Kernel Weighted GMM Estimators for Linear Time Series Models

Guido M. Kuersteiner*

This Version: December 2011

Abstract

This paper analyzes the higher order asymptotic properties of Generalized Method of Moments (GMM) estimators for linear time series models using many lags as instruments. A data dependent moment selection method based on minimizing the approximate mean squared error is developed. In addition, a new version of the GMM estimator based on kernel weighted moment conditions is proposed. It is shown that kernel weighted GMM can reduce the asymptotic bias compared to standard GMM. Kernel weighting also helps to simplify the problem of selecting the optimal number of instruments. A feasible procedure similar to optimal bandwidth selection is proposed for the kernel weighted GMM estimator.

Key Words: time series, feasible GMM, number of instruments, kernel weights, higher order MSE, bias reduction

* *Georgetown University, Dept. of Economics, Washington D.C.* http://www9.georgetown.edu/faculty/gk232/. This paper is a revised version of a manuscript circulated under the title "Mean Squared Error Reduction For GMM Estimators of Linear Time Series Models". I wish to thank Xiaohong Chen, Ronald Gallant, Jinyong Hahn, Jerry Hausman, Whitney Newey and Ken West for helpful comments. Stavros Panageas and Sungju Chun provided excellent research assistance. The comments of three anonymous referees and a Co-editor have greatly improved the exposition. Financial support from NSF grant SES-0095132 and SES-0523186 is gratefully acknowledged.

1. Introduction

This paper analyzes the higher order asymptotic properties of GMM estimators for linear time series models where the number of lagged instruments is potentially large. It is well known in the cross-sectional literature that using a large number of instruments can result in substantial second order bias of GMM estimators. This limits the implementation of efficient procedures. Similar results are obtained in this paper for the time series case. The analysis of the higher order mean squared error shows that the trade off between bias and variance can be manipulated by using a kernel weighting procedure for the instruments.

Expansions similar to the ones of Donald and Newey (2001) are obtained for the case of GMM estimators in models with lagged dependent right hand side variables and many lagged instruments. Based on these expansions, approximations for the higher order asymptotic mean squared error are obtained. The time series case is more difficult to analyze than the situation with cross-sectional samples and exogenous instruments because the expansion cannot be developed conditional on the instruments. Most work on many instrument asymptotics therefore considers situations where the data are sampled independently. Nevertheless, moment selection for the time series case when the number of instruments is fixed was considered by Inoue (2006) and bias properties of GMM estimators for time series models with a fixed number of instruments were investigated by Anatolyev (2005).

Minimizing the asymptotic approximation to the MSE with respect to the number of lagged instruments leads to a feasible GMM estimator for time series models. The trade-off in adding additional instruments is between asymptotic efficiency and bias. Because of the endogeneity of lagged instruments, the Mallow's criterion employed by Donald and Newey (2001) cannot be readily applied to the time series case. The approach taken here is to estimate nuisance parameters that enter the MSE formula with a VAR approximation to the reduced form data-generating process. The approximating VAR is allowed to grow in dimension as the sample size expands.

A second contribution of the paper is to propose a new kernel weighted version of GMM. Kernel weighting is based on the insight that in a time series context where lagged endogenous variables are used as instruments, more distantly lagged instruments are typically less informative for the first stage and contribute more to estimator bias in proportion to their information content than more recent instruments. Kernel weighting then modifies the GMM weight matrix to put less weight on moment conditions with farther lagged instruments. Since the first version of this paper was circulated[1] this idea has been adapted to cross-sectional settings by Okui (2005), Canay (2006) and Kuersteiner and Okui (2010). Regularization methods to address the many instrument problem are considered by Carrasco

[1] See Kuersteiner (2002b)

(2011) for iid models. Here, it is shown that the asymptotic bias can be reduced if suitable kernel weights are applied to the moment conditions. An added benefit of kernel weighting is that it simplifies the instrument selection problem to a bandwidth selection problem similar to the one encountered in the HAC estimation literature. The downside of kernel weighting is a higher order efficiency loss. However, Monte Carlo experiments indicate that the benefits of kernel weighting outweigh the efficiency loss for a wide range of data-generating processes.

The paper is organized as follows. Section 2 presents the time series models and introduces notation. Section 3 introduces the kernel weighted GMM estimator, contains the analysis of higher order asymptotic MSE terms and derives a selection criterion for the optimal number of instruments. Section 4 discusses implementation of the procedure, in particular consistent estimation of the criterion function for optimal bandwidth selection. Section 5 contains a small Monte Carlo experiment. Technical definitions, assumptions and proofs are collected in the appendix. Additional proofs and lemmas are collected in an auxiliary appendix published on the author's web-page.

2. Linear Time Series Models

The econometric model considered in this paper is similar to the linear time series framework of Hansen and Singleton (1991). Let $y_t \in \mathbb{R}^p$ be a strictly stationary stochastic process. It is assumed that y_t satisfies a structural econometric equation implied by restrictions obtained from economic theory. In order to describe this structural equation partition $y_t = \left[y_{t,1}, y_{t,2}', y_{t,3}'\right]$. Here, $y_{t,1}$ is the scalar left hand side variable, $y_{t,2}$ are the included and $y_{t,3}$ are the excluded contemporaneous variables. The vector X_t is defined to contain, possibly a subset, of the lagged variables $y_{t-1}, ..., y_{t-r}$ where r is known and fixed. The structural equation then takes the form

$$y_{t,1} = \alpha_0 + \beta_0' y_{t,2} + \beta_1' X_t + \varepsilon_t. \tag{2.1}$$

The structural model also imposes restrictions on the innovations ε_t. More specifically, ε_t is strictly stationary with $E\left[\varepsilon_t\right] = 0$ and follows a moving average (MA) process of order $m-1$ for $m \geq 1$, where m is assumed known and finite. Denote the autocovariance function of ε_t by $\gamma_j^{\varepsilon} = E\left[\varepsilon_t \varepsilon_{t-j}\right]$ with $\gamma_j^{\varepsilon} = 0$ for $|j| \geq m$.

Letting $\beta = \left(\beta_0', \beta_1'\right)' \in \mathbb{R}^d$ and collecting all the regressors in x_t where $x_t = \left(y_{t,2}', X_t'\right)'$ one can write (2.1) as $y_{t,1} = \alpha_0 + \beta' x_t + \varepsilon_t$. An alternative representation of (2.1) is obtained by setting $a(z, \beta) = a_0 + a_1 z + ... + a_r z^r$ with $1 \times p$ vectors a_i and $z \in \mathbb{C}$, the set of complex numbers, such that $a(L, \beta) y_t = \alpha_0 + \varepsilon_t$ where L is the lag operator. Note that a_i are subject to exclusion and normalization restrictions implied by (2.1).

The assumption of strict stationarity together with assumptions that guarantee the existence of moments of sufficiently high order imply that y_t admits an infinite order moving average representation

$$y_t = \mu_y + B(L)u_t \tag{2.2}$$

by the Wold representation theorem. Here, $\mu_y \in \mathbb{R}^p$ is a constant, u_t is a strictly stationary white noise sequence and $B(z) = \sum_{j=0}^{\infty} B_j z^j$ with $B_0 = I_p$ where I_p is the p-dimensional identity matrix. It is assumed that $B(z)^{-1}$ exists for $|z| \leq 1$ and has a convergent expansion $\pi(z) = I - \sum_{j=1}^{\infty} \pi_j z^j$. The coefficient matrices π_j are the coefficients of the infinite order AR representation of y_t and are assumed to satisfy $\sum_{j=k}^{\infty} \|\pi_j\| = O(\nu^k)$ for some $\nu \in [0, 1)$. The structural equation (2.1) imposes certain constraints on $B(z)$. To see this, partition $B(z) = [B_1(z)', B_2(z)']'$ where $B_1(z)$ is a $1 \times p$ vector, $B_2(z)$ is a $(p-1) \times p$ matrix of lag polynomials and $\mu_y = [\mu_{y1}, \mu'_{y2}]$ is partitioned conformingly. Define $\alpha_0 \equiv a(1, \beta)\mu_y$ and let

$$A(L) = \begin{bmatrix} a(L,\beta) \\ 0 \quad I_{p-1} \end{bmatrix}.$$

Now premultiply both sides of (2.2) with $A(L)$ such that

$$A(L)y_t = \begin{bmatrix} \alpha_0 \\ \mu_{y,2} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ B_2(L)u_t \end{bmatrix}$$

with $\varepsilon_t = b(L)u_t$. The polynomial $b(L)$ is of order $m-1$. The restrictions imposed on $B(L)$ then are that $a(L,\beta)B(L) = b(L)$ and $a(L,\beta)\mu = \alpha_0$. Note that (2.1) does not impose any constraints on the elements of B_j related to the excluded variables $y_{t,3}$.

The restrictions imposed on $B(L)$ do not imply that there is a reduced form for x_t that depends on a finite dimensional parameter, as would be the case for example if x_t had a finite order VAR representation. The higher order asymptotic expansions of this paper therefore do not apply to GMM estimators such as the ones by West and Wilcox (1996), Kuersteiner (2001, 2002a) and West, Wong and Anatolyev (2009) that explicitly rely on such a parameterization. These estimators depend on the correct specification of the reduced from for x_t to achieve the first order asymptotic efficiency bounds while the estimators considered in this paper achieve the bounds irrespective of the data generating mechanism for x_t.

The economic model (2.1) implies moment restrictions of the form

$$E[\varepsilon_{t+m} y_{t-j}] = 0 \text{ for all } j \geq 0 \tag{2.3}$$

because $\varepsilon_{t+m} = b(L)u_{t+m}$ depends on $u_{t+m}, ..., u_{t+1}$ while the instruments $y_t, y_{t-1}, ...$ only depend on $u_t, u_{t-1}, ...$ and because u_t is white noise. The moment restrictions (2.3) are the basis for the formulation of GMM estimators using an Mp dimensional vector of instruments $\tilde{z}_{t,M} = \left(y_t', y_{t-1}', ..., y_{t-M+1}'\right)$. Alternatively, the moment restrictions (2.3) are often implied by economic theory and then lead to the formulation of a structural model of the form (2.1). A well known example are the asset pricing models discussed in Hansen and Singleton (1996).

The innovations u_t are not assumed to be martingale difference sequences. As a result the moment condition (2.3) is an unconditional rather than a conditional moment restriction and the only valid instruments are linear in y_{t-j} for $j \geq 0$. This is in contrast to efficiency results for models with conditional moment restrictions due to Hansen (1985) and Hansen, Heaton and Ogaki (1988) who develop GMM efficiency bounds achieved by instruments that are possibly non-linear in y_{t-j}.

One of the main contributions of this paper is to develop a data-dependent method for selecting the parameter M based on a higher order approximation to the mean squared error of the GMM estimator. Asymptotically, M needs to tend to infinity in order to exploit all moment conditions in (2.3) and to achieve first order efficiency. In finite samples, the choice of M is limited by data-availability, but more importantly by a need to balance bias and variance when including additional instruments. The higher order analysis in Section 3 provides the tools to chose M optimally in a way that lets M tend to infinity with the sample size for efficiency reasons but does so slowly enough to control for higher order and finite sample bias.

Detailed technical assumptions are listed in the appendix. It is assumed that u_t is a homoskedastic white noise sequence. In addition, a restriction proposed by Hayashi and Sims (1983) is imposed, namely that $E(u_t u_s' | \tilde{z}_{t-1,M}) = E\left(u_t u_s'\right)$ for all $t \geq s$ and M. More specifically, this implies that $E(u_t u_t' | \tilde{z}_{t-1,M}) = \Sigma$ for some positive definite and nonrandom matrix Σ and $E(u_t u_s' | \tilde{z}_{t-1,M}) = 0$ for $t > s$. The assumption that the covariance structure of the process u_t does not depend on $\tilde{z}_{t,M}$ is restrictive as it rules out time changing variances and conditional heteroskedasticity. Relaxing this restriction results in more complicated GMM weight matrices of the type analyzed in Kuersteiner (2001). Homoskedasticity both simplifies the higher order expansion and thus expressions that determine the optimal choice of M and allows to establish theoretical properties of data-dependent choices for M that are more difficult to establish in a more general setting. In principle, the higher order moment restriction implied by conditional homoskedasticity could be used for estimation in addition to the conditions (2.3). The resulting estimator is however nonlinear and will not be considered here.

The regularity conditions listed in the Appendix also involve the existence of moments of up to order 12 for the innovation sequence. Existence of moments up to order 12 is not necessary to establish the first order asymptotic properties of the GMM estimator but is required to derive the approximation to the higher order MSE of the estimator. The assumptions on higher order moments are more stringent than the ones typically found in the cross-sectional literature. The reason is that the MSE of the GMM estimator cannot be computed conditionally on the instruments $\tilde{z}_{t,M}$ unless they are strictly exogenous.[2]

3. Kernel Weighted GMM

In this paper a generalized class of GMM estimators based on kernel weighted moment restrictions is introduced. In Theorem 3.1 it is shown that certain kernel functions reduce the higher order bias of GMM estimators. In other words, kernel weighting is a way to change the higher order variance-bias trade-off of an increasing number of overidentifying restrictions that is known to affect GMM.

It is assumed that a sample of size n, $y_1, ..., y_n$, generated by the process (2.2) and subject to the restrictions imposed by (2.1) is observed. The kernel weighted GMM estimator $\hat{\beta}_{n,M}$ is now discussed. Define the instrument vector $\tilde{z}_{t,M} = (y_t', y_{t-1}', ..., y_{t-M+1}')'$. An instrument selection matrix $S_M(t) = \operatorname{diag}(\mathbf{1}\{t \geq 1\}, ..., \mathbf{1}\{t \geq M\})$, where $\mathbf{1}\{.\}$ is the indicator function, is introduced to exclude instruments for which there is no data in the sample. The vector of available instruments is denoted by $z_{t,M} = (S_M(t) \otimes I_p)(\tilde{z}_{t,M} - \mathbf{1}_M \otimes \bar{y})$ where $\bar{y} = n^{-1}\sum_{t=1}^n y_t$, I_p is the p-dimensional identity matrix and $\mathbf{1}_M = (1, ..., 1)'$ is a vector of length M with all elements equal to 1. Let Z_M be the matrix of stacked instruments $Z_M = [z_{\max(1,r-m+1),M}, ..., z_{n-m,M}]'$ and $X = [x_{\max(m+1,r+1)} - \bar{x}, ..., x_n - \bar{x}]'$ the matrix of regressors. Also, Y is the stacked vector of the first demeaned element in y_t. Then define the $d \times Mp$ matrix $\hat{P}_M' = n^{-1}X'Z_M$ as well as the $Mp \times 1$ vector $\hat{P}_M^y = n^{-1}Z_M'Y$. Let $\hat{\Omega}_M$ be an estimator of the optimal weight matrix $\Omega_M = \sum_{l=-m+1}^{m-1} \gamma_l^\varepsilon \operatorname{Cov}[\tilde{z}_{t,M}, \tilde{z}_{t+l,M}]$. Assuming that M is such that $M \geq d/p$, where d is the dimension of β, the GMM estimator $\hat{\beta}_{n,M}$ can now be written as

$$\hat{\beta}_{n,M} = \left(\hat{P}_M' W_M \hat{\Omega}_M^{-1} W_M \hat{P}_M\right)^{-1} \hat{P}_M' W_M \hat{\Omega}_M^{-1} W_M \hat{P}_M^y. \tag{3.1}$$

The weight matrix W_M is defined as $W_M = (w_M \otimes I_p)$ where w_M is a diagonal matrix

$$w_M = \operatorname{diag}(k(0), ..., k((M-1)/M))'$$

having weight $k((j-1)/M)$ in the j-th diagonal element and zeros otherwise and where $k(.)$ is a kernel function satisfying properties outlined in Assumption A below. The general kernel weighted approach covers standard GMM as a special case when the truncated kernel $k(j/M) = \mathbf{1}\{|j/M| \leq 1\}$ is used. In that case $W_M = I_{Mp}$ and $\hat{\beta}_{n,M}$ is the conventional GMM estimator based on the instrument vector $z_{t,M}$.

[2] See for example Donald and Newey (2001) or Kuersteiner and Okui (2010).

The formulation of $\hat{\beta}_{n,M}$ does not depend on homoskedasticity. The weight matrix Ω_M can be estimated by a HAC type estimator such as Andrews (1991) or Newey and West(1994) which will lead to improved first order asymptotic efficiency when u_t is heteroskedastic. The optimal data-dependent choice for M developed in this paper on the other hand does depend on homoskedasticity. If M is chosen according to the methods of this paper but homoskedasticity fails in the data then the choice of M will be suboptimal. However, since Ω_M affects first order properties and M only affects higher order properties of $\hat{\beta}_{n,M}$ it is plausible that an estimator $\hat{\beta}_{n,M}$ with robust weight matrix and with a data-dependent M that ignores heteroskedasticity will still do well in finite samples.

