{M}E\left(x_i^{j_3}x_i^{j_4}\right)$$
$$= \left(\ell'\sigma_{u\varepsilon}\right)^2\left(\int_0^1k^2(x)\mathrm{d}x\right)^2C_1^2+o(1) \tag{B.4}$$

where I am using that

$$\frac{1}{M}\sum_{l,j}^{M}k\left(\frac{l}{M}\right)k\left(\frac{j}{M}\right)\vartheta_{lj}^{M}E\left(x_i^lx_i^j\right) = \frac{1}{M}\sum_{l}^{M}k^2\left(l/M\right)\vartheta_{ll}^{M}E(x_i^l)^2 + \frac{1}{M}\sum_{\substack{l=1,\\ j\neq l}}^{M}k\left(\frac{l}{M}\right)k\left(\frac{j}{M}\right)\vartheta_{lj}^{M}E\left(x_i^lx_i^j\right) = \left(\int_0^1k^2(x)\mathrm{d}x\right)C_1+o(1)<\infty$$

since the second term is $o(1)$ by the summability of $E\left(x_i^lx_i^j\right)$ and the first term converge to the roughness of the kernel times a constant based on arguments similar to those in Kuersteiner (2002, page 55) and Parzen (1957, page 342).

The next two terms, (B) and (D), are $o(M^2/n)$. To see this note that,

$$\|(B)\| \le \left\|\frac{1}{n^3}\sum_{i_1,i_2,i_3,i_4}^{n}\sum_{j_1,j_2,j_3,j_4}^{M}\prod_{l=1}^{4}k\left(\frac{j_l}{M}\right)\vartheta_{j_1j_2}^{M}\vartheta_{j_3j_4}^{M}\times E(\ell'\gamma_{i_1}^{j_1}\upsilon_{i_3}^{j_3})E(\upsilon_{i_2}^{j_2\prime}\gamma_{i_4}^{j_4\prime}\ell)\right\|$$
$$\le \frac{1}{n}\times\left(\ell'\sigma_{\varepsilon u}\right)^2\sum_{j_1,j_2}^{M}k\left(\frac{j_1}{M}\right)k\left(\frac{j_2}{M}\right)\vartheta_{j_1j_2}^{M} \times\left\|\sum_{j_3,j_4}^{M}k\left(\frac{j_3}{M}\right)k\left(\frac{j_4}{M}\right)\vartheta_{j_3j_4}^{M}E\left(x_i^{j_1}x_i^{j_3}\right)E\left(x_i^{j_2}x_i^{j_4}\right)\right\| \le O\left(\frac{M}{n}\right)$$

and

$$\|(D)\| \le \left\|\frac{1}{n^3}\sum_{i_1,i_2,i_3,i_4}^{n}\sum_{j_1,j_2,j_3,j_4}^{M}\prod_{l=1}^{4}k\left(\frac{j_l}{M}\right)\vartheta_{j_1j_2}^{M}\vartheta_{j_3j_4}^{M}\times E(\ell'\gamma_{i_1}^{j_1}\gamma_{i_4}^{j_4\prime}\ell)E(\upsilon_{i_2}^{j_2}\upsilon_{i_3}^{j_3})\right\|$$
$$\le \left\|\frac{1}{n}\sum_{j_1,j_4}^{M}E(\ell'\gamma_{i_1}^{j_1}\gamma_{i_4}^{j_4\prime}\ell)\sum_{j_2,j_3}^{M}\vartheta_{j_1j_2}^{M}\omega_{j_2j_3}\vartheta_{j_3j_4}^{M}\right\| \le \left\|\frac{1}{n}\sum_{j_1,j_4}^{M}E(\ell'\gamma_{i_1}^{j_1}\gamma_{i_4}^{j_4\prime}\ell)\vartheta_{j_1j_4}^{M}\right\| = O\left(\frac{M}{n}\right)$$

where I used that $\sum_{j_2,j_3}^{M}\vartheta_{j_1j_2}^{M}\omega_{j_2j_3}\vartheta_{j_3j_4}^{M}=\vartheta_{j_1j_4}^{M}$. Finally, (G) is $o(M^2/n)$ due to Assumption 3.1. ■

Lemma B.7. *Suppose Assumptions 3.1 to 3.5 hold and let d_0, d_3, H_{13}, H_{14} and D be defined as in Appendix A. Then, $M^{\alpha+1}E(d_0d_3')=M^{\alpha+1}E(d_0d_0')D^{-1}(H_{13}+H_{14})+o(M^{-\alpha})$ which means that these terms cancel out in the expression for the AMSE.*

Proof. I use $\bar{k}_\alpha$ as defined in (B.2). Write $d_3=d_{31}+d_{32}$ and then look at Ed_0d_{31}' and Ed_0d_{32}'.