The constant α_0 in (2.1) can be estimated as $\hat{\alpha}_0 \equiv \bar{y} - \bar{x}'\hat{\beta}_{n,M}$. The estimator $\left(\hat{\alpha}_0, \hat{\beta}'_{n,M}\right)'$ has the same first order limiting distribution as an alternative GMM estimator that includes a constant both in x_t and $z_{t,M}$ rather than working with the demeaned variables as is done here. The latter estimator has a stochastic expansion with additional higher order terms because it implicitly estimates the mean of the instruments with less than n observations.

To describe estimation of the weight matrix $\hat{\Omega}_M$, let $\hat{\gamma}^{\varepsilon}(l) = \frac{1}{n}\sum_{t=m+r+1}^{n} \hat{\varepsilon}_t \hat{\varepsilon}_{t-l}$ for $l \geq 0$ with $\hat{\varepsilon}_t = a(L, \tilde{\beta}_{n,M})(y_t - \bar{y})$, some consistent first stage estimator $\tilde{\beta}_{n,M}$ and $\hat{\gamma}^{\varepsilon}(l) = \hat{\gamma}^{\varepsilon}(-l)$ for $l < 0$.[3] Then define $\hat{\Omega}_M(l) = \frac{1}{n}\sum_{t=1}^{n} z_{t,M} z'_{t-l,M}$ for $l \geq 0$, $\hat{\Omega}_M(l) = \hat{\Omega}_M(-l)'$, for $l < 0$, $\hat{\Omega}^*_M = \sum_{l=-m+1}^{m-1} \hat{\gamma}^{\varepsilon}(l)\hat{\Omega}_M(l)$ and

$$\hat{\Omega}_M = \hat{\Omega}^*_M \mathbf{1}\left\{\min \hat{\xi}_\Omega \geq 0\right\} + \left(1 - \mathbf{1}\left\{\min \hat{\xi}_\Omega \geq 0\right\}\right) \sum_{l=-m+1}^{m-1} \left(1 - \frac{|l|}{m}\right) \hat{\gamma}^{\varepsilon}(l)\hat{\Omega}_M(l) \tag{3.2}$$

where $\min \hat{\xi}_\Omega$ is the smallest eigenvalue of $\hat{\Omega}^*_M$. The estimator $\hat{\Omega}_M$ is equal to the simpler form $\hat{\Omega}^*_M$ with probability approaching one (wpa1), but unlike $\hat{\Omega}^*_M$, is guaranteed to be positive semi-definite in finite samples.

The effects of using kernel weighted moments can be inferred from (3.1). The kernel matrix W_M distorts the variance of the estimator by using $W_M \hat{\Omega}_M^{-1} W_M$ instead of the optimal $\hat{\Omega}_M^{-1}$ as weight matrix. As is shown below, these effects are second order for suitable choices of the kernel function $k(j/M)$ and bandwidth M. Regularity conditions for kernel weights $k(.)$ are introduced next.

Assumption A. *Let* $\mathcal{K} = \{k(.)|k(.) : \mathbb{R} \to [-1, 1], k(0) = 1, k(x) = 0$ *for* $|x| > 1, k(x) = k(-x), k(.)$ *is continuous except at a countable number of points*$\}$. *Define* $k_q = \lim_{x\to 0}(1 - k(x))/|x|^q$. *In addition* $k(.) \in \mathcal{K}$ *satisfies one of the following two assumptions:*

[3] For M fixed and possibly small, a first stage estimator can be obtained from standard inefficient GMM procedures where $\hat{\Omega}_M = I_{Mp}$. In the Monte Carlo simulations the first stage estimator is based on the choice M that is selected as the optimal lag length in a VAR(p) approximation to $\pi(z)$.

A1 (Truncated Kernel). *For all $q \in [0, \infty)$ it follows that $k_q = 0$.*

A2 (Smooth Kernel). *There exists a smallest number $q \in (0, \infty)$ such that $0 < k_q < \infty$.*

The most important kernel satisfying A1 is the truncated kernel $k(x) = \mathbf{1}\{|x| \leq 1\}$. Assumption A2 rules out certain parametric kernel functions such as the Quadratic Spectral kernel but is satisfied by a number of well known kernels such as the Truncated, Bartlett, Parzen and Tukey-Hanning kernels.

Assumption A1 and A2 correspond to the assumptions made in Andrews (1991) except that here an additional requirement is that $k(x) = 0$ for $|x| > 1$. This ensures that only a finite number of moment conditions, controlled by the bandwidth parameter, are used and therefore simplifies estimation of the weight matrix $\hat{\Omega}_M$. The constraint $k(0) = 1$ is introduced for notational convenience only since $\hat{\beta}_{n,M}$ is invariant to the scale of W_M.

3.1. Higher Order Approximations

The higher order approximation to the MSE of $\hat{\beta}_{n,M}$ is derived from a Nagar (1959) type approximation similar to the one used in Donald and Newey (2001) for the iid case. Let $\hat{\beta}_{n,M}$ be stochastically approximated by $b_{n,M}$ such that

$$n^{1/2}(\hat{\beta}_{n,M} - \beta) = b_{n,M} + r_{n,M} \tag{3.3}$$

where $r_{n,M}$ is an error term with properties discussed below. To that end define $P_M = \text{Cov}(z_{t,M}, x_{t+m})$ and let $D = \lim_{M \to \infty} D_M$ where $D_M = P_M' \Omega_M^{-1} P_M$ such that D^{-1} is the efficiency lowerbound for $\hat{\beta}_{n,M}$.[4,5] For $\ell \in \mathbb{R}^d$ with $\ell'\ell = 1$ define the approximate mean squared error $\varphi_n(M, \ell, k(.))$ of $\ell'\hat{\beta}_{n,M}$ as in Donald and Newey (2001) as

$$\ell' E\left[b_{n,M} b_{n,M}'\right] \ell = \ell' D^{-1} \ell + \varphi_n(M, \ell, k(.)) + R_{n,M}$$

[4] The matrix D depends on the infinite dimensional inverse of Ω, Ω^{-1}. The existence of this inverse is established in Lemma A.7 by showing that the elements of Ω^{-1} have closed form expressions. The proof uses a similar result established by Lewis and Reinsel (1985). The representation in Lemma A.7 depends on the homoskedasticity of the error term. Invertibility of Ω when the errors are heteroskedastic is a more delicate problem - see Kuersteiner (2001) for some discussion.

[5] The efficiency lowerbound depends on infinitely many lagged instruments $z_t, z_{t-1}, ...$It can only be achieved asymptotically as $M, n \to \infty$. However, by Kolmogorov's existence theorem the process z_t exists for $t \in \{..., -1, 0, 1, ...\}$ which allows to define D.

and require that the error terms $r_{n,M}$ and $R_{n,M}$ satisfy[6]

$$\text{(3.4)}\qquad \frac{\|r_{n,M}\|^2 + R_{n,M}}{\varphi_n(M,\ell,k(.))} = o_p(1) \text{ as } M \to \infty, n \to \infty, M/n^{1/3} \to 0.$$

The main difference to Donald and Newey (2001) is that in the time series case $\varphi_n(M,\ell,k(.))$ is an unconditional expectation. As noted by Donald and Newey, the approximation is only valid for $M \to \infty$. Theorems 3.1 and 3.3 below establish that the approximate MSE, $\varphi_n(M,\ell,k(.))$, consists of a bias term which is of order $O(M/\sqrt{n})$ and a variance term which is of order $O(\|D - D_M\|)$ for kernels satisfying Assumption A1 and of order $O\left(M^{-2q}\right)$ for kernels satisfying Assumption A2. By assumption $M, n \to \infty$ and $M/\sqrt{n} \to 0$ such that $\varphi_n(M,\ell,k(.)) = o(1)$. This means that $\varphi_n(M,\ell,k(.))$ is a higher order MSE term that disappears asymptotically. To first order, the asymptotic distribution of $\hat{\beta}_{n,M}$ is unbiased with variance D^{-1}.

The first result concerns the exact nature of the bias term which does depend on the kernel function.

Theorem 3.1. *Suppose $k(.)$ satisfies Assumption A and Assumptions B and C in the Appendix hold. Let $\Gamma^{\varepsilon x}_{t-s} = E[\varepsilon_t x_s]$, $\Gamma^{\varepsilon y}_{t-s} = E[\varepsilon_{t+m} y_{s+1}]$, $\Gamma^{xy}_j = cov(x_{t+m}, y_{t-j+1})$ and define $f_{\varepsilon x}(\lambda) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\Gamma^{\varepsilon x}_j e^{-i\lambda j}$, $f_\varepsilon(\lambda) = \frac{1}{2\pi}\sum_{j=-m+1}^{m-1}\gamma^\varepsilon_j e^{-i\lambda j}$ and $f_{\varepsilon y}(\lambda) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\Gamma^{\varepsilon y}_j e^{-i\lambda j}$. Let $f^a(\lambda) = \sum_{j_1,j_2=1}^{\infty}\Gamma^{xy}_{j_1}\vartheta_{j_1,j_2}e^{-i\lambda j_2}$ where $\vartheta_{j,k}$ is the j,k-th block of Ω^{-1} which is shown to exist in Lemma (A.7). Define*

$$\text{(3.5)}\qquad \mathcal{A}_1 = (4\pi)^{-1}\int_{-\pi}^{\pi} f_{\varepsilon x}(\lambda) f_\varepsilon^{-1}(\lambda) d\lambda,\ \mathcal{A}_2 = 2^{-1}\int_{-\pi}^{\pi} f^a(\lambda) f_{\varepsilon y}(\lambda) d\lambda$$

Assume that $M, n \to \infty$ and $M/n^{1/3} \to 0$. Then, for $b_{n,M}$ defined in (3.3),

$$\text{(3.6)}\qquad \lim_{n\to\infty}\frac{\sqrt{n}}{M}E[b_{n,M} - \beta] = pD^{-1}\left(\mathcal{A}_1\int_{-1}^{1}k^2(x)dx + \mathcal{A}_2\int_{-1}^{1}k(x)dx\right).$$

To further analyze the bias components let $v_{t,i} = \varepsilon_{t+m}\left(y_{t+1-i} - \mu_y\right)$, $\psi_{t,M} = \left(v'_{t,1}, ..., v'_{t,M}\right)'$ and $V_M = n^{-1/2}\sum_{t=1}^{n-m}\psi_{t,M}$. The two components $\mathcal{A}_1$ and $\mathcal{A}_2$ in (3.5) can be shown to equal

$$\text{(3.7)}\qquad p\mathcal{A}_1\int_{-1}^{1}k^2(x)dx = \lim_{n\to\infty}\frac{\sqrt{n}}{M}E\left[\left(\check{P}_M - P_M\right)' W_M\Omega_M^{-1}W_M V_M\right]$$

where $\check{P}_M$ is the same as $\hat{P}_M$ except that $\bar{y}$ is replaced with μ_y and

$$\text{(3.8)}\qquad p\mathcal{A}_2\int_{-1}^{1}k(x)dx = \lim_{n\to\infty}\frac{\sqrt{n}}{M}E\left[P'_M W_M\Omega_M^{-1}\Omega_M^{\Delta}\Omega_M^{-1}W_M V_M\right]$$

[6] The requirement that $M/n^{1/3} \to 0$ is slightly stronger than $M/n^{1/2} \to 0$ imposed by Donald and Newey (2001). This is without consequence because the optimal M^* is $M^* = O(\log n)$ for the truncated kernel and $M^* = O\left(n^{1/4}\right)$ for kernels with $q \geq 1$ such that $M^*/n^{1/3} \to 0$. The rate $M/n^{1/3} \to 0$ is only used at one point to show that a certain remainder term is small.

where $\Omega_M^{\Delta} = \sum_{l=-m+1}^{m-1} \gamma_l^{\varepsilon} \left(\hat{\Omega}_M(l) - \Omega_M(l)\right) + O_p(M/n)$. These expressions correspond to, yet are somewhat different from, expressions in Donald and Newey (2001) and Newey and Smith (2004). In Donald and Newey (2001) $\mathcal{A}_2 = 0$ because in their case $z_{t,M}$ is exogenous and all expectations are taken conditional on $z_{t,M}$. Here, Ω_M^{Δ} and V_M are correlated because of correlation between ε_s and $z_{t,M}$ and $z_{t,M}$ appearing in both terms. Newey and Smith (2004) consider non-linear GMM estimators with a finite number of moment conditions. Thus, the decomposition given in their Theorem 4.1 applies here with the following modifications: Using their notation it follows that $a = 0$ because ε_t only depends linearly on β and the remaining terms in B_I are of lower order when $M \to \infty$. Similarly B_W is of lower order. The term $\Sigma E\left[G_i \Omega^{-1} g_i\right]$ in B_G corresponds to $D^{-1} p \mathcal{A}_1 \int_{-1}^{1} k^2(x) dx$ while the term $HE\left[g_i g_i' \Omega^{-1} g_i\right]$ corresponds to $D^{-1} p \mathcal{A}_2 \int_{-1}^{1} k(x) dx$. The remaining terms in B_G and B_Ω are of lower order when $M \to \infty$. To better understand the relationship between the two notations note that $\psi_{t,M}$ corresponds to the moment vector g_i and $\partial \psi_{t,M}/\partial \beta'$ corresponds to G_i with $n^{-1} \sum_{t=1}^{n} \partial \psi_{t,M}'/\partial \beta = -\check{P}_M'$. For the truncated kernel, the right hand side of (3.7) then can be written as

(3.9)
$$E\left[\left(\check{P}_M - P_M\right)' \Omega_M^{-1} V_M\right] = E\left[\check{P}_M' \Omega_M^{-1} V_M\right] = n^{-1/2}\left(n^{-1} \sum_{t,s=1}^{n-m} E\left[\frac{\partial \psi_{t,M}'}{\partial \beta} \Omega_M^{-1} \psi_{s,M}\right]\right) + o\left(M/\sqrt{n}\right)$$

which reduces to the term $E\left[G_i \Omega^{-1} g_i\right]$ of Newey and Smith (2004) under independence and when M is fixed. The term B_G is the dominating bias term in GMM estimators but is absent in certain Generalized Empirical Likelihood (GEL) estimators. This is the reason for the smaller bias of GEL compared to GMM as discussed in Donald, Imbens and Newey (2009). However, GEL estimators are typically difficult to implement for time series models.

For (3.8) note that the estimator $\hat{\Omega}_M$ imposes homoskedasticity. As a result, $\hat{\Omega}_M$ does not involve averages $n^{-1} \sum_{t=1}^{n} \psi_{t,M} \psi_{t,M}'$ as is the case for the GMM estimator analyzed by Newey and Smith (2004). The appendix shows that the estimation error related to the estimator of $\gamma^{\varepsilon}(l)$ is of smaller order when $M \to \infty$. Then, (3.8) depends on

$$E\left[P_M' \Omega_M^{-1} \Omega_M^{\Delta} \Omega_M^{-1} V_M\right] = n^{-1/2}\left(P_M' \Omega_M^{-1} n^{-1} \sum_{t,s=1}^{n-m} \sum_{l=-m+1}^{m-1} \gamma^{\varepsilon}(l) E\left[z_{t,M} z_{t-l,M}' \Omega_M^{-1} \psi_{s,M}\right]\right) + o\left(M/\sqrt{n}\right)$$

where, under independence and M fixed, $\gamma^{\varepsilon}(0) E\left[z_{t,M} z_{t,M}' \Omega_M^{-1} \psi_{t,M}\right]$ corresponds to $E\left[g_i g_i' \Omega^{-1} g_i\right]$ aside from the differences due to the imposed homoskedasticity in $\hat{\Omega}_M$.

Anatolyev (2005) provides results for the Bias of GMM estimators with a fixed number of instruments for non-linear time series models. It can be seen from (3.9) that (3.7) corresponds to the term $-\sum_{u=-\infty}^{\infty} \Sigma E\left[m_{\theta t} V^{-1} m_{t-u}\right]$ in Anatolyev (2005, Theorem 1). The remaining terms in $B_{\partial m \Omega m}(u)$, again using Anatolyev's notation, are of smaller order under asymptotics with $M \to \infty$. Similarly, (3.8) corresponds to $\sum_{u,v=-\infty}^{\infty} \Xi E\left[m_t m_{t-u}' V^{-1} m_{t-v}\right]$ with the remaining terms in $B_{m^3}(u)$ being of smaller order. The remaining terms corresponding to Anatolyev (2005, Theorem 1) involving $B_W(u)$, $B_{\partial m \Xi m}$ and $B_{\partial^2 m}$ are either zero or of smaller order for the same reasons as in the comparison with the results of Newey and Smith (2004).

The following result characterizes conditions under which kernel weighted GMM is less biased than standard GMM.