$$M^{\alpha+1}E(d_0d_{31}') = M^{\alpha+1}E(-Q'\Omega^{-1}VV_M'\Omega_M^{-1}(I-\mathcal{K}_M)Q_M)$$
$$= -M^{\alpha+1}\sum_{j_1,j_2=1}^{\infty}\sum_{j_3,j_4=1}^{M}\Gamma_{j_1}\vartheta_{j_1j_2}E(x_i^{j_2}x_i^{j_3}\varepsilon_i^2)\vartheta_{j_3j_4}^{M}\left(1-k(j_4/M)\right)\Gamma_{j_4}'$$
$$= -M^{\alpha+1}\sum_{j_1,j_2=1}^{\infty}\sum_{j_3,j_4=1}^{M}\Gamma_{j_1}\vartheta_{j_1j_2}\omega_{j_2j_3}\vartheta_{j_3j_4}^{M}\left(1-k(j_4/M)\right)\Gamma_{j_4}'$$
$$= -M^{\alpha+1}\sum_{j_1=1}^{\infty}\sum_{j_4=cM+1}^{M}\Gamma_{j_1}\vartheta_{j_1j_4}^{M}\left(1-k(j_4/M)\right)\Gamma_{j_4}' = (A)+(B)$$

where

$$(A) = -M\sum_{j_1=1}^{M}\sum_{j_4=cM+1}^{M}\Gamma_{j_1}\vartheta_{j_1j_4}^{M}\bar{k}_\alpha(j_4/M)j_4^\alpha\Gamma_{j_4}' = M^{\alpha+1}H_{13}$$

$$(B) = -M\sum_{j_1=M+1}^{\infty}\sum_{j_4=cM+1}^{M}\Gamma_{j_1}\vartheta_{j_1j_4}^{M}\bar{k}_\alpha(j_4/M)j_4^\alpha\Gamma_{j_4}' = o(M^{-\alpha}).$$

The last equality follows by similar arguments to those in (C.3). Since for all $j_1>j_4$ we have $|\vartheta_{j_1j_4}^{M}|\le C/j_4^{1+\epsilon}$ for some C and $\epsilon>0$, it follows that

$$M\sum_{j_4=cM+1}^{M}\bar{k}_\alpha(j_4/M)j_4^\alpha\Gamma_{j_4}'\times\frac{C}{j_4^{1+\epsilon}}\to 0, \tag{B.5}$$

as $M \to \infty$. Using this into the expression for (B) we get,

$$|(B)| = \left| M \sum_{j_1=M+1}^{\infty} \sum_{j_4=cM+1}^{M} \Gamma_{j_1} \vartheta_{j_1 j_4}^{M} \bar{k}_\alpha(j_4/M) j_4^\alpha \Gamma_{j_4}' \right| \le M^{-\alpha} \times M^\alpha \sum_{j_1=M+1}^{\infty} \|\Gamma_{j_1}\| \times \underbrace{M \sum_{j_4=cM+1}^{M} \bar{k}_\alpha(j_4/M) j_4^\alpha \|\Gamma_{j_4}'\| |\vartheta_{j_1 j_4}^{M}|}_{=o(1) \text{ by (B.5)}} \le M^{-\alpha} \times \underbrace{M^\alpha \sum_{j_1=M+1}^{\infty} \|\Gamma_{j_1}\|}_{=o(1) \text{ by Lemma C.7}} \times o(1) = o(M^{-\alpha}).$$

The same argument is valid for $Ed_0 d_{32}'$ and $Ed_0 d_0' D^{-1} H_{14}$. ■

Lemma B.8. *Let Assumptions 3.1 to 3.5 hold and let d_0, d_2, H_{12}, and D be defined as in Appendix A. Then, $E(d_0 d_2') = E(d_0 d_0') D^{-1} H_{12} + o(M^{-(2\alpha+1)})$ which means that these terms cancel out.*

Proof. Directly evaluate $Ed_0 d_2'$

$$E(d_0 d_2') = E(Q' \Omega^{-1} V V_M' (I - \mathcal{K}_M) \Omega_M^{-1} (I - \mathcal{K}_M) Q_M) = \sum_{j_1, j_2=1}^{\infty} \sum_{j_3, j_4=1}^{M} \Gamma_{j_1} \vartheta_{j_1 j_2} E(x_i^{j_2} x_i^{j_3} \varepsilon_i^2) \times (1 - k(j_3/M)) \vartheta_{j_3 j_4}^{M} (1 - k(j_4/M)) \Gamma_{j_4}' = \sum_{j_1, j_2=1}^{\infty} \sum_{j_3, j_4=1}^{M} \Gamma_{j_1} \vartheta_{j_1 j_2} \omega_{j_2 j_3} (1 - k(j_3/M)) \vartheta_{j_3 j_4}^{M} (1 - k(j_4/M)) \Gamma_{j_4}'.$$

Note that $\sum_{j_2=1}^{\infty} \vartheta_{j_1 j_2} \omega_{j_2 j_3} = 1$ if $j_1 = j_3$ and 0 otherwise. Then, since $E(d_0 d_0') D^{-1} = I$,

$$M^{2\alpha+1} E(d_0 d_2') = M \sum_{j_3, j_4=cM+1}^{M} j_3^\alpha \Gamma_{j_3} \bar{k}_\alpha(j_3/M) \vartheta_{j_3 j_4}^{M} \bar{k}_\alpha(j_4/M) j_4^\alpha \Gamma_{j_4}' + o(1) = M^{2\alpha+1} E(d_0 d_0') D^{-1} H_{12} + o(1). \quad ■$$