Corollary 3.2. *Fix $\ell \in \mathbb{R}^d$ with $\ell'\ell = 1$. Let $c_1 = \ell' D^{-1} \mathcal{A}_1$ and $c_2 = \ell' D^{-1} \mathcal{A}_2$. Suppose Assumptions B and C in the Appendix hold, $k(.)$ satisfies Assumption A2 and $n, M \to \infty$, $M/n^{1/3} \to 0$. If $k(.)$ satisfies the additional constraint*

$$\left| c_1 \int_{-1}^{1} k^2(x) dx + c_2 \int_{-1}^{1} k(x) dx \right| \leq 2 \left| c_1 + c_2 \right| \tag{3.10}$$

then

$$\lim_{n \to \infty} \left| \sqrt{n}/M E \ell' (b_{n,M} - \beta) \right| \leq \lim_{n \to \infty} \left| \sqrt{n}/M E \ell' (b_{n,M}^T - \beta) \right| \tag{3.11}$$

where $b_{n,M}^T$ is the stochastic approximation of the GMM estimator based on the truncated kernel. If the inequality in (3.10) is strict then the inequality in (3.11) is also strict.

The result of the Corollary is based on the fact that for the truncated kernel $\int_{-1}^{1} k^2(x) dx = \int_{-1}^{1} k(x) dx = 2$ such that the bias in that case is $2pD^{-1}(\mathcal{A}_1 + \mathcal{A}_2)$. Note that Condition (3.10) is satisfied if $\int_{-1}^{1} k^2(x) dx = \int_{-1}^{1} k(x) dx \leq 2$ although much weaker conditions may hold for many values of c_1 and c_2. In particular, if $c_1 > 0$ and $c_2 > 0$ then it is enough to have $\int_{-1}^{1} k^2(x) dx \leq 2$ and $0 < \int_{-1}^{1} k(x) dx \leq 2$ which is satisfied for many standard kernels. For well known kernels such as the Bartlett, Parzen or Tukey-Hanning $\int_{-1}^{1} k(x)^2 dx$ is equal to 2/3, .53 and 3/4 respectively. The linear component $\int_{-1}^{1} k(x) dx$ is equal to 1, 3/4 and 1 for the Bartlett, Parzen and Tukey-Hanning kernel.

Bias properties are only one aspect of estimator performance. By construction, the weight matrix W_M increases the variability of $\hat{\beta}_{n,M}$ relative to the case where $W_M = I_{Mp}$. This happens despite the fact that the first order limiting distribution of $\sqrt{n}\left(\hat{\beta}_{n,M} - \beta_0\right)$ is unaffected by the choice of W_M as long as Assumption A holds and $M \to \infty$. The next result allows to quantify the trade-off between higher order bias and variance of $\hat{\beta}_{n,M}$.

Theorem 3.3. *Suppose Assumptions B and C in the Appendix hold and $\ell \in \mathbb{R}^d$ with $\ell'\ell = 1$ is fixed. Let $\sigma_{1M} = D - P_M' \Omega_M^{-1} P_M = D - D_M$. Assume that $n, M \to \infty$ such that $M/n^{1/3} \to 0$. Then,*
i) for $k(x) = \mathbf{1}\{|x| < 1\}$, let $\mathcal{A}_0 = 2(\mathcal{A}_1 + \mathcal{A}_2)$ and define $\mathcal{A} = \ell' D^{-1} \mathcal{A}_0 \mathcal{A}_0' D^{-1} \ell$,

$$\varphi_n(M, \ell, k(.)) = \frac{(Mp)^2}{n} \mathcal{A} + \ell' D^{-1} \sigma_{1M} D^{-1} \ell, \tag{3.12}$$

ii) for $k(.)$ such that Assumption A2 is satisfied define $\mathcal{A}_0 = \mathcal{A}_1 \int_{-1}^{1} k(x)^2 dx + \mathcal{A}_2 \int_{-1}^{1} k(x) dx$ and $\mathcal{A} = \ell' D^{-1} \mathcal{A}_0 \mathcal{A}_0' D^{-1} \ell$. It follows that

$$\varphi_n(M, \ell, k(.)) = \frac{(Mp)^2}{n} \mathcal{A} + M^{-2q} k_q^2 \ell' D^{-1} \mathcal{B}^{(q)} D^{-1} \ell$$

with $\mathcal{B}^{(q)} = \left(\mathcal{B}_2^{(q)} - \mathcal{B}_1^{(q)} D^{-1} \mathcal{B}_1^{(q)\prime} \right)$ where $\mathcal{B}_2^{(q)}$ is defined as

$$\begin{aligned} \mathcal{B}_2^{(q)} &= \sum_{k=1,j=1}^{\infty} |k|^q |j|^q \Gamma_k^{xy} \vartheta_{k,j} \Gamma_{-j}^{yx} \\ &+ \sum_{j_1,\ldots,j_4=1}^{\infty} \Gamma_{j_1}^{xy} \vartheta_{j_1,j_2} |j_2|^q \omega_{j_2,j_3} |j_3|^q \vartheta_{j_3,j_4} \Gamma_{-j_4}^{yx} + \mathcal{B}_1^{(2q)} \end{aligned} \tag{3.13}$$

and $\mathcal{B}_1^{(q)}$ is defined as $\mathcal{B}_1^{(q)} = \sum_{j_1,j_2=1}^{\infty} \left(\Gamma_{j_1}^{xy} \vartheta_{j_1,j_2} |j_2|^q \Gamma_{-j_2}^{yx} + \Gamma_{j_1}^{xy} |j_1|^q \vartheta_{j_1,j_2} \Gamma_{-j_2}^{yx} \right)$.

Remark 1. *Note that ε_t and the elements of $z_{t,M}$ enter both with up to second powers in $\hat{\Omega}_M$, while they enter as a product in V_M. Consequently, the expansion of $\hat{P}_M' W_M \hat{\Omega}_M^{-1} W_M V_M$ has a mean that to first order involves 8-th moments and a variance that involves 16-th moments. A further analysis shows that $\hat{P}_M' W_M \hat{\Omega}_M^{-1} W_M V_M$ can be decomposed into terms that are stochastically smaller by replacing $\hat{\Omega}_M^{-1}$ with $\Omega_M^{-1} \Omega_M^{\Delta} \Omega_M^{-1}$, or in other words $\hat{\gamma}_j^{\varepsilon}$ with γ_j^{ε}. This reduces the required moments of the highest order terms to 6-th and 12-th moments respectively.*

The higher order MSE $\varphi_n(M, \ell, k(.))$ of $\hat{\beta}_{n,M}$ consists of a bias component that depends on the constant $\mathcal{A}$ and on the kernel as well as a variance term that depends on the constants $\mathcal{B}_1^{(q)}$ and $\mathcal{B}_2^{(q)}$. The terms involving $\mathcal{B}_1^{(q)}$ and $\mathcal{B}_2^{(q)}$ measure the asymptotic discrepancy between the variance of GMM based on a kernel, and the asymptotic variance of GMM based on lag truncation.

Note that $\left(P_M' \Omega_M^{-1} P_M \right)^{-1}$ is the asymptotic variance of $\hat{\beta}_{n,M}$ based on the truncated kernel when M is held fixed as $n \to \infty$, while

$$\Xi_M = \left(P_M' \Omega_{W,M}^{-1} P_M \right)^{-1} \left(P_M' \Omega_{W,M}^{-1} \Omega_M \Omega_{W,M}^{-1} P_M \right) \left(P_M' \Omega_{W,M}^{-1} P_M \right)^{-1}$$

with $\Omega_{W,M}^{-1} = W_M \Omega_M^{-1} W_M$ is the corresponding variance of $\hat{\beta}_{n,M}$ based on a kernel satisfying Assumption A2 for M fixed and $n \to \infty$. The term $M^{-2q} k_q^2 \mathcal{B}^{(q)}$ then can be understood as the approximate, asymptotic difference $P_M'\Omega_M^{-1}P_M - \Xi_M^{-1}$ as $M, n \to \infty$. Because the weighting matrix W_M distorts the optimal weighting matrix Ω_M, the higher order variance of $\hat{\beta}_{n,M}$ also depends on the kernel through the constant k_q. The constant k_q measures the higher order loss in efficiency due to the kernel function. The loss for instrument y_{t+1-i} is proportional to $1 - k(i/M) = k_q M^{-q} |i|^q$ for large M.

For the truncated kernel analyzed in Theorem 3.3i) the difference $P_M'\Omega_M^{-1}P_M - \Xi_M^{-1} = 0$. The next largest variance term then depends on $D - D_M$. The constant D^{-1} is the variance lowerbound for $\hat{\beta}_{n,M}$. This implies that $\sigma_{1M} = D - D_M$ can be interpreted as a measure of how close $\hat{\beta}_{n,M}$ is to achieving the efficiency bound. The approximation to the MSE thus reveals a second order trade-off of the instrument choice between more bias and reduced asymptotic variance. Note that the term $D^{-1}\sigma_{1M}D^{-1}$ is similar to the result in Donald and Newey (2001, Proposition 1) for the two stage least squares estimator in cross-sectional settings.

When kernels satisfying A2 are used the term $M^{-2q}k_q^2\mathcal{B}^{(q)}$ dominates σ_{1M}. This implies that kernels satisfying A2 lead to a higher order MSE that goes to zero at a slower rate than the higher order MSE in case i) of Theorem 3.3. Whether the slower rate of convergence is of practical importance is investigated by Monte Carlo experiments reported in Section 5.

3.2. Special Cases

This section considers a number of special cases of model (2.1) to illustrate the results in Theorem 3.1 and relating them to similar results in the literature.

A special case of interest arises when $m = 1$. Then, ε_t is serially uncorrelated and $f_\varepsilon(\lambda) = \gamma_0^\varepsilon/(2\pi)$. This implies that $\mathcal{A}_1 = (2\gamma_0^\varepsilon)^{-1}\int_{-\pi}^{\pi} f_{\varepsilon x}(\lambda)d\lambda = E[\varepsilon_t x_t]/(2\gamma_0^\varepsilon)$. With a truncated kernel, $\int_{-1}^{1} k^2(x)dx = 2$ such that the first term in (3.6) equals $2pD^{-1}\mathcal{A}_1 = pD^{-1}E[\varepsilon_t x_t]/\gamma_0^\varepsilon$. In addition, $\Omega_M = \gamma_0^\varepsilon \operatorname{Var}(\tilde{z}_{t,M})$ such that $D = H/\gamma_0^\varepsilon$ where $H = \lim_{M\to\infty}(P_M' \operatorname{Var}(\tilde{z}_{t,M}) P_M)$. This reduces the first part of the bias to $pH^{-1}E[\varepsilon_t x_t]$.

Next, assume that $z_{t,M}$ is based on $y_{t,3}$ only and that $y_{t,3}$ is strictly exogenous. If in addition it also holds that $m = 1$, the bias formula can be further simplified. Note that for $\tilde{z}_{t,\infty} = \left[y_{t,3}', y_{t-1,3}', ...\right]'$ it follows that $E[\varepsilon_t x_t] = E[\varepsilon_t v_t]$ where $v_t = x_t - \gamma_0^\varepsilon P'\Omega^{-1}\tilde{z}_{t,\infty}$ with $P = E[x_t\tilde{z}_{t,\infty}]$ and $\Omega = E\left[\varepsilon_{t+1}^2 z_{t,\infty}z_{t,\infty}'\right]$ is the residual from the reduced form equation relating x_t to $z_{t,\infty}$. In addition, because for $m = 1$, $\hat{\Omega}_M = \hat{\gamma}_0^\varepsilon\hat{\Omega}_M(0)$, the term $\hat{\gamma}_0^\varepsilon$ cancels in $\hat{\beta}_{n,M}$. The term Ω_M^Δ in $\mathcal{A}_2$ then only depends on the exogenous $z_{t,M}$ which means that $2p\mathcal{A}_2 = \lim_{n\to\infty} E\left[P_M'W_M\Omega_M^{-1}\Omega_M^\Delta\Omega_M^{-1}W_M E[V_M|Z_M]\right] = 0$. The bias in (3.6) thus reduces to

$$MpH^{-1}E[\varepsilon_t v_t]/\sqrt{n} \tag{3.14}$$

where Mp is the number of instruments. This expression corresponds to the formula given by Donald and Newey (2001) for the 2SLS estimator in a cross-sectional setting.

An even closer analogy to Donald and Newey (2001) can be obtained by imposing further restrictions on $B(L)$ and Σ. To that end, assume that for lag polynomials $\pi_{xz}(L)$ and $\pi_z(L)$ satisfying Assumption C, the structural model and associated reduced form can be written as $y_{1t} = \beta_0' y_{t,2} + u_{t,1}$, $y_{t,2} = \pi_{xz}(L) y_{t,3} + u_{t,2}$ and $\pi_z(L) y_{t,3} = u_{t,3}$ where $u_t = \left(u_{t,1}, u_{t,2}', u_{t,3}'\right)$ is partitioned in accordance with $y_t = \left(y_{t,1}, y_{t,2}', y_{t,3}'\right)$ and $u_{t,3}$ is independent of $u_{1,t}, u_{t,2}$. In this case $B(L)$, partitioned conformingly with $y_t = \left(y_{t,1}, y_{t,2}', y_{t,3}'\right)$, satisfies the constraints

$$B(L) = \begin{bmatrix} 1 & \beta_0' & -\beta_0' \pi_{xz}(L) \pi_z^{-1}(L) \\ 0 & I & -\pi_{xz}(L) \pi_z^{-1}(L) \\ 0 & 0 & \pi_z^{-1}(L) \end{bmatrix}$$

which implies that $\beta_1 = 0$ in (2.1) and $m = 1$. If $z_{t,M}$ is based on $y_{t,3}$ only as before, it follows that $v_t = u_{t,2} = y_{t,2} - \pi_{xz}(L) y_{t,3}$ and the bias formula in (3.14) then is $MpH^{-1}\rho/\sqrt{n}$ where $\rho = cov(u_{t,1}, u_{t,2})$.

It is also of interest to consider the bias approximation for the model used in the Monte Carlo experiments in Section 5. The parameters ϕ and ρ control instrument strength and correlation between reduced from and structural errors and θ is the parameter of the MA polynomial. Detailed calculations discussed in the auxiliary appendix show that $2\mathcal{A}_1 = \rho$ as before while $\mathcal{A}_2$ is of a more complicated form. Contour plots reported in the auxiliary appendix depict the bias as a function of ϕ and ρ and are given for various values of θ. These graphs confirm findings in the Monte Carlo simulations that the bias generally increases with ρ, decreases with ϕ and decreases in θ. In particular, the bias is largest when θ is close to -1 and smallest when θ is close to 1. The auxiliary appendix also contains closed form expressions for the approximate MSE $\varphi_n(M, \ell, k(.))$ in Model 5.1. Manageable closed from expressions are only available when $M = 1$. For that case, the variance component $D^{-1}\sigma_{1M}D^{-1}$ of $\varphi_n(M, \ell, k(.))$ can be analyzed in detail. The exact expressions are again too complex to report here but contour plots of $D^{-1/2}\sigma_{1M}D^{-1/2}$ show that the variance component increases in both ρ and ϕ as well as in the absolute value of θ. One reason for these effects lies in the correlation between the explanatory variable and the instruments at lag j which is

$$\Gamma_j^{xy} = \begin{bmatrix} \phi^{1+j}(1 - \phi\theta)\rho + \phi^{1+j}/\left(1 - \phi^2\right) \\ \phi^{1+j}/\left(1 - \phi^2\right) \end{bmatrix}.$$

Thus, Γ_j^{xy} is an increasing function in both ρ and ϕ. This means that the larger these parameters are, the bigger the discrepancy σ_{1M} between the efficiency bound D and the inverse of the asymptotic variance of the GMM estimator.

3.3. Bias Reducing Kernels

The result in Corollary 3.2 that kernels reduce the higher order bias depends on the constants c_1 and c_2 taking values in a certain range. Corollary 3.2 also implies that kernels satisfying the additional constraint $\int_{-1}^{1} k^2(x)dx = \int_{-1}^{1} k(x)dx$ reduce the bias for all values of the constants c_1 and c_2. Such kernels can be constructed from the class of bias reducing kernels introduced by Bierens (1987). Let

$$k_r(x) = (2\pi)^{-1/2} \sum_{j=1}^{r} a_j \left|\sigma_j\right|^{-1} \exp\left(-1/2x^2/\sigma_j^2\right)$$

and consider the transformation to the interval $[-1,1]$ given by

$$k_{BR}(x) = \; k_r\left(\tan\left(\tfrac{\pi}{2}x\right)\right) \tfrac{\pi}{2} \sec\left(\tfrac{\pi}{2}x\right)^2, \quad x \in [-1,1] \tag{3.15}$$

and $k_{BR}(x) = 0$ for $x \notin [-1,1]$ where $\sec(x) = \cos(x)^{-1}$. For certain choices of σ_j it is possible to solve the system of equations

$$\begin{aligned} \int_{-1}^{1} k_{BR}(x)\, dx &= \int_{-1}^{1} k_{BR}(x)^2\, dx \\ k_{BR}(0) &= 1 \end{aligned}$$

for the parameters a_j. A specific choice of parameters that produces a well behaved kernel is $r = 2$, $\sigma_1 = 1/\sqrt{2}$ and $\sigma_2 = 2/\sqrt{2}$. The parameter values a_j that solve the two integral equations are $a_1 = 1.37621$ and $a_2 = -0.495668$ with a value of $\int_{-1}^{1} k(x)dx = 0.880545$ which is well below 2. For these parameter values, $k_q = 0.40645$ for $q = 2$.

When the constant $c_2 \neq 0$, which typically is the case unless the instruments are strictly exogenous, and if $c_1 \neq 0$, kernel functions can be constructed to eliminate the bias term. To see this let $a = (a_1, ..., a_r)$ be the coefficients of $k_r(x)$ and define K_2 as the matrix with k, j-th element

$$\int_{-1}^{1} \frac{\pi}{8} \sec\left(\frac{\pi}{2}x\right)^4 \left|\sigma_k\right|^{-1} \left|\sigma_j\right|^{-1} \exp\left(-1/2\tan\left(\frac{\pi}{2}x\right)^2 \left(\sigma_k^{-2} + \sigma_j^{-2}\right)\right) dx.$$

The $r \times 1$ vector K_1 is defined similarly with typical element j

$$\int_{-1}^{1} \frac{\sqrt{\pi}}{2\sqrt{2}} \sec\left(\frac{\pi}{2}x\right)^2 \left|\sigma_j\right|^{-1} \exp\left(-1/2\tan\left(\frac{\pi}{2}x\right)^2 \sigma_j^{-2}\right) dx.$$

It then follows that $\int_{-1}^{1} k_{BR}(x)^2\, dx = a'K_2a$ is a quadratic form in a while $\int_{-1}^{1} k_{BR}(x)\, dx = a'K_1$ is linear in a. The approximate bias now is $c_1a'K_2a + c_2a'K_1$. For given choices of σ_j the optimal choice for a is

$$a^* = -\frac{c_2}{c_1} K_2^{-1} K_1.$$

Use the notation $k_{BR}^*(x)$ to denote a kernel with parameters a^*. For example, when $r = 2$,

$$k_{BR}^*(x) = \frac{\sqrt{\pi}}{2\sqrt{2}} \sec\left(\frac{\pi}{2}x\right)^2 \left(\frac{a_1^*}{\sigma_1} \exp\left(-\frac{\tan\left(\frac{\pi x}{2}\right)^2}{2\sigma_1^2}\right) + \frac{a_2^*}{\sigma_2} \exp\left(-\frac{\tan\left(\frac{\pi x}{2}\right)^2}{2\sigma_2^2}\right)\right), \; x \in [-1, 1]. \tag{3.16}$$

A potential disadvantage of $k_{BR}^*(x)$ is that it depends on the unknown nuisance parameter c_2/c_1 which needs to be estimated and may affect the finite sample behavior of an estimator using $k_{BR}^*(x)$.

4. Fully Feasible GMM

Implementation of the estimator $\hat{\beta}_{n,M}$ based on the truncated kernel or a smooth kernel requires a data-dependent choice of M. The approximate higher order mean squared error of $\hat{\beta}_{n,M}$ for the truncated case is given by

$$\varphi_n(M, \ell, k_{TR}(.)) = \frac{(Mp)^2}{n}\mathcal{A} + \ell' D^{-1} \sigma_{1M} D^{-1} \ell. \tag{4.1}$$

The number of moment conditions Mp are selected such that the approximate mean squared error $\varphi_n(M)$ is minimized. Feasible choices of M are such that $M \in I$ where $I = \{[d/p]+1, [d/p]+2, ..., M_{\max}\}$ and $[a]$ denotes the largest integer smaller than a. The set I is constructed to guarantee identification of β for all $M \in I$. It should be noted that I depends on n and $M_{\max} \to \infty$ as $n \to \infty$. The optimal M^* then is defined as

$$M^* = \arg\min_{M \in I} \varphi_n(M, \ell, k_{TR}(.)). \tag{4.2}$$

Note that M^* is a function of n with $M^* \to \infty$ as $n \to \infty$. Implementing any criterion based on $\varphi_n(M, \ell, k_{TR}(.))$ is complicated by the fact that asymptotically the choice of M depends on the rate at which σ_{1M} tends to zero, a parameter that is difficult to estimate. A sieve type approximation to the data-distribution is used to achieve this task. The details of the procedure are laid out in the auxiliary appendix and only a brief description is given here.

A finite order VAR(h) approximation to the infinite order reduced form process of y_t is used to estimate the parameters D_M and D. The approximate model with VAR coefficient matrices $\pi_{1,h}, ..., \pi_{h,h}$ is given by

$$y_t = \mu_{y,h} + \pi_{1,h} y_{t-1} + ... + \pi_{h,h} y_{t-h} + u_{t,h} \tag{4.3}$$

where $\Sigma_h = E\left[u_{t,h} u_{t,h}'\right]$ is the mean squared prediction error of the approximating model. Using a parametric model to approximate the autocovariance function of y_t simplifies the estimation of parameters such as D which depend on the entire autocovariance function of y_t. Kuersteiner (2005) discusses data dependent rules for the selection of h and the asymptotic validity of the procedure used here is based on these results. In particular, the sequential testing procedure proposed in Ng and Perron (1995) is used to select $h \in [h_{\min}, h_{\max}]$ where $\hat{h}$ is used to denote the data-dependent choice of h.[7] Let $\left(\hat{\pi}'_{1,\hat{h}}, \hat{\pi}'_{2,\hat{h}}, ..., \hat{\pi}'_{\hat{h},\hat{h}}\right)$ be the parameter estimates of the approximating VAR and denote by $\hat{\Sigma}_h$ the estimated error covariance matrix of the approximating VAR where $\hat{\Sigma}_h = n^{-1}\sum_{t=h+1}^{n} \hat{u}_{t,h}\hat{u}'_{t,h}$and $\hat{u}_{t,h} = \hat{\pi}_{\hat{h}}(L)\, y_t$ where $\hat{\pi}_{\hat{h}}(z) = I - \hat{\pi}_{1,h}L - ... - \hat{\pi}_{h,h}L^h$. In order to estimate the autocovariance function of y_t note that $\Gamma_j^{yy} = \Gamma(j, \pi(z))$ is a functional of $\pi(z)$ such that the estimator $\hat{\Gamma}_{j,\hat{h}}^{yy}$ is obtained as $\hat{\Gamma}_{j,\hat{h}}^{yy} = \Gamma\left(j, \hat{\pi}_{\hat{h}}(z)\right)$. For numerical evaluation, define the companion matrix

$$\hat{H}_h = \begin{bmatrix} \hat{\pi}_{1,h} & \hat{\pi}_{2,h} & \cdots & \hat{\pi}_{h,h} \\ I_p & 0 & \cdots & 0 \\ 0 & \ddots & & \\ \vdots & & I_p & 0 \end{bmatrix},$$

where $\hat{H}_h$ is of dimension $ph \times ph$ and let $E_h = (I_p, 0, ..., 0)'$ be a $ph \times p$ selector matrix. For a data-dependent choice $\hat{h}$ of h, an approximation to the autocovariance function Γ_j^{yy} is obtained by first computing

$$\operatorname{vec} \hat{G}_{0,\hat{h}}^{yy} = \left(I_{p\hat{h}} - \hat{H}_{\hat{h}} \otimes \hat{H}_{\hat{h}}\right)^{-1} \operatorname{vec} \begin{bmatrix} \hat{\Sigma}_{\hat{h}} & 0 \\ 0 & 0 \end{bmatrix}$$

and

$$\hat{\Gamma}_{j,\hat{h}}^{yy} = E'_{\hat{h}} \hat{G}_{0,\hat{h}}^{yy} \hat{H}_{\hat{h}}^{j} E_{\hat{h}} \tag{4.4}$$

for $0 \leq j \leq k_{\max}$ where $k_{\max} = O\left(\sqrt{n/\log n}\right)$ and $\hat{\Gamma}_{j,\hat{h}}^{yy} = \hat{\Gamma}_{-j,\hat{h}}^{yy\prime}$ for $-k_{\max} \leq j < 0$.[8] The autocovariance matrices Γ_j^{xy} can be estimated by selecting the appropriate elements from $\hat{\Gamma}_{s,\hat{h}}^{yy}$. From these estimates construct the matrix $\hat{P}'_{M,\hat{h}}$.

Let $\tilde{\beta}_{n,M}$ be a consistent first step estimate and obtain estimated residuals $\hat{\varepsilon}_t = y_t - \bar{y} - \tilde{\beta}_{n,M}(x_t - \bar{x})$. The estimated residuals are then used to obtain consistent estimates of the parameter $\theta = (\theta_1, .., \theta_{m-1})'$,

[7] The procedure of Ng and Perron (1995) uses downward testing starting at $h = h_{\max}$. In the implementation used here the search is stopped once h reaches $h_{\min}$. For the simulations and theory in this paper $h_{\max}$ is set to $h_{\max} = (\log n)^2$, and $h_{\min} = ((\log(\log n)) \log n)/10$.

[8] The proof of Theorem 4.1 requires establishing certain rates of convergence which can only be obtained uniformly in $j \leq k_{\max}$. In practice, σ_{1M} is a function of M only with regard to Γ_j^{yy} for $j \leq M_{\max}$. Thus, $k_{\max} = M_{\max}$ is sufficient in practice. The choice of $k_{\max}$ and $M_{\max}$ is expected to only affect higher order asymptotic terms of the estimator. The sensitivity to changes in $k_{\max}$ is investigated in the Monte Carlo section. The results reported in Section 5 use $k_{\max} = M_{\max} = 10\sqrt{n/\log n}$. In the auxiliary appendix Tables 10-12 report results comparing $k_{\max} = 10\sqrt{n/\log n}$ to $k_{\max} = 20\sqrt{n/\log n}$. The differences are negligible.

denoted by $\hat{\theta}$, by fitting a univariate MA(m-1) model to the time series $\hat{\varepsilon}_t$. The proof of Theorem 4.1 only requires that $\hat{\theta} - \theta = O_p\left(n^{-1/2}\right)$. A variety of estimators for the MA(m-1) model, including nonlinear least squares, satisfy this requirement.[9] Let $\hat{f}_{\varepsilon}(\lambda) = \hat{\sigma}_{\varepsilon}^2 \left|\hat{\theta}\left(e^{-i\lambda}\right)\right|^2$ where $\hat{\theta}(L) = 1 - \hat{\theta}_1 L - ... - \hat{\theta}_{m-1} L^{m-1}$ and $\hat{\sigma}_{\varepsilon}^2 = n^{-1} \sum_{t=r+m}^{n} \left(\hat{\varepsilon}_t - \hat{\theta}_1 \hat{\varepsilon}_{t-1} - ... - \hat{\theta}_{m-1} \hat{\varepsilon}_{t-m+1}\right)^2$. Obtain $\hat{\zeta}_j = (2\pi)^{-1} \int_{-\pi}^{\pi} \hat{f}_{\varepsilon}^{-1}(\lambda) e^{i\lambda j} d\lambda$ for all $j \leq k_{\max}$ where $\hat{\zeta}_j$ is used to compute $\widehat{\mathcal{A}}_1$. Denote the estimated autocovariance function of ε_t by $\hat{\gamma}_{\hat{\theta}}^{\varepsilon}(j) = (2\pi)^{-1} \int_{-\pi}^{\pi} \hat{f}_{\varepsilon}(\lambda) e^{i\lambda j}$ for $j = 0, .., m-1$ and $\hat{\gamma}_{\hat{\theta}}^{\varepsilon}(j) = \hat{\gamma}_{\hat{\theta}}^{\varepsilon}(-j)$ for $j < 0$. Define $\hat{\Omega}_{\hat{h}}(l)$ with typical k,j-th block $\hat{\Gamma}_{k-j-l,\hat{h}}^{yy}$, $\hat{\Omega}_{M,\hat{h}}^{*}$ is estimated as $\hat{\Omega}_{M,\hat{h}}^{*} = \sum_{l=-m+1}^{m-1} \hat{\gamma}_{\hat{\theta}}^{\varepsilon}(l) \hat{\Omega}_{\hat{h}}(l)$. The estimator $\hat{\Omega}_{M,\hat{h}}$ of Ω_M is then obtained from (3.2) after replacing $\hat{\Omega}_M^{*}$ with $\hat{\Omega}_{M,\hat{h}}^{*}$, $\hat{\Omega}_M(l)$ with $\hat{\Omega}_{\hat{h}}(l)$ and $\hat{\xi}_{\Omega}$ with $\min \hat{\xi}_{\hat{h}}$, the smallest eigenvalue of $\hat{\Omega}_{M,\hat{h}}^{*}$. An estimate of D_M can be formed as $\hat{D}_{M,\hat{h}} = \hat{P}_{M,\hat{h}}' \hat{\Omega}_{M,\hat{h}}^{-1} \hat{P}_{M,\hat{h}}$. Estimate D by $\hat{D}_{k_{\max},\hat{h}} = \hat{P}_{k_{\max},\hat{h}}' \hat{\Omega}_{k_{\max},\hat{h}}^{-1} \hat{P}_{k_{\max},\hat{h}}$.

The constants $\mathcal{A}_1$ and $\mathcal{A}_2$ are now obtained from

$$\widehat{\mathcal{A}}_1 = \frac{1}{2} \sum_{j=-(n-1)/2}^{(n-1)/2} \hat{\zeta}_j \hat{\Gamma}_j^{\varepsilon x} \tag{4.5}$$

with $\hat{\Gamma}_j^{\varepsilon x} = n^{-1} \sum_{t=\max(j,r)+1}^{\min(n,n+j)} \hat{\varepsilon}_t x_{t-j}$ and, letting $\hat{\Gamma}_j^{\varepsilon y} = n^{-1} \sum_{t=\max(j,r-m)}^{\min(n-m,n+j)} \hat{\varepsilon}_{t+m} y_{t+1-j}$,

$$\widehat{\mathcal{A}}_2 = \frac{1}{2} \sum_{j_1,j_2=1}^{k_{\max}} \hat{\Gamma}_{j_1,\hat{h}}^{xy} \hat{\vartheta}_{j_1 j_2,\hat{h}} \hat{\Gamma}_{-j_2}^{\varepsilon y}. \tag{4.6}$$

The constant $\mathcal{B}^{(q)}$ can be computed in a similar way by replacing population quantities for $\Gamma_{j_1}^{xy}$ and $\vartheta_{j_1 j_2}$ in (3.13) with the estimates proposed here. Feasible optimal bandwidth parameters $\hat{M}^{*}$ for the truncated kernel are then obtained by solving

$$\hat{M}^{*} = \arg\min_{M \in I} \hat{\varphi}_n(M, \ell, k_{TR}(.)) \tag{4.7}$$

for $\hat{\varphi}_n(M, \ell, k_{TR}(.)) = \frac{(Mp)^2}{n} \widehat{\mathcal{A}} + \ell' \hat{D}_{k_{\max},\hat{h}}^{-1} \left(\hat{D}_{k_{\max},\hat{h}} - \hat{D}_{M,\hat{h}}\right) \hat{D}_{k_{\max},\hat{h}}^{-1} \ell$, where $\widehat{\mathcal{A}}$ is the plug-in estimator using $\widehat{\mathcal{A}}_1, \widehat{\mathcal{A}}_2$ and $\hat{D}_{k_{\max},\hat{h}}$. The following theorem establishes the sense in which $\hat{\varphi}_n$ approximates φ_n and $\hat{M}^{*}$ approximates M^{*} defined in (4.2).

Theorem 4.1. *Suppose Assumptions B, C and D hold. In addition assume that* $E(u_t | u_{t-1}, ...) = 0$ *and* $\hat{\theta} - \theta = O_p\left(n^{-1/2}\right)$, $k_{\max} = O\left(\sqrt{n / \log n}\right)$, $h_{\min} = c_{\min}((\log(\log n)) \log n)$ *for some* $0 < c_{\min} < \infty$, $h_{\max} = O\left((\log n)^2\right)$ *and* $\nu^{-M} \ell' D^{-1} \sigma_{1M} D^{-1} \ell \geq \varepsilon > 0$ *for some* ε *and all* M. *Assume there exists a twice differentiable function* $g(M)$ *such that* $\lim_M \ell' D^{-1} \sigma_{1M} D^{-1} \ell / g(M) = 1$ *and* $\liminf_M M^2 \partial^2 g(M) / (\partial M)^2 / g(M) > 0$. *Let* $\hat{M}^{*}$ *be as defined in (4.7) and* M^{*} *as defined in (4.2). Then, for any* $\delta > 0$, $M_{\max} = O\left(n^{\frac{1}{2\tau}} (\log n)^{-\frac{5+\delta}{2\tau}}\right)$ *and* $\tau \geq 3$

$$\frac{\hat{\varphi}_n(\hat{M}^{*}, \ell, k_{TR}(.))}{\varphi_n(M^{*}, \ell, k_{TR}(.))} - 1 = o_p(1) \tag{4.8}$$

[9] In the Monte Carlo simulations the parameter θ is estimated using the MATLAB routine 'armax'.

and $\left(\hat{M}^*/M^*-1\right)=o_p(1)$.