Lemma B.9. *Suppose Assumptions 3.1 to 3.5 hold and let d_1, d_0, H_{11} and D be defined as in Appendix A. Then $E(d_1 d_0') = E(d_0 d_0') D^{-1} H_{11}$*

Proof. Recall that $d_1 = d_{11} + d_{12}$, where $d_{11} = Q'(\Omega_M^{*-1} - \Omega^{-1}) V$ and $d_{12} = Q_M' \Omega_M^{-1} V_M - Q' \Omega_M^{*-1} V$. Then,

$$E(d_{11} d_0') = Q'(\Omega_M^{*-1} - \Omega^{-1}) E(VV') \Omega^{-1} Q = Q'(\Omega_M^{*-1} - \Omega^{-1}) Q = H_{111}.$$

On the other hand, note that $Q' \Omega_M^{*-1} V$ can be written as:

$$\left[Q_M' \mid Q_{M+1}'^{\infty} \right] \begin{bmatrix} \Omega_M^{-1} & 0 \\ 0 & \sigma^{-2} I_\infty \end{bmatrix} \begin{bmatrix} V_M \\ \text{---} \\ V_{M+1}^{\infty} \end{bmatrix} = Q_M' \Omega_M^{-1} V_M + \sigma^{-2} Q_{M+1}'^{\infty} V_{M+1}^{\infty}$$

which means that $d_{12} = -\sigma^{-2} \widetilde{Q}' \widetilde{V}$ where $\widetilde{Q}' = \left[\mathbf{0}_M' \mid Q_{M+1}'^{\infty} \right]$ and $\widetilde{V}' = \left[\mathbf{0}_M' \mid V_{M+1}^{\infty} \right]$. Next, I compute $Ed_{12} d_0' = -\sigma^{-2} \widetilde{Q}' E(\widetilde{V} V') \Omega^{-1} Q$ and note that $E(\widetilde{V} V') = \widetilde{\Omega}$ is just the matrix Ω except that the first $M \times M$ block is replaced by zeros. Denote by $\widetilde{\omega}_{jl}$ an element of $\widetilde{\Omega}$ and note that for $j > M$, $\widetilde{\omega}_{jl}$ is just ω_{jl}. Using all this we can write

$$E(d_{12} d_0') = -\sigma^{-2} \widetilde{Q}' E(\widetilde{V} V') \Omega^{-1} Q = -\sigma^{-2} \sum_{j_1=1}^{\infty} \sum_{j_2, j_3=1}^{\infty} \widetilde{\Gamma}_{j_1} \widetilde{\omega}_{j_1 j_2} \vartheta_{j_2 j_3} \Gamma_{j_3}' = -\sigma^{-2} \sum_{j_1=1}^{M} \sum_{j_2, j_3=1}^{\infty} \widetilde{\Gamma}_{j_1} \widetilde{\omega}_{j_1 j_2} \vartheta_{j_2 j_3} \Gamma_{j_3}' - \sigma^{-2} \times \sum_{j_1=M+1}^{\infty} \sum_{j_2, j_3=1}^{\infty} \widetilde{\Gamma}_{j_1} \omega_{j_1 j_2} \vartheta_{j_2 j_3} \Gamma_{j_3}'.$$

The first term is zero since $\widetilde{\Gamma}_{j_1} = 0$ for $j_1 \le M$. For the second term, I use that $\sum_{j_2}^{\infty} \omega_{j_1 j_2} \vartheta_{j_2 j_3} = 1$ if $j_1 = j_3$ and is zero otherwise. As a consequence,

$$E(d_{12} d_0') = -\sigma^{-2} \sum_{j_1=M+1}^{\infty} \Gamma_{j_1} \Gamma_{j_1}' = H_{112}.$$

The fact that $E(d_0 d_0') = D$ completes the proof. ■

Lemma B.10. *Suppose Assumptions 3.1 to 3.5 hold and let d_1 be defined as in (A.14). Then, $M^{2\alpha+1} Ed_1 d_1'$ can be approximated by $\mathbf{C}_3$ where $\mathbf{C}_3$ is some matrix with finite norm.*

Proof. As in the previous Lemma I use $d_1 = d_{11} + d_{12}$, where $d_{11} = Q'(\Omega_M^{*-1} - \Omega^{-1}) V$ and $d_{12} = Q_M' \Omega_M^{-1} V_M - Q' \Omega_M^{*-1} V = -\sigma^{-2} \widetilde{Q}' \widetilde{V}$ where $\widetilde{Q}' = \left[\mathbf{0}_M' \mid Q_{M+1}'^{\infty} \right]$ and $\widetilde{V}' = \left[\mathbf{0}_M' \mid V_{M+1}^{\infty} \right]$. Then, $d_1 d_1' = d_{11} d_{11}' + d_{12} d_{12}' + d_{11} d_{12}' + (d_{11} d_{12}')'$. By Looking at each of these terms I get,

$$E(d_{12} d_{12}') = \sigma^{-4} \widetilde{Q}' E(\widetilde{V} \widetilde{V}') \widetilde{Q} = \sigma^{-4} \widetilde{Q}' \widetilde{\Omega} \widetilde{Q} = \sigma^{-4} \times \sum_{j,l=M+1}^{\infty} \Gamma_j \omega_{jl} \Gamma_l' = O(M^{2\alpha+1})$$