Remark 2. *The additional condition that $E\left(u_t|u_{t-1},...\right)=0$, i.e. that u_t is a martingale difference sequence (mds), is needed because the proof of Theorem 4.1 uses results established in Hannan and Deistler (1988) and Kuersteiner (2005) that rely on this assumption. Relaxing the mds condition may be possible but is beyond the scope of this paper.*

Remark 3. *A possible alternative estimator for φ_n is one where population moments are replaced by sample analogs. However, such an estimator does not have desirable properties. The reason is that $\sup_j \left\|\hat{\Gamma}_j^{yy}-\Gamma_j^{yy}\right\|=O_p\left(\sqrt{\log n/n}\right)$ is the best possible uniform rate at which autocovariances can be estimated. In order to achieve (4.8) one needs to establish*

$$|\hat{\varphi}_n(M,.,.))-\varphi_n(M,.,.)|=\varphi_n(M,.,.)o_p\left(1\right) \tag{4.9}$$

uniformly in $M\leq M_{\max}$. Since the left hand side cannot be shown to be smaller than $O_p\left(\sqrt{\log n/n}\right)$ while at the same time $\varphi_n(M,.,.)=O\left(\nu^M\right)$, it follows that $M_{\max}$ can be at most $M_{\max}=o\left(\log n\right)$ to achieve the necessary uniform convergence in (4.9). Such a slow rate for $M_{\max}$ is not sufficient to guarantee that $M^*\in I$. On the other hand, the approximation of Γ_j^{yy} by a parametric model proposed here implicitly imposes additional smoothness properties of the population spectral density of y_t on the estimated $\hat{\Gamma}_j^{yy}$ in Assumption D. The smoothness of the estimated spectrum is the key to establishing (4.8). It should be noted that Assumption D is not required for data-dependent moment selection of the kernel weighted GMM estimator in (4.11). Thus, feasible kernel weighted GMM is justified under weaker assumptions, giving it a further advantage over standard GMM.

Remark 4. *The restrictions imposed on the distribution of u_t, in particular homoskedasticity are needed to express the elements of Ω^{-1} in terms of parameters of an approximating VAR (see Lemma A.7). Such an approximation is not possible in a more general setting with heteroskedastic errors. To see this let $v_{t,i}=\varepsilon_{t+m}\left(y_{t+1-i}-\mu_y\right)'$ and $\psi_{t,M}=\left(v_{t,1}',...,v_{t,M}'\right)'$ such that even when imposing the additional assumption that u_t is an mds sequence, $\Omega_M=\sum_{l=-m+1}^{m-1}E\left[\psi_{t-l,M}\psi_{t,M}'\right]$. Without homoskedasticity, the j,k-th block of Ω_M, $E\left[v_{t,j}v_{t,k}'\right]$, is essentially unrestricted. Because of Remark 3 it appears impossible to establish sufficiently fast rates of convergence for an estimator of Ω_M under these more general conditions.*

One advantage of using a smooth kernel is that it leads to a closed form expression for the optimal bandwidth M. The optimal bandwidth for smooth kernels is given by

$$M_k^*=\left(\frac{nqk_q^2\ell D^{-1}\mathcal{B}^{(q)}D^{-1}\ell}{p^2\mathcal{A}}\right)^{\frac{1}{2+2q}}. \tag{4.10}$$

The empirical counterpart of M_k^* is obtained by plugging in estimators for $\mathcal{A}$ and $\mathcal{B}^{(q)}$ into formula (4.10) such that

$$\hat{M}_k^* = \left(\frac{nqk_q^2 \ell' \hat{D}^{-1}_{k_{\max},\hat{h}} \widehat{\mathcal{B}}^{(q)} \hat{D}^{-1}_{k_{\max},\hat{h}} \ell}{p^2 \widehat{\mathcal{A}}} \right)^{\frac{1}{2+2q}}. \tag{4.11}$$

Expression 4.10 depends on constants $\mathcal{A}$ and $\mathcal{B}^{(q)}$ that are easier to estimate than the term σ_{1M} which needs to be evaluated uniformly in M over some range of permissible values for M. Consequently, establishing that $\hat{M}_k^*/M_k^* - 1 = o_p(1)$ is much simpler than proving the result in Theorem (4.1). All that is required is that $\widehat{\mathcal{A}} - \mathcal{A} = o_p(1)$ and $\widehat{\mathcal{B}}^{(q)} - \mathcal{B}^{(q)} = o_p(1)$. This can be achieved with a variety of estimators, including the spectral density estimators discussed in Andrews (1991) or Newey and West (1994).

5. Monte Carlo Simulations

A small Monte Carlo experiment is conducted to assess the performance of the proposed moment selection methods. For the simulations the following data generating process is used

$$\begin{aligned} y_{t,1} &= \beta y_{t,2} + u_{t,1} - \theta u_{t-1,1} \\ y_{t,2} &= \phi y_{t-1,2} + u_{t,2} \end{aligned} \tag{5.1}$$

with $u_t = (u_{t,1}, u_{t,2}) \sim N(0, \Sigma)$ where Σ has elements $\sigma_1^2 = \sigma_2^2 = 1$ and ρ. The parameter β is the parameter to be estimated and is set to $\beta = 1$ in all simulations. All remaining parameters are nuisance parameters not explicitly estimated. The parameter ρ is one of the determinants of the small sample bias of both Ordinary Least Squares (OLS) and GMM estimators and is varied over $\rho = \{.1, .5, .9\}$. The parameter ϕ controls the quality of lagged instruments and is chosen in $\{.1, .3, .5\}$. Low values of ϕ imply that the model is poorly identified. The parameter θ finally is set to $\{-.9, -.5, 0, .5, .9\}$.

Samples of size $n = \{128, 512\}$ from Model (5.1) are generated. Starting values are $y_0 = 0$ and $u_0 = 0$. In each sample the first 1,000 observations are discarded to eliminate dependence on initial conditions. The simulations are based on 1,000 replications.

In order to estimate Ω_M an inefficient but consistent estimate $\tilde{\beta}_{n,\tilde{M}}$ based on (3.1) setting $W_{\tilde{M}} = I_{\tilde{M}p}$ and $\Omega_{\tilde{M}} = I_{\tilde{M}p}$ is used. The parameter $\tilde{M}$ is selected as the number of lags chosen for the approximate VAR in (4.3) by the sequential testing procedure proposed in Ng and Perron (1995).[10] Then construct residuals $\tilde{\varepsilon}_t = y_{1t} - \tilde{\beta}_{n,M} y_{2t}$ and estimate $\tilde{\Omega}_{\tilde{M}}$ as described in (3.2). The initial estimator $\tilde{\beta}_{n,\tilde{M}}$ then

[10] This choice of M attempts to capture the most relevant lags for the implicit first stage used in $\tilde{\beta}_{n,M}$. It is similar in spirit to the use of the first stage cross-validation criterion in Donald and Newey (2001, p.1173).

is reestimated using $\tilde{\Omega}_{\tilde{M}}$ as weight matrix. Residuals used in subsequent steps are based on the reestimated first step estimator $\tilde{\beta}_{n,M}$, leading to an estimate $\hat{\Omega}_M$ where M now is a parameter to be determined by the automatic procedures described in the previous section. In the second stage apply (3.1) with $\hat{\Omega}_M$ and the matrix W_M for the Bartlett, Tukey-Hanning as well as the bias reducing kernel developed in (3.15). Feasible conventional GMM with data-dependent moment selection is implemented using the selection method described in (4.7). The estimated number of instruments $\hat{M}^*$ is then substituted in the estimator (3.1) by setting $W_{\hat{M}^*} = I_{\hat{M}^* p}$. Feasible kernel weighted estimators are constructed in the same way with $\hat{M}^*$ selected based on (4.11) and substituted into Formula (3.1) where $W_{\hat{M}^*}$ now is a matrix with corresponding kernel weights.[11]

The results of the Monte Carlo experiment are reported in Tables 1a-3c in the appendix. For each estimator the median bias, decile range, mean squared error (MSE), mean absolute error (MAE), the size of a two sided t-test of $H_0 : \beta = 1$ at nominal level 5% (Size) and the median number of instruments or the median bandwidth for kernel based estimators is reported. OLS stands for the ordinary least squares estimator, GMM-1 for the standard GMM estimator with one lag of the full set of instruments $\tilde{z}_{t,1} = (y_{t,1}, y_{t,2})$, GMM-25 is the standard GMM estimator with 25 lags of instruments, such that $\tilde{z}_{t,25} = (y_t, ..., y_{t-24})$ and $y_t = (y_{t,1}, y_{t,2})$. GMM-Tuk-Han is the kernel weighted GMM estimator based on the Tukey-Hanning and GMM-BR is based on the bias reducing kernel introduced in (3.15) where in both cases $\hat{M}^*$ is selected based on (4.11).[12] GMM-Trunc is the standard GMM estimator with data-dependent number of instrument selection as defined in (4.7). CUE-1 and CUE-25 are the continuous updating estimators of Hansen, Heaton and Yaron (1996) based on one and 25 lagged instruments respectively. Finally, WWA is the estimator proposed by West, Wong and Anatolyev (2009).[13]

To gain some insight into how estimator bias depends on the parameters ϕ, θ and ρ, it is useful to consider the approximate large sample bias of the OLS estimator. When identification is weak GMM can be expected to be as biased as the OLS estimator. Simple calculations show that asymptotically, OLS tends to $\left(1 - \phi^2\right)(1 - \phi\theta)\rho$. Thus, the OLS bias is increasing in ρ and is decreasing both in ϕ and

[11] A more detailed step-by-step description of the estimators is given in Section 7 of the Auxiliary Appendix.

[12] Tables in the Auxiliary Appendix report additional results for the Bartlett kernel, the Bias minimal kernel (3.16) and a bias corrected version of the truncated kernel developed in the Auxiliary Appendix. To save space these results are not discussed in detail. The Bartlett has similar properties as the Tukey-Hanning but performs generally a bit less well, especially when identification is weak. Bias correction methods reduce bias, in particular when the model is well identified. However, the associated increase in variability often off-sets these gains and the MAE seldomly improves.

[13] The results reported for WWA are based on the same implementation of the estimator as the one considered in West, Wong and Anatolyev (2009). Their implementation estimates a heteroskedasticity robust weight matrix to construct the instruments. In the Auxiliary Appendix an additional alternative version of WWA that imposes homoskedasticity is considered for the Monte Carlo designs with homoskedastic errors. At least for the MC designs where weak identification is a prevalent feature, the performance of WWA is not significantly affected by imposing homoskedasticity.

θ over the range of parameter values considered here. This is consistent with the results in the Monte Carlo simulations.

Table 1a-1c contains results for the case when $\phi = .1$. This constitutes a situation where the instruments are weak, especially when the sample size is $n = 128$. As is well known from the cross-sectional literature (c.f. Hahn, Hausman and Kuersteiner, 2004), the Nagar approximation to the MSE of the 2SLS estimator may become unreliable under these circumstances. One way in which this manifests itself is in the fact that estimator bias does not increase with the number of instruments as predicted by (3.12). This feature, most notably visible by comparing GMM-1 and GMM-25, can be observed for all values of ρ and $n = 128$. A larger value of ρ implies that the endogeneity problem, and thus the bias of OLS, is more severe. When instruments are weak, a higher or lower value of ρ does however not seem to affect the bias trade-off of adding more instruments. As a consequence, moment selection approaches which essentially are based on the idea of optimizing this trade-off do not work that well. Their MSE is often higher than the MSE of a GMM estimator with a large, but fixed number of instruments such as GMM-25. These findings are analogous to findings for the cross-sectional case in Donald and Newey (2001). Amongst the data-dependent methods, KGMM with the Tukey-Hanning kernel performs best while GMM-Trunc often performs least well, especially in terms of the MSE and MAE criteria.

Similar results obtain for the case of the large sample with $n = 512$ in Tables 1a-1c. GMM-Trunc continues to perform poorly compared to the other data-dependent methods. GMM-BR based on the bias reducing kernel now does quite well, sometimes performing at par with the Tukey-Hanning kernel.

CUE and more so WWA are generally less biased than GMM based estimators, although their bias is still significant when $\rho \geq .5$. While the bias properties of CUE and WWA are more favorable, they show more dispersion, as measured by the decile range and MAE, than the best data-dependent GMM estimators in most designs. The CUE performs particularly poorly in this regard. Size distortions are generally mild when $\rho = .1$ except for WWA and CUE-25 when $|\theta| \geq .5$. With increasing values of ρ the size distortions of GMM based tests become very severe. WWA is a bit less sensitive in this regard, however size is heavily distorted for this estimator as well. The best performance is achieved by GMM-1 and CUE-1. At least for $\rho = .5$ the size of tests based on these estimators remains relatively accurate while for $\rho = .9$ it is distorted but less so than for GMM based tests.

In Tables 2a-2c the case where $\phi = .3$ is considered. Here, the instruments are more informative about the parameter β. Nevertheless, when $n = 128$ the asymptotic approximation to the MSE overestimates the effect more instruments have on the bias. Consequently, GMM-25 continues to do well compared to the data-dependent procedures when $\rho = .1$ and $\rho = .5$. When $\rho = .9$ and $\theta \leq 0$ the

Tukey-Hanning kernel outperforms GMM-25. When $n = 512$ the asymptotic approximation is more relevant and better captures the bias variance trade-off. The Tukey-Hanning kernel continues to perform best in terms of MAE amongst the data-dependent procedures. When $n = 512$ in Table 2b and more so in Table 2c data-dependent procedures dominate the fixed instrument estimators GMM-1 and GMM-25. When $\theta \leq 0$ this dominance in terms of MSE and MAE criteria is quite pronounced. CUE and WWA continue to have similar properties as when $\phi = .1$. They are generally less biased than GMM based procedures but have more variation. The net result on MAE is that in most cases the best data-dependent GMM based estimator dominates. The only exception are the results on Table 2c with $n = 512$ where bias reduction pays off and identification is sufficiently strong for WWA to work well. However, even in this case, data-dependent GMM procedures do quite well too and the difference to WWA in terms of MAE is not that large. The size properties of a t-test are similar to the ones reported in Tables 1a-1c. GMM-1 and CUE-1 are overall the best procedures. WWA has some size problems when $\rho = .1$ and θ is large, while data-dependent GMM generally works well for $\rho = .1$. With increasing correlation, size deteriorates dramatically. However, the data-dependent GMM procedures perform generally much better than GMM-25 and CUE-25.

In Tables 3a-3c finally, the case where $\phi = .5$ is considered. For this parameter value, even samples with $n = 128$ are quite informative about the parameter β and the asymptotic approximation to the MSE is more accurate. The Tukey-Hanning kernel dominates the other data-dependent procedures in terms of MAE in many cases in Tables 3a-3c and for $n = 128$. When $n = 512$, WWA is best when $\rho = .5$ or $\rho = .9$. WWA is generally less biased than the other estimators which, combined with strong identification, explains the good performance in Tables 3b and 3c and $n = 512$. As before, the data-dependent methods perform significantly better than GMM-25 when $\rho \geq .5$. The results for the size of t-tests remain similar to the previous cases. GMM-1 and CUE-1 are generally most reliable across the range of different designs, while GMM-25 and CUE-25 are worst. The data-dependent GMM methods and WWA partly correct for these large size distortions but do not always completely remove them.

Overall, the results demonstrate the advantages of data-dependent moment selection in models with moderate to severe endogeneity and sufficiently strong instruments to validate the asymptotic approximations to the MSE. Amongst the data-dependent procedures the Tukey-Hanning kernel shows the best overall performance. The standard GMM estimator with data-dependent instrument selection generally performs less well than the kernel weighted GMM estimators. CUE and WWA do well in terms of bias but often have inflated variances. In some designs the increase in variability is severe. WWA does not perform well in terms of MAE when identification is weak. In Tables 3b and 3c, with strong identification and large enough samples, it does well as long as $\theta \leq .5$. However, in those circumstances the data-dependent GMM estimators are not far behind. When the goal of the analysis is to have tests with accurate size, the preferred choice is to use CUE-1 or GMM-1. It is expected however, that this choice will come at the cost of reduced power. GMM-25 and CUE-25 cannot be recommended in general, both from a testing and estimation point of view. Tests based on these estimators can have very severe size distortions. GMM-25 can be severely biased while CUE-25 suffers from large small sample variability.

6. Conclusions

The higher order asymptotic properties of GMM estimators for time series models with many instruments are analyzed. Using expressions for the asymptotic mean squared error a selection rule for the optimal number of lagged instruments is derived. Fully feasible GMM estimators where the number of instruments are based on a data-dependent selection rule are developed and investigated in a Monte Carlo study.

A new version of the GMM estimator for linear time series models is proposed where the moment conditions are weighted by a kernel function. It is shown that suitably chosen kernel weights of the moment restrictions reduce the asymptotic bias. A fully automatic procedure to chose the number of instruments through an automated bandwidth choice is developed. Monte Carlo experiments demonstrate the advantages of kernel weighting relative to conventional GMM.

Instrument choice is based on the assumption of homoskedastic errors. If the data are heteroskedastic the selected number of instruments will in general be suboptimal. Monte Carlo experiments as well as the fact that the choice of the number of instruments is based on higher order properties of the estimators suggest that the homoskedasticity assumption may lead to reasonable simplifications in practice. Whether this statement can be established theoretically remains an open question that would require the development of selection procedures under heteroskedasticity. The latter is a topic for future research.