where the last step uses Lemma C.7. Moreover,

$$E(d_{11} d_{11}') = Q'(\Omega_M^{*-1} - \Omega^{-1}) E(VV') (\Omega_M^{*-1} - \Omega^{-1}) Q = Q'(\Omega_M^{*-1} - \Omega^{-1}) \Omega (\Omega_M^{*-1} - \Omega^{-1}) Q = Q' \Omega_M^{*-1} (\Omega - \Omega_M^{*}) \Omega^{-1} \Omega \Omega_M^{*-1} (\Omega - \Omega_M^{*}) \Omega^{-1} Q = Q' \Omega_M^{*-1} (\Omega - \Omega_M^{*}) \Omega_M^{*-1} (\Omega - \Omega_M^{*}) \Omega^{-1} Q$$

which is $O(M^{-(2\alpha+1)})$ by using the same arguments as those in Lemma B.1 for the term H_{111}. The same is true for the cross terms. ■

Appendix C. Secondary lemmas

Lemma C.1. *Let Ω_M be $E(x_{iM} x_{iM}' \varepsilon_i^2)$ and Ω be the infinite operator associated with Ω_M. Then, under Assumptions 3.1 and 3.2 Ω_M^{-1} exists for all M, $\Omega \in L(l^2, l^2)$ and Ω^{-1} exists and is bounded.*

Proof. This proof follows Kuersteiner (2001, page 399). I first prove that $\Omega \in L(l^2, l^2)$. Write $\Omega = S + \sigma^2 I$, so that by Assumption 3.2, $\sum_{l=1,j=1}^{\infty} |s_{lj}| < \infty$. Now take $z \in l^2$. Then $\Omega z = Sz + \sigma^2 Iz = w$. Each of the elements in w are bounded since $|w_l| \le \sum_{j=1}^{\infty} |s_{lj} z_j| + |z_l| \sigma^2 \le (\sum_{j=1}^{\infty} |s_{lj}|^2)^{1/2} \|z\| + |z_l| \sigma^2 < \infty$. In fact, $\sum_{l=1}^{\infty} |w_l|^2 \le \|z\| \sum_{l=1,j=1}^{\infty} |s_{lj}|^2 + \|z\|^2 \sigma^4 < \infty$ since $\sum_{l=1,j=1}^{\infty} |s_{lj}|^2 < \infty$. Thus, $w \in l^2$. Notice also that $\sup_{\|z\| \le 1} \|\Omega z\| < \infty$, so that Ω is a bounded linear operator on l^2. By Theorem II.3.1 in Gohberg and Goldberg (1981) Ω is uniformly continuous and it is self-adjoint by its symmetry. To prove invertibility I use the fact that Ω is self-adjoint. This implies that the operator is invertible if the null space (denoted by $\ker \Omega$) contains only the null element in l^2. Assume on the contrary that there is $z \in l^2$ such that $z \neq (0, 0, \ldots)$ and

$\Omega z = 0$. Then $z'\Omega z = \sum_{j=1,l=1}^{\infty} z_j E(x_i^j x_i^l \varepsilon_i^2) z_l = E[\sum_{j=1}^{\infty} z_j x_i^j \varepsilon_i]^2 = 0$ which implies that $\sum_{j=1}^{\infty} z_j x_i^j \varepsilon_i = 0$ w.p.1. Then, either (a) $\sum_{j=1}^{\infty} z_j x_i^j = 0$ a.s. or (b) $x_i^j \varepsilon_i = 0$ a.s. $\forall j$. Condition (a) is ruled out by Assumption 3.1. On the other hand, (b) means that $x_i^{j2}\varepsilon_i^2 = 0$ a.s. and then $E(x_i^{j2}\varepsilon_i^2) = 0$ which contradicts Assumption 3.2. Thus, $\ker \Omega = \mathbf{0}$ and Ω^{-1} exists. Since Ω is closed (just for being a linear operator), the Closed Graph Theorem (Theorem X.4.2 in Gohberg and Goldberg, 1981) says that Ω^{-1} is bounded on l^2.

Finally, since $z_M' \Omega_M z_M = E[\sum_{j=1}^{M} z_{M,j} x_{M,i}^j \varepsilon_i]^2 > 0$ by Assumptions 3.1 and 3.2, Ω_M is positive definite and therefore invertible for all M. ■

Lemma C.2. *Let σ^2 be the constant defined in Assumption 3.2 and define*

$$\Omega_M^* = \begin{bmatrix} \Omega_M & 0 \\ 0 & \sigma^2 I_\infty \end{bmatrix}.$$

Then, under Assumptions 3.1 and 3.2 Ω_M^{*-1} exists and $\left\|\Omega_M^{*-1} - \Omega^{-1}\right\|_* \to 0$ as $M \to \infty$.

Proof. This proof follows Kuersteiner (2001, page 400). The fact that Ω_M^{*-1} exist is direct since Ω_M and $\sigma^2 I_\infty$ are invertible. Then I prove that $\left\|\Omega_M^* - \Omega\right\|_* \to 0$ as $M \to \infty$. Define S as in the previous lemma, $\Omega = S + \sigma^2 I$, and note that

$$\begin{aligned}\left\|\Omega_M^* - \Omega\right\|_* &\le \sup_{\|z\|\le 1}\left(\sum_{k=1}^{M}\left|\sum_{l=M+1}^{\infty} s_{kl} z_l\right|^2 + \sum_{k=M+1}^{\infty}\left|\sum_{l=1}^{\infty} s_{kl} z_l\right|^2\right)^{1/2} \\ &\le \sup_{\|z\|\le 1}\sum_{k=1}^{M}\sum_{l=M+1}^{\infty}|s_{kl}|\,|z_l| + \sum_{k=M+1}^{\infty}\sum_{l=1}^{\infty}|s_{kl}|\,|z_l| \\ &\le 2\sum_{k=1}^{\infty}\sum_{l=M+1}^{\infty}|s_{kl}| \to 0.\end{aligned}$$

Write $(\Omega^{-1} - \Omega_M^{*-1}) = \Omega^{-1}(\Omega_M^* - \Omega)\Omega_M^{*-1}$ so that $\left\|\Omega^{-1} - \Omega_M^{*-1}\right\|_* \le \left\|\Omega^{-1}\right\|_* \left\|\Omega_M^* - \Omega\right\|_* \left\|\Omega_M^{*-1}\right\|_*$. This goes to zero as $M \to \infty$ if $\left\|\Omega^{-1}\right\|_*$ and $\left\|\Omega_M^{*-1}\right\|_*$ are finite. $\left\|\Omega^{-1}\right\|_* < \infty$ since Ω^{-1} is bounded. On the other hand, $\left\|\Omega_M^{*-1}\right\|_* \le \left\|\Omega_M^{-1}\right\| + \sigma^{-2} < \infty$ since all the eigenvalues of Ω_M are bounded away from zero by Assumption 3.2 and Lemma C.1. ■

Lemma C.3. *Let $\widehat{\Omega}_M = \frac{1}{n}\sum_{i=1}^{n} x_{Mi} x_{Mi}' \widehat{\varepsilon}_i^2$ where $\widehat{\varepsilon}_i = y_i - \widetilde{\delta}' W_i$ and the preliminary estimator $\widetilde{\delta}$ satisfies $\sqrt{n}(\widetilde{\delta} - \delta) = O_p(1)$. Then, under Assumptions 3.1–3.3, $\|\widehat{\Omega}_M - \Omega_M\| = O_p(M/\sqrt{n})$.*

Proof. By Assumption 3.3, $E\left|\widehat{\omega}_{jl} - \omega_{jl}\right|^2 = O(n^{-1})$ uniformly in j and l. Then, $\|\widehat{\Omega}_M - \Omega_M\|^2 = \sum_{j,l=1}^{M}\left|\widehat{\omega}_{jl} - \omega_{jl}\right|^2 = O(M^2/n)$ and the result follows. ■

Lemma C.4. *Assumption 3.5 implies that $\lim_{M\to\infty}\sum_{j=cM+1}^{M} j^\alpha \left\|\Gamma_j\right\| = -\kappa \log(c) < \infty$ for $c \in (0, 1)$ and that $\sum_{j=1}^{M} j^s \left\|\Gamma_j\right\|$ converges for all $s < \alpha$.*

Proof. Using the change of variables $\xi = j/M$ with $\Delta\xi = 1/M$ we have:

$$\begin{aligned}\lim_{M\to\infty}\sum_{j=cM+1}^{M} j^\alpha \left\|\Gamma_j\right\| &= \lim_{M\to\infty}\sum_{j=cM+1}^{M}\frac{\kappa}{j} \\ &= \lim_{M\to\infty}\sum_{\xi=c+1/M}^{1}\frac{\kappa}{\xi M} = \kappa\int_c^1 \frac{1}{\xi}\mathrm{d}\xi = -\kappa \log(c) < \infty.\end{aligned}$$

Moreover, since $\sum_{j=1}^{M}\frac{1}{j^p} < \infty$ for $p > 1$, the second result follows. ■

Lemma C.5. *Define $\bar{k}_\alpha(x) = \frac{c}{1-c}\left(\frac{1/c}{x^{\alpha-1}} - \frac{1}{x^\alpha}\right)$ for $c < x \le 1$. Then, $\bar{k}_\alpha(j/M)$ is bounded when $cM < j \le M$ since $\max_{c<x\le 1}\bar{k}_\alpha(x) = 1$ if $c > 1 - 1/\alpha$ and $\max_{c<x\le 1}\bar{k}_\alpha(x) = \frac{c^{1-\alpha}}{1-c}\frac{(\alpha-1)^{\alpha-1}}{\alpha^\alpha} < \infty$ otherwise.*

Proof. The maximizer is given by $x^* = \min[1, c\alpha/(\alpha-1)]$ and this implies the result. ■

Lemma C.6. *Assumption 3.5 implies that $\lim_{M\to\infty}\sum_{j=cM+1}^{M} j^\alpha \Gamma_j \bar{k}_\alpha (j/M) < \infty$.*