A. Proofs

A.1. Definitions

Definition A.1. *Let* $u_t \in \mathbb{R}^p$ *be a strictly stationary vector process with elements* u_t^i *such that* $E\left[u_t^i\right] = 0$ *and* $E\left[\left(u_t^i\right)^k\right] < \infty$. *Let* $\varsigma = (\varsigma_1, ..., \varsigma_k) \in \mathbb{R}^k$ *and* $u = (u_{t_1}^{i_1}, ..., u_{t_k}^{i_k})$ *then* $\phi_{i_1,...i_k,t_1,...,t_k}(\varsigma) = E\left[e^{i\varsigma' u}\right]$ *is the joint characteristic function of* u. *The joint* k*-th order cumulant is*

$$\mathrm{cum}^*_{i_1,...,i_k}(t_1, ..., t_k) = \frac{\partial^k}{\partial\varsigma_1 \cdots \partial\varsigma_k}|_{\varsigma=0} \ln \phi_{i_1,...i_k,t_1,...,t_k}(\varsigma).$$

Alternatively the notation $\mathrm{cum}^*(u_{t_1}^{i_1}, ..., u_{t_k}^{i_k})$ *is used where more convenient. By stationarity it is enough to define*

$$\mathrm{cum}_{i_1,...,i_k}(t_1, ..., t_{k-1}) = \mathrm{cum}^*_{i_1,...,i_k}(t_1, ..., t_{k-1}, 0).$$

Definition A.2. *Let* $\mu_x = E[x_t]$. *Define* $w_{t,i} = (x_{t+m} - \mu_x)\left(y_{t-i+1} - \mu_y\right)'$, $\Gamma_i^{xy} = E[w_{t,i}]$ *and* $\Gamma_{-i}^{yx} = E\left[w_{t,i}'\right]$ *and let* $\breve{w}_{t,i} = w_{t,i} - \Gamma_i^{xy}$. *Next define* $w_{t,j}^y = \left(y_t - \mu_y\right)\left(y_{t-j} - \mu_y\right)'$ *with* $E\left[w_{t,j}^y\right] = \Gamma_j^{yy}$. *Let* $\breve{w}_{t,j}^y = w_{t,j}^y - \Gamma_j^{yy}$. *Define* $v_{t,i} = \varepsilon_{t+m}(y_{t-i+1} - \mu_y)$ *and* $E[\varepsilon_{t+m} y_{s+1}] = \Gamma_{t-s}^{\varepsilon y}$. *Define the infinite dimensional instrument vector* $\tilde{z}_{t,\infty} = (y_t', y_{t-1}', ...)'$ *and let* $P' = \mathrm{Cov}(x_{t+m}, \tilde{z}_{t,\infty})'$. *Define the infinite dimensional matrix* $\Omega = \sum_{l=-m+1}^{m-1} \gamma_l^\varepsilon \,\mathrm{Cov}\left[\tilde{z}_{t,\infty}, \tilde{z}_{t+l,\infty}\right]$ *by its typical* j,k*-th block* $\omega_{j,k}$ *where*

$$\omega_{j,k} = \textstyle\sum_{l=-m+1}^{m-1} \gamma^\varepsilon(l)\, \Gamma_{k-j-l}^{yy}.$$

Denote by $\vartheta_{j,k}$ *and* $\vartheta_{j,k}^M$ *the* j,k*-th block of* Ω^{-1} *and* Ω_M^{-1}. *Let* $D = P'\Omega^{-1}P$ *and* $d_0 = P'\Omega^{-1}V$ *where* $V = n^{-1/2}\sum_{t=1}^n \varepsilon_{t+m}(z_{t,\infty} - \mathbf{1}_\infty \otimes \mu_y)$.

Definition A.3. *Let* $f_\Omega(\lambda) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\sum_{l=-m+1}^{m-1} \gamma_l^\varepsilon \Gamma_{j-l}^{yy} e^{-i\lambda j}$ *which can be represented as* $f_\Omega(\lambda) = 2\pi f_\varepsilon(\lambda) f_y(\lambda)$ *where* $f_y(\lambda) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty} \Gamma_j^{yy} e^{-i\lambda j}$.

Definition A.4. *For a matrix* A, $\|A\|^2 = \mathrm{tr}\, AA'$. *The matrix norm* $\|A\|_2^2$ *is given by*

$$\|A\|_2^2 = \sup_{x \neq 0} x'A'Ax/x'x.$$

The $p^2 \times p^2$ *commutation matrix* $K_{pp} = \sum_{i,j=1}^p e_i e_j' \otimes e_j e_i'$ *where* $\otimes$ *is the Kronecker product and* e_i *is the* i*-th unit* p*-vector; see Magnus and Neudecker (1979).*

A.2. Assumptions

In addition to the structural restrictions of Equation (2.1) the following formal assumptions on u_t and $B(L)$ are imposed.

Assumption B. *Let $u_t \in \mathbb{R}^p$ be strictly stationary and ergodic, with $E[u_t] = 0$, $E(u_t u_t' | \tilde{z}_{t-1,M}) = \Sigma$ for some positive definite and nonrandom matrix Σ and $E(u_t u_s' | \tilde{z}_{t-1,M}) = 0$ for $t > s$. Let u_t^i be the i-th element of u_t and $\operatorname{cum}_{i_1,...,i_k}(t_1, ..., t_{k-1})$ the k-th order cross cumulant of $u_{t_1}^{i_1}, ..., u_{t_k}^{i_k}$. Assume that*

$$\sum_{t_1=-\infty}^{\infty} \cdots \sum_{t_{k-1}=-\infty}^{\infty} |\operatorname{cum}_{i_1,...,i_k}(t_1, ..., t_{k-1})| < \infty \text{ for } k \leq 12.$$

Assumption C. *The lag polynomial $B(z)$ with coefficient matrices B_j satisfies $\det B(z) \neq 0$ for $|z| \leq 1$. Define $B(z)^{-1} = \pi(z) = I - \sum_{j=1}^{\infty} \pi_j z^j$. Moreover, let $b(z) = \alpha(z, \beta) B(z)$ and assume that $b(z) = \sum_{j=0}^{m-1} b_j z^j$ with $b(z) \neq 0$ for $|z| \leq 1$. Let $f_\varepsilon(\lambda) = (2\pi)^{-1} b(e^{i\lambda})' \Sigma b(e^{-i\lambda})$ and assume that there exists a constant σ_ε^2 and lag polynomial $\theta(z) = 1 - \theta_1 z - ... - \theta_{m-1} z^{m-1}$ such that $f_\varepsilon(\lambda)$ can be represented as $f_\varepsilon(\lambda) = (2\pi)^{-1} \sigma_\varepsilon^2 \left|\theta(e^{i\lambda})\right|^2$. Let $\theta(z)^{-1} = \sum_{j=0}^{\infty} \zeta_j^\theta z^j$. For some ν with $\nu \in (0,1)$ and some generic constant C, assume that $\sum_{j=k}^{\infty} \|\pi_j\| \leq C\nu^k$, $\sum_{j=k}^{\infty} \zeta_j^\theta \leq C\nu^k$ and $\sum_{j=k}^{\infty} \|B_j\| \leq C\nu^k$ uniformly in $k = 1, 2, ...$. Assume that P has full column rank.*

Assumption D. *Assume that*

$$\left\|\Gamma\left(j, \hat{\pi}_{\hat{h}}(z)\right) - \Gamma(j, \pi(z))\right\| = j\nu^j O_p\left(\sup_{|z|\leq 1} \left\|\hat{\pi}_{\hat{h}}(z) - \pi(z)\right\|\right) + j^3 \nu_*^j O_p\left(\sup_{|z|\leq 1} \left\|\hat{\pi}_{\hat{h}}(z) - \pi(z)\right\|^3\right)$$

uniformly in j for $\nu < \nu_* < 1$. Let $K_{\hat{h}}^y\left(e^{i\lambda}\right)$ be the spectral density with Fourier coefficients $\Gamma\left(j, \hat{\pi}_{\hat{h}}(z)\right)$ and assume that $K_{\hat{h}}^y\left(e^{i\lambda}\right) = \tilde{K}_{\hat{h}}^y\left(e^{i\lambda}\right) \tilde{K}_{\hat{h}}^y\left(e^{-i\lambda}\right)'$ where $\tilde{K}_{\hat{h}}^y(z)$ is a matrix valued (infinite order) polynomial in z.

Remark 5. *The column rank assumption for P is needed for identification (see Kuersteiner (2001) for an extensive discussion of this point). Assumption C guarantees that $f_\varepsilon(\lambda) \neq 0$ for $\lambda \in [-\pi, \pi]$. Then $1/f_\varepsilon(\lambda)$ exists and corresponds to the spectral density of an AR(m-1) model.*

A.3. Auxiliary Lemmas

Proofs for the following Lemmas are given in the Auxiliary Appendix.

Lemma A.5. *Let X, Y, W, Z be random matrices with all elements $x_{ij}, y_{ij}, w_{ij}, z_{ij}$ such that $E[x_{ij}] = ... = E[z_{ij}] = 0$ and $E\left[|x_{ij}|^4\right] < \infty, ..., E\left[|z_{ij}|^4\right] < \infty$. Let A, B, C and D be matrices with fixed coefficients such that the matrix products $CWAXY'BZD$ and DC are well defined. Then*

$$\begin{aligned} E\left[\operatorname{tr} CWAXY'BZD\right] &= \left(\operatorname{vec} B'\right)' E\left[Y \otimes Z\right] (I \otimes DC) E\left[X' \otimes W\right] \operatorname{vec} A \\ &\quad + \operatorname{tr}\left(E\left[D'Z' \otimes AX\right] E\left[\operatorname{vec}(Y'B) \operatorname{vec}(W'C')'\right]\right) \\ &\quad + \operatorname{tr}\left[(E\left[AXY'B\right])(E\left[ZDCW\right])\right] + \mathcal{K}_4 \end{aligned}$$

where $\mathcal{K}_4 = \sum_k \sum_{j_1,...,j_7} ... \sum a_{j_2,j_3} b_{j_5,j_6} c_{k,j_1} d_{j_7,k} \operatorname{cum}^*(w_{j_1,j_2}, x_{j_3,j_4}, y_{j_4,j_5}, z_{j_6,j_7})$.

Lemma A.6. *If $v_{t,i} = \varepsilon_{t+m}(y_{t-i+1} - \mu_y)$ and $w_{t,i} = (x_{t+m} - \mu_x)\left(y_{t-i+1} - \mu_y\right)'$ and $\ell \in \mathbb{R}^d$ is a vector of constants such that $\ell'\ell = 1$ then*
i) $E\left[v_{t,i} \otimes \breve{w}'_{s,j}\ell\right] = ((\mathrm{vec}(\Gamma^{yy}_{s-t+i-j}) \otimes \left(\Gamma^{\varepsilon x}_{t-s}\right)') + K_{pp}(\Gamma^{\varepsilon y}_{t-s+j} \otimes \Gamma^{yx}_{t-i-s}) + \mathcal{K}^1_4)(I \otimes \ell)$ *where $\mathcal{K}^1_4$ is a $p^2 \times d$ matrix with typical element (a,b) equal to*

$$\left[\mathcal{K}^1_4\right]_{a,b} = \mathrm{cum}^*(\varepsilon_{t+m}, y^{[(a-1)/p]+1}_{t-i+1}, y^{a \bmod (p-1)}_{s-j+1}, x^b_{s+m}),$$

where $[a]$ is the largest integer smaller than a, $a \bmod p$ is the remainder on division of a by p, and K_{pp} is defined in (A.4),
ii) $E\left[v_{t,i}\ell' w_{s,j}\right] = (\ell'\Gamma^{\varepsilon x}_{t-s})\Gamma^{yy}_{t-s+j-i} + (\ell'\Gamma^{xy}_{s-t+i})'\Gamma^{\varepsilon y\prime}_{t-s+j} + \mathcal{K}^2_4$ *where $\mathcal{K}^2_4$ is a $p \times p$ matrix with typical element (a,b)*

$$\left[\mathcal{K}^2_4\right]_{a,b} = \mathrm{cum}^*(\varepsilon_{t+m}, y^b_{t-i+1}, y^a_{s-j+1}, \ell' x_{s+m}),$$

iii) $E\left[v_{t,i}v'_{s,j}\right] = \gamma^{\varepsilon}_{t-s}\Gamma^{yy}_{t-i+j-s}$,
iv) $E\left[w_{t,i}w'_{s,j}\right] = \Gamma^{xy}_i\Gamma^{yx}_{-j} + \gamma^{yy}_{t-i+j-s}\Gamma^{xx}_{t-s} + \Gamma^{xy}_{t-i-s}\Gamma^{yx}_{t-s+j} + \mathcal{K}^4_4$ *where $\mathcal{K}^4_4$ is a $p \times p$ matrix with typical element (a,b)*

$$\left[\mathcal{K}^4_4(t,s,i,j)\right]_{a,b} = \sum_l \mathrm{cum}^*(x^a_{s+m}, x^b_{t+m}, y^l_{t-i+1}, y^l_{s-j+1})$$

and $\gamma^{yy}_{s-t} = E\left[(y_{t-j} - \mu_y)'(y_{s-j} - \mu_y)\right]$.

Lemma A.7. *Let $\theta(L)$ be as defined in Assumption C and $f_\varepsilon(\lambda) = (2\pi)^{-1}\sigma^2_\varepsilon\left|\theta(e^{i\lambda})\right|^2$. Define $\tilde{\pi}(L) = \theta(L)^{-1}\pi(L)$ where $\tilde{\pi}(L) = \sum_{j=0}^\infty \tilde{\pi}_j L^j$. Then it follows that the j,k-th block element $\vartheta_{j,k}$ of Ω^{-1}, where Ω is defined in Definition A.2, is given by*

$$\vartheta_{j,k} = \sigma_\varepsilon^{-2}\sum_{l=0}^{j-1}\tilde{\pi}'_l\Sigma^{-1}\tilde{\pi}_{l+k-j} = \sigma_\varepsilon^{-2}\sum_{l=0}^{k-1}\tilde{\pi}'_{l+j-k}\Sigma^{-1}\tilde{\pi}_l \quad \text{(A.1)}$$

where $\tilde{\pi}_j = 0$ for $j < 0$.

Lemma A.8. *Let $f_\Omega(\lambda) = 2\pi f_\varepsilon(\lambda)f_y(\lambda)$. Define $\vartheta^\infty_j = (2\pi)^{-2}\int_{-\pi}^{\pi} f_\Omega(\lambda)^{-1} e^{i\lambda j}d\lambda = \sigma_\varepsilon^{-2}\sum_{l=0}^\infty \tilde{\pi}'_l\Sigma^{-1}\tilde{\pi}_{l+j}$ for $j \geq 0$ and $\vartheta^\infty_j = \vartheta^\infty_{-j}$ for $j < 0$. Let $\vartheta^M_{k,j}$ be the k,j-th element of Ω^{-1}_M. Then, as $M \to \infty$,*

$$\left\|\vartheta^M_{zM,zM-j} \to \vartheta^\infty_{-j}\right\| = o(1) \text{ uniformly in } z \in (0,1) \text{ and } j > 0.$$

Lemma A.9. *Let $\hat{\Omega}_M$ be defined in (3.2). Let $\bar{M} \to \infty$ such that $\bar{M}/n^{1/3} \to 0$. Then $\Pr\left(\min\hat{\xi}_\Omega \geq 0\right) \to 1$ and*

$$\Pr\left(\hat{\Omega}^*_M = \hat{\Omega}_M\right) \to 1$$

uniformly in $M \leq \bar{M}$. In addition, $\left\|\hat{\Omega}_M - \Omega_M\right\|^2 = O_p\left(M^2/n\right)$ uniformly in $M \leq \bar{M}$.

Lemma A.10. *Let $\hat{\Omega}_M$ be defined in (3.2) and $\sqrt{n}(\tilde{\beta}_n - \beta_0) = O_p(1)$. Let $\bar{M} \to \infty$ such that $\bar{M}/n^{1/3} \to 0$. Then $\left\|\hat{\Omega}_M^{-1} - \Omega_M^{-1}\right\|_2 = O_p(M/n^{1/2})$ and $\left\|\hat{\Omega}_M^{-1}\right\|_2 = O_p(1)$ uniformly in $M \leq \bar{M}$.*

A.4. Lemmas

Lemma A.11. *Define $\rho_{n,M} = O\left(\sqrt{\varphi_n(M,\ell,k(.))}\right)$. Assume that Conditions A, B and C hold. Let $M \to \infty$ such that $M/n^{1/3} \to 0$. Then*

$$\sqrt{n}(\beta_{n,M} - \beta) = b_{n,M} + o_p(\rho_{n,M}) \tag{A.2}$$

with

$$b_{n,M} = D^{-1}\textstyle\sum_{i=0}^{9} d_i - D^{-1}\sum_{i=1}^{4}\sum_{j=0}^{9} H_i D^{-1} d_j \tag{A.3}$$

where d_i are defined in (A.22)-(A.30) below, D and d_0 are defined in Definition A.2 and H_i are defined in (A.4)-(A.6) and (A.20) below.