Proof. To check this notice that

$$\begin{aligned}&\lim_{M\to\infty}\sum_{j=cM+1}^{M} j^\alpha \left\|\Gamma_j\right\| \left|\bar{k}_\alpha(j/M)\right| \\ &= \lim_{M\to\infty}\sum_{j=cM+1}^{M}\frac{\kappa}{j}\left|\frac{c}{1-c}\left(\frac{1/c}{(j/M)^{\alpha-1}} - \frac{1}{(j/M)^\alpha}\right)\right| \\ &= \lim_{M\to\infty}\frac{c\kappa}{1-c}\sum_{j=cM+1}^{M}\left(\frac{M^{\alpha-1}}{cj^\alpha} - \frac{M^\alpha}{j^{\alpha+1}}\right) \\ &= \lim_{M\to\infty}\frac{c\kappa}{1-c}\sum_{\xi=c+1/M}^{1}\left(\frac{M^{\alpha-1}}{c(\xi M)^\alpha} - \frac{M^\alpha}{(\xi M)^{\alpha+1}}\right) \\ &= \lim_{M\to\infty}\frac{c\kappa}{1-c}\sum_{\xi=c+1/M}^{1}\left(\frac{M^{-1}}{c\xi^\alpha} - \frac{M^{-1}}{\xi^{\alpha+1}}\right) \\ &= \frac{c\kappa}{1-c}\left(\frac{c^{-(\alpha-1)}-1}{c(\alpha-1)} - \frac{c^{-\alpha}-1}{\alpha}\right) \\ &= \frac{\kappa}{\alpha(\alpha-1)}\frac{(1-c^\alpha)}{(1-c)c^{\alpha-1}} - \frac{\kappa}{(\alpha-1)} < \infty\end{aligned}$$

where I used the change of variables $\xi = j/M$ with $\Delta\xi = 1/M$ and the fact that

$$\sum_{\xi=c+1/M}^{1}\frac{1}{\xi^\alpha}\frac{1}{M} \to \int_c^1 \xi^{-\alpha}\mathrm{d}\xi = \frac{c^{-(\alpha-1)}-1}{(\alpha-1)}. \quad ■ \tag{C.1}$$

Lemma C.7. *Assumption 3.5 implies $M^\alpha \sum_{j=M+1}^{\infty}\left\|\Gamma_j\right\| \to \frac{\kappa}{\alpha}$ and $M^{2\alpha+1}\sum_{j=M+1}^{\infty}\left\|\Gamma_j\right\|^2 \to \frac{\kappa}{2\alpha+1}$.*

Proof. For the first case notice that I can write $\sum_{j=M+1}^{\infty}\left\|\Gamma_j\right\| = \sum_{k=1}^{\infty}\sum_{j=kM+1}^{(k+1)M}\left\|\Gamma_j\right\|$. Doing the change of variables $\xi = j/M$ I have

$$\begin{aligned}M^\alpha\sum_{j=kM+1}^{(k+1)M}\left\|\Gamma_j\right\| &= M^\alpha\sum_{j=kM+1}^{(k+1)M}\frac{\kappa}{j^{\alpha+1}} = \sum_{\xi=k+1/M}^{(k+1)}\frac{\kappa}{\xi^{\alpha+1}}\frac{1}{M} \\ &\to \int_k^{k+1}\kappa\xi^{-(\alpha+1)}\mathrm{d}\xi = \frac{\kappa}{\alpha}\left(\frac{1}{k^\alpha} - \frac{1}{(k+1)^\alpha}\right)\end{aligned}$$

and the results follows from $\sum_{k=1}^{\infty}\left(\frac{1}{k^\alpha} - \frac{1}{(k+1)^\alpha}\right) = 1$. For the second case,

$$M^{2\alpha+1}\sum_{j=kM+1}^{(k+1)M}\left\|\Gamma_j\right\|^2 \to \frac{\kappa}{2\alpha+1}\left(\frac{1}{k^{2\alpha+1}} - \frac{1}{(k+1)^{2\alpha+1}}\right),$$

using the same change of variables. The result follows from $\sum_{k=1}^{\infty}\left(\frac{1}{k^{2\alpha+1}} - \frac{1}{(k+1)^{2\alpha+1}}\right) = 1$. ■

Lemma C.8. *Assumption 3.5 implies that*

$$\lim_{M\to\infty} M \sum_{j=cM+1}^{M}\sum_{l=cM+1}^{M} \frac{\kappa}{j}\bar{k}_\alpha(j/M)\bar{k}_\alpha(l/M)\frac{\kappa}{l}\vartheta_{jl}^M = F_{\alpha,\sigma,\kappa}(c)$$

where $F_{\alpha,\sigma,\kappa}(c) = \frac{\kappa^2c^2}{\sigma^2(1-c)^2}\left[\frac{1}{c^{2\alpha+1}(4\alpha^2-1)\alpha}+\frac{c(2\alpha-1)-\alpha}{c^2\alpha(2\alpha-1)}-\frac{1}{2\alpha+1}\right] < \infty$ *for* $c \in (0,1)$.

Proof. Note that I can write the expression as,

$$\lim_{M\to\infty} M \sum_{j=cM+1}^{M} \frac{\kappa^2}{j^2}\bar{k}_\alpha^2(j/M)\vartheta_{jj}^M$$
$$+2\lim_{M\to\infty} M \sum_{j=cM+1}^{M}\sum_{l>j}^{M} \frac{\kappa}{j}\bar{k}_\alpha(j/M)\bar{k}_\alpha(l/M)\frac{\kappa}{l}\vartheta_{jl}^M.$$

Consider first the case where $j = l$ and note that $j^{-2}\bar{k}_\alpha^2(j/M) = \frac{c^2}{(1-c)^2}\left[\frac{1}{c^2}\frac{M^{2\alpha-2}}{j^{2\alpha}}+\frac{M^{2\alpha}}{j^{2\alpha+2}}-\frac{2M^{2\alpha-1}}{cj^{2\alpha+1}}\right]$. As in the previous Lemma I use the approximation in (C.1) into the sum $M\kappa^2\sum_{j=cM+1}^{M} j^{-2}\bar{k}_\alpha^2(j/M)\vartheta_{jj}^M$ and the fact that $\left|\vartheta_{jj}^M-\frac{1}{\sigma^2}\right| = o\left(\frac{1}{j}\right)$ to get

$$M\kappa^2\sum_{j=cM+1}^{M} j^{-2}\bar{k}_\alpha^2(j/M)\vartheta_{jj}^M$$
$$= \frac{\kappa^2c^2}{\sigma^2(1-c)^2}M\sum_{j=cM+1}^{M}\left[\frac{1}{c^2}\frac{M^{2\alpha-2}}{j^{2\alpha}}+\frac{M^{2\alpha}}{j^{2\alpha+2}}-\frac{2M^{2\alpha-1}}{cj^{2\alpha+1}}\right]+o(1)$$
$$= \frac{\kappa^2c^2}{\sigma^2(1-c)^2}M\sum_{\xi=c+1/M}^{1}\left[\frac{1}{c^2}\frac{1}{\xi^{2\alpha}M^2}+\frac{1}{\xi^{2\alpha+2}M^2}-\frac{2}{c\xi^{2\alpha+1}M^2}\right]+o(1)$$
$$\to \frac{\kappa^2c^2}{\sigma^2(1-c)^2}\left[\frac{1}{c^2}\frac{c^{-(2\alpha-1)}-1}{(2\alpha-1)}+\frac{c^{-(2\alpha+1)}-1}{(2\alpha+1)}-\frac{2}{c}\frac{c^{-2\alpha}-1}{2\alpha}\right]$$

as $M\to\infty$, which is equal to $F_{\alpha,\sigma,\kappa}(c)$ after some algebra.

Now consider the case where $l > j$. The claim is that,

$$\lim_{M\to\infty} M \sum_{j=cM+1}^{M}\sum_{l>j}^{M} \frac{\kappa}{j}\bar{k}_\alpha(j/M)\bar{k}_\alpha(l/M)\frac{\kappa}{l}|\vartheta_{jl}^M| = 0. \quad \text{(C.2)}$$

To show this I use that $\sum_{l>j}^{\infty}\left|\vartheta_{jl}\right| = o(1/j)$, which is implied by $\sum_{j=1}^{\infty}\sum_{l>j}^{\infty}\left|\vartheta_{jl}\right| < \infty$. Note that if this were not the case, say, $\sum_{l>j}^{\infty}\left|\vartheta_{jl}\right| \geq \frac{C}{j}$, for $C > 0$ then $\sum_{j=1}^{\infty}\sum_{l>j}^{\infty}\left|\vartheta_{jl}\right| \geq \sum_{j=1}^{\infty}\frac{C}{j} = \infty$. Now define $G_j = \sum_{l>j}^{M}\bar{k}_\alpha(l/M)\frac{\kappa}{l}\vartheta_{jl}^M$ and note that since $\frac{\kappa}{l}\leq\kappa$ and $\bar{k}_\alpha(l/M)\leq C$ for some finite C,

$$G_j \leq C\sum_{l>j}^{M}|\vartheta_{jl}^M| = C\times o\left(\frac{1}{j}\right),$$

so $\exists\epsilon > 0$ such that $G_j \leq \frac{C}{j^{1+\epsilon}}$. Using this into (C.2),

$$\lim_{M\to\infty} M\sum_{j=cM+1}^{M}\frac{\kappa}{j}\bar{k}_\alpha(j/M)G_j \leq \lim_{M\to\infty} M\sum_{j=cM+1}^{M}\frac{\kappa}{j}\bar{k}_\alpha(j/M)\frac{C}{j^{1+\epsilon}}$$
$$\leq \lim_{M\to\infty}\frac{C}{M^\epsilon}\sum_{j=cM+1}^{M}\left(\frac{M}{j}\right)^{2+\epsilon}\frac{1}{M} = 0, \quad \text{(C.3)}$$

since $\sum_{j=cM+1}^{M}\left(\frac{M}{j}\right)^{2+\epsilon}\frac{1}{M} < \infty$ by (C.1). This completes the proof. ∎

Lemma C.9. *Let* $a_l = \sum_{j=1}^{M}\Gamma_j\vartheta_{jl}$, *then* $\sum_{l=cM+1}^{M} l^\alpha\|a_l\| \leq \sum_{l=cM+1}^{M} l^\alpha\|\Gamma_l\||\vartheta_{ll}|+o(1)$. *Furthermore, letting* $b_l = \sum_{j=1}^{\infty}\Gamma_j\vartheta_{jl}$ *we get that* $\sum_{l=M+1}^{\infty}\|b_l\| \leq \sum_{l=M+1}^{\infty}\|\Gamma_l\||\vartheta_{ll}|+o(1)$.