Proof. First note that $\varphi_n(M,\ell,k(.)) \geq \frac{(Mp)^2}{n}\mathcal{A}$ and

$$\varphi_n(M,\ell,k(.)) \geq \max\left(\ell' D^{-1}\sigma_{1M} D^{-1}\ell, M^{-2q}k_q^2 \ell' D^{-1}\mathcal{B}^{(q)} D^{-1}\ell\right).$$

Because $M \to \infty$ and $M/n^{1/3} \to 0$ it follows that $M^2/n = o(1)$, $\ell' D^{-1}\sigma_{1M} D^{-1}\ell = o(1)$ and $M^{-2q} = o(1)$. Then, any stochastic sequence $T_n = o_p(M/\sqrt{n})$ satisfies $T_n = o_p\left(\rho_{n,M}\right)$. Similarly, $T_n = o_p\left(\max\left\{\sqrt{\ell'\sigma_{1M}\ell}, M^{-q}\right\}\right)$ implies $T_n = o_p\left(\rho_{n,M}\right)$. Consider a second order Taylor approximation of $\hat{D}_M^{-1}$ around D^{-1} such that for $\hat{d}_M = \hat{P}_M' W_M \hat{\Omega}_M^{-1} W_M n^{-1/2}\sum_{t=1}^{n-m}\varepsilon_{t+m} z_{t,M}$,

$$\sqrt{n}(\hat{\beta}_{n,M} - \beta) = D^{-1}[I - (\hat{D}_M - D)D^{-1} + (\hat{D}_M - D)D^{-1}(\hat{D}_M - D)D^{-1}]\hat{d}_M + o_p(\rho_{n,M})$$

where for $M/n^{1/3} \to 0$ the error term is $o_p(\rho_{n,M})$ by the Taylor theorem, and the fact that $\det D \neq 0$, $\hat{D}_M - D = O_p(M/n^{1/2})$ and $\hat{d}_M = O_p(1)$ as shown in the auxiliary appendix. It is shown below that $\hat{d}_M = \sum_{i=0}^{9} d_i + o_p\left(\rho_{n,M}\right)$ and $\hat{D}_M - D = \sum_{i=1}^{4} H_i$ which establishes (A.2). To see this consider the decomposition of $\hat{d}_M$ and $\hat{D}_M - D$. Decompose the expansion into $\hat{D}_M - D = H_1 + ... + H_4$ where

$$H_1 = P_M' W_M \Omega_M^{-1} W_M P_M - P'\Omega^{-1}P, \tag{A.4}$$

$$H_2 = \hat{P}_M' W_M \Omega_M^{-1} W_M \hat{P}_M - P_M' W_M \Omega_M^{-1} W_M P_M, \tag{A.5}$$

$$H_3 = -\hat{P}_M' W_M \Omega_M^{-1}(\hat{\Omega}_M - \Omega_M)\Omega_M^{-1} W_M \hat{P}_M \tag{A.6}$$

and H_4 is defined in (A.20). Also, $\hat{d}_M = d_0 + d_1 + ... + d_9$ with d_i defined in (A.21)-(A.30). The terms H_3 and H_4 contain an expansion of $\hat{\Omega}_M^{-1}$ around Ω_M^{-1} given by

$$\hat{\Omega}_M^{-1} = \Omega_M^{-1} - \Omega_M^{-1}(\hat{\Omega}_M - \Omega_M)\Omega_M^{-1} + B \tag{A.7}$$

where $B = \hat{\Omega}_M^{-1}(\hat{\Omega}_M - \Omega_M)\Omega_M^{-1}(\hat{\Omega}_M - \Omega_M)\Omega_M^{-1}$. The term $\left\|\hat{\Omega}_M - \Omega_M\right\|^2 = o_p(1)$ by Lemma A.9 and $\left\|\hat{\Omega}_M^{-1}\right\|_2 = O_p(1)$ by Lemma A.10. In the auxiliary appendix it is shown that $H_1 = H_{11}+H_{12}+H_{13}+H_{14}$ is

$$H_{11} \equiv P_M'\Omega_M^{-1}P_M - P'\Omega^{-1}P = -\sigma_{1M}, \tag{A.8}$$

$$H_{12} \equiv P_M'(I - W_M)\Omega_M^{-1}(I - W_M)P_M = O(M^{-2q}), \tag{A.9}$$

$$H_{13} \equiv -P_M'\Omega_M^{-1}(I - W_M)P_M = O(M^{-q}), \tag{A.10}$$

$$H_{14} \equiv -P_M'(I - W_M)\Omega_M^{-1}P_M = O(M^{-q}), \tag{A.11}$$

where $\equiv$ means 'equal by definition' and $H_{12} = H_{13} = H_{14} = 0$ for the truncated kernel. In the auxiliary appendix the term $H_2 = H_{211} + H_{212} + H_{221} + H_{222}$ is analyzed to be

$$H_{211} \equiv -\left(\hat{P}_M - \check{P}_M\right)' W_M\Omega_M^{-1}W_M(\hat{P}_M - \check{P}_M) = O_p(M/n^2), \tag{A.12}$$

$$\begin{aligned} H_{212} &\equiv \hat{P}_M'W_M\Omega_M^{-1}W_M(\hat{P}_M - \check{P}_M) + (\hat{P}_M - \check{P}_M)'W_M\Omega_M^{-1}W_M\hat{P}_M \\ &= O_p(M/n^{3/2}) + O_p\left(n^{-1}\right), \end{aligned} \tag{A.13}$$

$$H_{221} \equiv -\left(\check{P}_M - P_M\right)' W_M\Omega_M^{-1}W_M\left(\check{P}_M - P_M\right) = O_p(M/n), \tag{A.14}$$

$$\begin{aligned} H_{222} &\equiv \check{P}_M'W_M\Omega_M^{-1}W_M(\check{P}_M - P_M) + (\check{P}_M - P_M)'W_M\Omega_M^{-1}W_M\check{P}_M \\ &= O_p(M/n) + O_p\left(n^{-1/2}\right), \end{aligned} \tag{A.15}$$

where $\hat{P}_M$ is defined in Section 3 and $\check{P}_M$ is defined from $\check{\Gamma}_j^{xy} = n^{-1}\sum_{t=j+1}^{n-m} w_{t,j}$ as $\check{P}_M = \left[\check{\Gamma}_1^{xy}, ..., \check{\Gamma}_M^{xy}\right]'$.

The auxiliary appendix shows that $H_3 = H_{31} + H_{32} + H_{33} + H_{34}$ is

$$H_{31} \equiv (\hat{P}_M - P_M)'W_M\Omega_M^{-1}(\hat{\Omega}_M - \Omega_M)\Omega_M^{-1}W_M(\hat{P}_M - P_M) = o_p(M/n^{1/2}), \tag{A.16}$$

$$H_{32} \equiv P_M'W_M\Omega_M^{-1}(\hat{\Omega}_M - \Omega_M)\Omega_M^{-1}W_M(\hat{P}_M - P_M) = o_p(M/n^{1/2}), \tag{A.17}$$

$$H_{33} \equiv (\hat{P}_M - P_M)'W_M\Omega_M^{-1}(\hat{\Omega}_M - \Omega_M)\Omega_M^{-1}W_MP_M = o_p(M/n^{1/2}), \tag{A.18}$$

$$H_{34} \equiv P_M'W_M\Omega_M^{-1}(\hat{\Omega}_M - \Omega_M)\Omega_M^{-1}W_MP_M = O_p(n^{-1/2}), \tag{A.19}$$

and H_4 which is a remainder term defined as

$$H_4 \equiv \hat{P}_M'W_M(\hat{\Omega}_M^{-1} - \Omega_M^{-1} + \Omega_M^{-1}(\hat{\Omega}_M - \Omega_M)\Omega_M^{-1})W_M\hat{P}_M = o_p(M/n^{1/2}) \tag{A.20}$$

where the last equality is established in the auxiliary appendix.

Define $V_M = \left[n^{-1/2}\sum_{t=1}^{n-m} v'_{t,1}, ..., n^{-1/2}\sum_{t=1}^{n-m} v'_{t,M}\right]'$ with $V \equiv V_\infty$ as in Definition A.2. In the auxiliary appendix it is shown that $\hat{d}_M = \sum_{j=0}^{9} d_j + o_p(\rho_{n,M})$ with

(A.21) $$d_0 \equiv P'\Omega^{-1}V = O_p(1),$$

(A.22) $$d_1 \equiv P'_M\Omega_M^{-1}V_M - P'\Omega^{-1}V = O_p(\|\sigma_{1M}\|),$$

(A.23) $$d_2 \equiv P'_M(I-W_M)\Omega_M^{-1}(I-W_M)V_M = O_p(M^{-2q}),$$

(A.24) $$d_3 \equiv -P'_M(I-W_M)\Omega_M^{-1}V_M - P'_M\Omega_M^{-1}(I-W_M)V_M = O_p(M^{-q}),$$

(A.25) $$d_4 \equiv \left(\hat{P}_M - \check{P}_M\right)' W_M\Omega_M^{-1}W_MV_M = O_p(M/n),$$

(A.26) $$d_5 \equiv \left(\check{P}_M - P_M\right)' W_M\Omega_M^{-1}W_MV_M = O_p(M/\sqrt{n}),$$

(A.27) $$d_6 \equiv \left(\hat{P}_M - P_M\right)' W_M\Omega_M^{-1}(\Omega_M - \hat{\Omega}_M)\Omega_M^{-1}W_MV_M = o_p(M/\sqrt{n}),$$

(A.28) $$d_7 \equiv P'_MW_M\Omega_M^{-1}(\Omega_M - \hat{\Omega}_M)\Omega_M^{-1}W_MV_M = O_p(M/\sqrt{n}),$$

(A.29) $$d_8 \equiv \hat{P}'_MW_MBW_MV_M = o_p(M/\sqrt{n}),$$

(A.30) $$d_9 \equiv \hat{P}'_MW_M\hat{\Omega}_M^{-1}W_M\left(\tilde{V}_M - V_M\right) = o_p(M/\sqrt{n})$$

where $\tilde{V}_M = n^{-1/2}\sum_{t=1}^{n-m}\tilde{\psi}_{t,M}$ with $\tilde{\psi}_{t,M} = \left(\tilde{v}'_{t,1}, ..., \tilde{v}'_{t,M}\right)'$ and $\tilde{v}_{t,i} = \varepsilon_{t+m}\left(y_{t+1-i} - \bar{y}\right)$ if $i \leq t$ and $\tilde{v}_{t,i} = 0$ otherwise. All rate results are established in the auxiliary appendix. ■

Lemma A.12. *Let d_j and H_{ij} be as defined in Lemma A.11. Then,*
i) $E\left[d_0d'_0\right]D^{-1}H_{11} = H_{11} + O\left(\|\sigma_{1M}\|^2 n^{-1}\right)$,
ii) $E\left[d_0d'_0\right]D^{-1}H_{12} = M^{-2q}k_q^2\mathcal{B}_0^{(q)} + o(M^{-q2})$ *where* $\mathcal{B}_0^{(q)} = \sum_{j_1,j_2=1}^{\infty}\Gamma_{j_1}^{xy}|j_1|^q\vartheta_{j_1,j_2}|j_2|^q\Gamma_{-j_2}^{yx}$,
iii) $E[d_1d'_1] = -H_{11} + O(n^{-1})$,
iv) $E\left[d_1d'_0\right] = H_{11} + O(n^{-1})$,
v) $E\left[d_1d'_3\right] = O(n^{-1})$,
vi) $E\left[d_0d'_2\right] = M^{-2q}k_q^2\mathcal{B}_0^{(q)} + O(n^{-1}) + o(M^{-2q})$,
vii) $E\left[d_0d'_3\right] = -M^{-q}k_q\mathcal{B}_1^{(q)} + O(n^{-1}) + o(M^{-2q})$ *where* $\mathcal{B}_1^{(q)}$ *is defined as*

$$\mathcal{B}_1^{(q)} = \sum_{j_1,j_2=1}^{\infty}\left(\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}|j_2|^q\Gamma_{-j_2}^{yx} + \Gamma_{j_1}^{xy}|j_1|^q\vartheta_{j_1,j_2}\Gamma_{-j_2}^{yx}\right),$$

viii) $H_{13} + H_{14} = -M^{-q}k_q\mathcal{B}_1^{(q)} + o(M^{-q})$,
ix) $E\left[d_3d'_3\right] = M^{-q2}k_q^2\mathcal{B}_2^{(q)} + o(M^{-q2})$ *where* $\mathcal{B}_2^{(q)}$ *is defined in (3.13),*
x) $\ell'D^{-1}E\left[d_5d'_5\right]D^{-1}\ell = M^2/n\left(\int_{-\infty}^{\infty}k(x)^2dx\right)^2\ell'D^{-1}\mathcal{A}_1\mathcal{A}'_1D^{-1}\ell + o(M^2/n)$ *where* $\mathcal{A}_1$ *is defined in (3.5),*
xi) $E\left[d_7d'_7\right] = M^2/n\left(\int_{-1}^{1}k(x)dx\right)^2\mathcal{A}_2\mathcal{A}'_2 + o(M^2/n)$ *where* $\mathcal{A}_2$ *is defined in (3.5),*
xii) $E\left[d_7d'_5\right] = M^2/n\left(\int_{-1}^{1}k(x)dx\int_{-1}^{1}k(x)^2dx\right)\mathcal{A}_2\mathcal{A}'_1 + o(M^2/n)$.

Proof. For i) note that $E[d_0] = 0$. Using Lemma (A.6iii)

$$\begin{aligned} E\left[d_0 d_0'\right] &= \frac{1}{n}\sum\nolimits_{t,s=1}^{n-m}\sum\nolimits_{j_1,\ldots,j_4=1}^{\infty}\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}\gamma_{t-s}^{\varepsilon}\Gamma_{t-s-j_2+j_3}^{yy}\vartheta_{j_3,j_4}\Gamma_{-j_4}^{yx} \\ &= \sum\nolimits_{l=-m+1}^{m-1}\frac{n-|l|}{n}\sum\nolimits_{j_1,\ldots,j_4=1}^{\infty}\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}\gamma_l^{\varepsilon}\Gamma_{l+j_3-j_2}^{yy}\vartheta_{j_3,j_4}\Gamma_{-j_4}^{yx} \\ &\to P'\Omega^{-1}P \text{ as } n\to\infty \end{aligned}$$

where the second line follows from the fact that $\gamma_l^{\varepsilon} = 0$ for $l \geq m$. Then, $E[d_0 d_0'] = D + O(n^{-1})$.

For ii) write $H_{12} = M^{-2q}\sum_{j_1,j_2=1}^{M}\Gamma_{j_1}^{xy}|j_1|^q \frac{1-k(j_1/M)}{|j_1|^q M^{-q}}\vartheta_{j_1,j_2}^M \frac{1-k(j_2/M)}{|j_2|^q M^{-q}}|j_2|^q\Gamma_{-j_2}^{yx}$. Note that $\vartheta_{j_1,j_2}^M \to \vartheta_{j_1,j_2}$ as $M\to\infty$ for all j_1, j_2 fixed and finite by Lemma 2.7i) in the Auxiliary Appendix. By the Dominated Convergence Theorem it follows that

$$\sum_{j_1,j_2=1}^{M}\Gamma_{j_1}^{xy}|j_1|^q\frac{1-k(j_1/M)}{|j_1|^q M^{-q}}\vartheta_{j_1,j_2}^M\frac{1-k(j_2/M)}{|j_2|^q M^{-q}}|j_2|^q\Gamma_{-j_2}^{yx}\to k_q^2\mathcal{B}_0^{(q)} \text{ as } M\to\infty$$

where we have used Assumption A such that $H_{12} = M^{-2q}k_q^2\mathcal{B}_0^{(q)} + o(M^{-2q}) = O(M^{-2q})$. The result then follows immediately from the same argument as in the proof of part i).