Proof. Notice that this is immediately true when Ω^{-1} is diagonal. Although this is not the case here, Assumptions 3.2 and 3.5 imply that the lower block of Ω^{-1} converges to a diagonal matrix and then the lemma holds. To see this for the case of a_l notice that

$$\sum_{l=cM+1}^{M} l^\alpha\|a_l\| \leq \sum_{l=cM+1}^{M} l^\alpha\|\Gamma_l\||\vartheta_{ll}| + \sum_{l=cM+1}^{M} l^\alpha\sum_{j\neq l}^{M}\|\Gamma_l\||\vartheta_{ll}|$$
$$= \sum_{l=cM+1}^{M} l^\alpha\|\Gamma_l\||\vartheta_{ll}|+o(1),$$

since $\sum_{j\neq l}^{M}\|\Gamma_l\||\vartheta_{ll}| \leq C\sum_{j\neq l}^{M}|\vartheta_{ll}| = C\times o\left(\frac{1}{l^{\alpha+1}}\right)$ by Assumption 3.5(b) and then $\sum_{l=cM+1}^{M} l^\alpha\times C\times o(\frac{1}{l^{\alpha+1}}) = o(1)$ by the same arguments as those in the previous lemma. A similar argument applies to b_l. ∎

References

Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59 (3), 817–858.
Barndorff-Nielsen, O.E., Cox, D.R., 1989. Asymptotic Techniques for Use in Statistics. Chapman and Hall.
Bekker, P.A., 1994. Alternative approximations to the distributions of instrumental variable estimators. Econometrica 62 (3), 657–681.
Brown, B.W., Newey, W., 2002. Generalized method of moments, efficient bootstrapping, and improved inference. Journal of Business and Economics Statistics 20 (4), 507–517.
Canay, I.A., 2009. Supplementary appendix to simultaneous selection and weighting of moments in GMM using a trapezoidal kernel, available at http://faculty.wcas.northwestern.edu/~iac879/.
Chamberlain, G., Imbens, G., 2004. Random effects estimators with many instrumental variables. Econometrica 72 (1), 295–306.
Donald, S.G., Imbens, G., Newey, W.K., 2008. Choosing instrumental variables in conditional moment restriction models. Journal of Econometrics 152 (1), 28–36.
Donald, S.G., Newey, W.K., 2001. Choosing the number of instruments. Econometrica 69 (5), 1161–1191.
Gohberg, I., Goldberg, S., 1981. Basic Operator Theory. Birkhauser.
Hahn, J., Hausman, J., 2002. A new specification test for the validity of instrumental variables. Econometrica 70 (1), 163–189.
Hahn, J., Hausman, J., Kuersteiner, G., 2004. Estimation with weak instruments: Accuracy of higher order bias and mse approximations. Econometrics Journal 7 (1), 272–306.
Han, C., Phillips, P.C.B., 2006. Gmm with many moment conditions. Econometrica 74 (1), 147–192.
Hansen, C., Hausman, J., Newey, W.K., 2006. Estimation with many instrumental variables, manuscript, M.I.T.
Hausman, J., Newey, W.K., Woutersen, T., Chao, J., Swanson, N., 2009. Instrumental variable estimation with heteroskedasticiy and many instruments, manuscript, M.I.T.
Inoue, A., 2006. A bootstrap approach to moment selection. Econometrics Journal 9, 48–75.
Kuersteiner, G., 2001. Optimal instrumental variable estimation for arma models. Journal of Econometrics 104 (2), 359–405.
Kuersteiner, G., 2002. Mean squared error reduction for gmm estimators of linear time series models, manuscript, Boston University.
Morimune, K., 1983. Approximate distribution of k-class estimators when the degree of overidentification is large compared with the sample size. Econometrica 51 (3), 821–841.
Okui, R., 2008. Instrumental variable estimation in the presence of many moment conditions, Journal of Econometrics (submitted for publication).
Parzen, E., 1957. On consistency estimates of the spectrum of a stationary time series. Econometrica 28 (2), 329–348.
Politis, D.N., 2007. Higher order accurate, positive semi-definite estimation of large sample covariance and spectral density matrices, manuscript, UCSD.
Politis, D.N., Romano, J.P., 1995. Bias-corrected nonparametric spectral estimation. Journal of Time Series Analysis 16, 67–103.
Politis, D.N., Romano, J.P., 1999. Multivariate density estimation with general flat-top kernels of infinite order. Journal of Multivariate Analysis 68, 1–25.
Rothenberg, T.J., 1984. Approximating the distributions of econometrics estimators and test statistics. In: Griliches, Z., Intriligator (Eds.), Handbook of Econometrics, vol. 2. Academic Press, New York.