For iii) consider

$$\begin{aligned} E\left[d_1 d_1'\right] &= P_M'\Omega_M^{-1}E\left[V_M V_M'\right]\Omega_M^{-1}P_M - P'\Omega^{-1}E\left[VV_M'\right]\Omega_M^{-1}P_M \\ &\quad -P_M'\Omega_M^{-1}E\left[V_M V'\right]\Omega^{-1}P + P'\Omega^{-1}E\left[VV'\right]\Omega^{-1}P \end{aligned}$$

where the i,j-th $p\times p$ block of $E[V_M V_M']$ is

$$n^{-1}\sum_{t,s=1}^{n-m}E\left[v_{t,i}v_{s,j}'\right] = n^{-1}\sum_{t,s=1}^{n-m}\gamma_{t-s}^{\varepsilon}\Gamma_{t-s+i-j}^{yy} = \omega_{i,j} + O(n^{-1}).$$

The same argument shows that the i,j-th $p\times p$ block of the $\infty\times Mp$ matrix $E[VV_M']$ is $\omega_{i,j}+O(n^{-1})$. It then follows that $E[V_M V_M']\Omega_M^{-1} = I_{Mp} + O(n^{-1})$ and $E[V_M V']\Omega^{-1} = [I_{Mp}, \mathbf{0}_{Mp\times\infty}] + O(n^{-1})$ with a similar expression for $\Omega^{-1}E[VV_M']$. This shows that

$$E\left[d_1 d_1'\right] = P'\Omega^{-1}P - P_M'\Omega_M^{-1}P_M + O(n^{-1}) = -H_{11} + O(n^{-1}).$$

For iv) now directly evaluate

$$\begin{aligned} E\left[d_1 d_0'\right] &= P_M'\Omega_M^{-1}E\left[V_M V'\right]\Omega^{-1}P - P'\Omega^{-1}E\left[VV'\right]\Omega^{-1}P \\ &= P_M'\Omega_M^{-1}E\left[V_M V'\right]\Omega^{-1}P - P'\Omega^{-1}P + O\left(n^{-1}\right) \end{aligned}$$

where the $M\times\infty$ matrix $E[V_M V']\Omega^{-1}$ has j,k-th block $\sum_{l=1}^{\infty}\omega_{j,l}\vartheta_{l,k} = I$ if $j=k$ and 0 otherwise. Thus, $P_M'\Omega_M^{-1}E[V_M V']\Omega^{-1}P = P_M'\Omega_M^{-1}P_M + O(n^{-1})$ which implies that $E[d_1 d_0'] = H_{11} + O(n^{-1})$.

For v) directly evaluate

$$\begin{aligned} E\left[d_1 d_3'\right] &= -P_M'\Omega_M^{-1}E\left[V_M V_M'\right]\Omega_M^{-1}(I-W_M)P_M + P'\Omega^{-1}E\left[VV_M'\right]\Omega_M^{-1}(I-W_M)P_M \\ &\quad -P_M'\Omega_M^{-1}E\left[V_M V_M'\right](I-W_M)\Omega_M^{-1}P_M + P'\Omega^{-1}E\left[VV_M'\right](I-W_M)\Omega_M^{-1}P_M \\ &= -P_M'\Omega_M^{-1}(I-W_M)P_M + P_M'\Omega_M^{-1}(I-W_M)P_M \\ &\quad -P_M'(I-W_M)\Omega_M^{-1}P_M + P_M'(I-W_M)\Omega_M^{-1}P_M + O(n^{-1}) \\ &= O(n^{-1}) \end{aligned}$$

by the same arguments as in the proof of iii) and iv).

For vi) directly evaluate

$$\begin{aligned} E\left[d_0 d_2'\right] &= \frac{1}{n}\sum\nolimits_{t,s=1}^{n-m}\sum\nolimits_{j_1,j_2=1}^{\infty}\sum\nolimits_{j_3,j_4=1}^{M}\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}E\left[v_{t,j_2}v_{s,j_3}'\right](1-k(j_3/M))\vartheta_{j_3,j_4}^{M}(1-k(j_4/M))\Gamma_{-j_4}^{yx} \\ &= M^{-2q}k_q^2\sum\nolimits_{j_1,j_2=1}^{\infty}\sum\nolimits_{j_3,j_4=1}^{M}\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}\omega_{j_2,j_3}\left|j_3\right|^q\vartheta_{j_3,j_4}^{M}\left|j_4\right|^q\Gamma_{-j_4}^{xy} + O(n^{-1}) \\ &= M^{-2q}k_q^2\sum\nolimits_{j_1,j_2=1}^{\infty}\Gamma_{j_1}^{xy}\left|j_1\right|^q\vartheta_{j_1,j_2}\left|j_2\right|^q\Gamma_{-j_2}^{xy} + o(M^{-q2}) + O(n^{-1}) \\ &= M^{-2q}k_q^2\mathcal{B}_0^{(q)} + o(M^{-q2}) + O(n^{-1}) \end{aligned}$$

where the Toeplitz Lemma is used for the second equality and dominated convergence for the third equality.

For vii) directly evaluate

$$\begin{aligned} E\left[d_0 d_3'\right] &= -\frac{1}{n}\sum\nolimits_{t,s=1}^{n-m}\sum\nolimits_{j_1,j_2=1}^{\infty}\sum\nolimits_{j_3,j_4=1}^{M}\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}E\left[v_{t,j_2}v_{s,j_3}'\right] \\ &\quad \times\left[(1-k(j_3/M))\vartheta_{j_3,j_4}^{M} + \vartheta_{j_3,j_4}^{M}(1-k(j_4/M))\right]\Gamma_{-j_4}^{yx} \\ &= -M^{-q}k_q\sum\nolimits_{j_1,j_2,j_3,j_4=1}^{\infty}\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}\omega_{j_2,j_3}\left[\left|j_3\right|^q\vartheta_{j_3,j_4} + \vartheta_{j_3,j_4}\left|j_4\right|^q\right]\Gamma_{-j_4}^{yx} + o(M^{-q}) + O(n^{-1}) \\ &= -M^{-q}k_q\sum\nolimits_{j_1,j_2=1}^{\infty}\left(\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}\left|j_2\right|^q\Gamma_{-j_2}^{yx} + \Gamma_{j_1}^{xy}\left|j_1\right|^q\vartheta_{j_1,j_2}\Gamma_{-j_2}^{yx}\right) + o(M^{-q}) + O(n^{-1}). \end{aligned}$$

For viii) write $H_{13} = -M^{-q}\sum_{j_1,j_2=1}^{M}\Gamma_{j_1}^{xy}k(j_1/M)\vartheta_{j_1,j_2}^{M}\frac{(1-k(j_2/M))}{|j_2|^q M^{-q}}\left|j_2\right|^q\Gamma_{-j_2}^{yx}$ such that

$$\begin{aligned} M^{-q}H_{13} &= k_q\sum\nolimits_{j_1,j_2=1}^{M}\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}^{M}\left|j_2\right|^q\Gamma_{-j_2}^{yx} + o(1) \\ &= k_q\sum\nolimits_{j_1,j_2=1}^{\infty}\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}\left|j_2\right|^q\Gamma_{-j_2}^{yx} + o(1). \end{aligned}$$

The second equality uses the convergence of $\vartheta_{j_1,j_2}^{M}\to\vartheta_{j_1,j_2}$ for all j_1, j_2 fixed as well as the fact that ϑ_{j_1,j_2}^{M} is bounded uniformly in M, j_1 and j_2 and that $\left\|\Gamma_{j_1}^{xy}\right\|$ is summable in j_1 by Kuersteiner (2005, Lemma 4.6) such that the result follows by a dominated convergence argument.

For ix) write $d_3 = d_{31} + d_{32}$ where

$$d_{31} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n-m}\sum_{j_1,j_2=1}^{M}\Gamma_{j_1}^{xy}(1-k(j_1/M))\vartheta_{j_1,j_2}^{M}v_{t,j_2} \tag{A.31}$$

and

$$d_{32} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n-m} \sum_{j_1,j_2=1}^{M} \Gamma_{j_1}^{xy} \vartheta_{j_1,j_2}^{M} \left(1 - k\left(j_2/M\right)\right) v_{t,j_2}. \tag{A.32}$$

Then

$$\begin{aligned} M^{2q} E\left[d_{31} d_{31}'\right] &= \sum\nolimits_{j_1,j_2,j_3,j_4=1}^{M} |j_1|^q |j_4|^q \Gamma_{j_1}^{xy} \frac{1-k\left(j_1/M\right)}{|j_1|^q M^{-q}} \vartheta_{j_1,j_2}^{M} \omega_{j_2,j_3} \vartheta_{j_3,j_4}^{M} \frac{1-k\left(j_4/M\right)}{|j_4|^q M^{-q}} \Gamma_{j_4}^{yx} + o(1) \\ &= k_q^2 \sum\nolimits_{j_1,j_4=1}^{\infty} |j_1|^q |j_4|^q \Gamma_{j_1}^{xy} \vartheta_{j_1,j_4} \Gamma_{j_4}^{yx} + o(1) \end{aligned}$$

and

$$\begin{aligned} M^{2q} E\left[d_{32} d_{32}'\right] &= \sum\nolimits_{j_1,j_2,j_3,j_4=1}^{M} \Gamma_{j_1}^{xy} \vartheta_{j_1,j_2}^{M} |j_2|^q \frac{1-k\left(j_2/M\right)}{|j_2|^q M^{-q}} \omega_{j_2,j_3} |j_3|^q \frac{1-k\left(j_3/M\right)}{|j_3|^q M^{-q}} \vartheta_{j_3,j_4}^{M} \Gamma_{-j_4}^{yx} + o(1) \\ &= k_q^2 \sum\nolimits_{j_1,j_2,j_3,j_4=1}^{\infty} \Gamma_{j_1}^{xy} \vartheta_{j_1,j_2} |j_2|^q \omega_{j_2,j_3} |j_3|^q \vartheta_{j_3,j_4} \Gamma_{-j_4}^{yx} + o(1). \end{aligned}$$

Finally we consider the cross-product

$$\begin{aligned} M^{2q} E\left[d_{32} d_{31}'\right] &= k_q^2 \sum\nolimits_{j_1,j_2,j_3,j_4=1}^{M} \Gamma_{j_1}^{xy} \vartheta_{j_1,j_2}^{M} |j_2|^q \omega_{j_2,j_3} \vartheta_{j_3,j_4}^{M} |j_4|^q \Gamma_{-j_4}^{yx} + o(1) \\ &= k_q^2 \sum\nolimits_{j_1,j_2=1}^{\infty} |j_2|^{2q} \Gamma_{j_1}^{xy} \vartheta_{j_1,j_2} \Gamma_{-j_2}^{yx} + o(1). \end{aligned}$$

For x) only consider the largest order term, while the remainder terms are analyzed in the Auxiliary Appendix,

$$d_{51} = n^{-3/2} \sum\nolimits_{j_1,j_2=1}^{M} \sum\nolimits_{t=\max(r-m,j_1)+1}^{n-m} \breve{w}_{t,j_1} k\left(j_1/M\right) \vartheta_{j_1,j_2}^{M} k\left(j_2/M\right) \sum\nolimits_{t=1}^{n-m} v_{t,j_2}$$

where $\breve{w}_{t,j_1} = w_{t,j_1} - \Gamma_{j_1}^{xy}$. Let $\tilde{\ell} = D^{-1}\ell$. Noting that

$$E\left[\tilde{\ell}' \breve{w}_{s_2,j_4} \otimes \vartheta_{j_1,j_2}^{M} v_{t_1,j_2}\right] E\left[\operatorname{vec}(v_{t_2,j_3}' \vartheta_{j_3,j_4}^{M\prime}) \operatorname{vec}(\breve{w}_{s_1,j_1}' \tilde{\ell})'\right] = \vartheta_{j_1,j_2}^{M} E\left[v_{t_1,j_2} \tilde{\ell}' \breve{w}_{s_2,j_4}\right] \vartheta_{j_3,j_4}^{M} E\left[v_{t_2,j_3} \tilde{\ell}' \breve{w}_{s_1,j_1}\right]$$

and using Lemma A.5 leads to

$$\begin{aligned} & E\left[\tilde{\ell}' d_{51} d_{51}' \tilde{\ell}\right] \\ &= \frac{1}{n^3} \sum_{j_1,..,j_4}^{M} \sum_{t_1,t_2 s_1,s_2}^{n-m} \prod_{l=1}^{4} k\left(j_l/M\right) \Big\{ \left(\operatorname{vec} \vartheta_{j_1,j_2}^{M\prime}\right)' E\left[\left(v_{t_1,j_2} \otimes \breve{w}_{s_1,j_1}'\right] (I \otimes \tilde{\ell}\tilde{\ell}') E\left[\breve{w}_{s_2,j_4} \otimes v_{t_2,j_3}'\right] \operatorname{vec} \vartheta_{j_3,j_4}^{M} \\ &\qquad + \operatorname{tr}\left[\vartheta_{j_1,j_2}^{M} E\left[v_{t_1,j_2} \tilde{\ell}' \breve{w}_{s_2,j_4}\right] \vartheta_{j_3,j_4}^{M} E\left[v_{t_2,j_3} \tilde{\ell}' \breve{w}_{s_1,j_1}\right]\right] \\ &\qquad + \operatorname{tr}\left[\vartheta_{j_1,j_2}^{M} E\left[v_{t_1,j_2} v_{t_2,j_3}'\right] \vartheta_{j_3,j_4}^{M} E\left[\breve{w}_{s_2,j_4}' \tilde{\ell}\tilde{\ell}' \breve{w}_{s_1,j_1}\right]\right] \Big\} + \tilde{\ell}' \mathcal{K}_8 \tilde{\ell} \end{aligned}$$

where the matrix of eighth order cumulant terms $\mathcal{K}_8$ contains elements of the form

$$\frac{1}{n^3} \sum_{j_1,..,j_4}^{M} \sum_{t_1,t_2 s_1,s_2}^{n-m} \left[\vartheta_{j_1,j_2}^{M}\right]_{i_2,i_3} \operatorname{cum}^*\left(\breve{w}_{t_1,j_1}^{i_1,i_2}, \breve{w}_{t_2,j_1}^{i_4,i_5}, v_{s_2,j_2}^{i_3}, v_{s_1,j_4}^{i_4}\right) \left[\vartheta_{j_1,j_2}^{M}\right]_{i_5,i_4} = O\left(\frac{M^2}{n^2}\right) = o\left(\frac{M^2}{n}\right)$$

which are of lower order due to Assumption (B). The first term can be written as

$$\frac{1}{n^3}\sum_{j_1,j_2}^{M}\sum_{t_1,s_1=1}^{n-m}\prod_{l=1}^{4}k(\tfrac{j_l}{M})\left(\operatorname{vec}\vartheta_{j_1,j_2}^{M\prime}\right)' E\left[\left(v_{t_1,j_2}\otimes\breve{w}'_{s_1,j_1}\right]\left(I\otimes\tilde{\ell}\tilde{\ell}'\right)\sum_{j_3,j_4}^{M}\sum_{t_2,s_2=1}^{n-m}E\left[\breve{w}_{s_2,j_4}\otimes v'_{t_2,j_3}\right]\operatorname{vec}\vartheta_{j_3,j_4}^{M}$$

where

$$E\left[v_{t_1,j_2}\otimes\breve{w}'_{s_1,j_1}\right]=\left(\operatorname{vec}\left(\Gamma^{yy}_{s_1-t_1-j_1+j_2}\right)\otimes\Gamma^{\varepsilon x\prime}_{t_1-s_1}\right)+K_{pp}(\Gamma^{\varepsilon y}_{t_1-s_1+j_1}\otimes\Gamma^{yx}_{t_1-s_1-j_2})+\mathcal{K}_4^1 \tag{A.33}$$

by Lemma A.6i) with $\mathcal{K}_4^1$ being a term of fourth order cumulants. Focusing on the first term one finds

$$\begin{aligned}&\frac{1}{nM}\sum_{t_1,s_1}^{n-m}\sum_{j_1,j_2}^{M}\prod_{l=1}^{2}k\left(j_l/M\right)\left(\operatorname{vec}\vartheta_{j_1,j_2}^{M\prime}\right)'\left(\operatorname{vec}\left(\Gamma^{yy}_{s_1-t_1-j_1+j_2}\right)\otimes\Gamma^{\varepsilon x\prime}_{t_1-s_1}\right)\\&=2\pi\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\frac{1}{M}\sum_{j_1,j_2}^{M}\prod_{l=1}^{2}k\left(j_l/M\right)\left(\operatorname{vec}\vartheta_{j_1,j_2}^{M\prime}\right)'\left(\operatorname{vec}f_y(\lambda)\otimes f_{\varepsilon x'}(\mu)\right)e^{i\lambda(-j_1+j_2)}K_n(\lambda)d\lambda d\mu+o\left(1\right)\end{aligned}$$

where $K_n(\lambda)=(2\pi)^{-1}n^{-1}\sum_{t_1,s_1=1}^{n}e^{i\lambda(t_1-s_1)}$ is the Fejer kernel (see Brockwell and Davis, 1991, p.70). Let $\eta_M(\lambda)=\frac{1}{M}\sum_{j_1,j_2}^{M}\prod_{l=1}^{2}k\left(j_l/M\right)\left(\operatorname{vec}\vartheta_{j_1,j_2}^{M\prime}\right)e^{i\lambda(-j_1+j_2)}$. Based on Parzen (1957) we set $u_1=j_1$, $u_2=j_1-j_2$ and $z=u_1/M$ such that

$$\eta_M(\lambda)=\frac{1}{M}\sum_{z=1/M}^{1}\sum_{u_2=-M+1}^{M-1}k(z-\frac{u_2}{M})k(z)\left(\operatorname{vec}\vartheta_{zM,zM-u_2}^{M\prime}\right)e^{-i\lambda u_2}$$

where $\sum_{z=1/M}^{1}$ stands for the sum over u_1/M for $u_1=1,...,M$. Let $\tilde{\pi}(L)=\theta(L)^{-1}\pi(L)$ such that $\omega_{j,k}=(2\pi)^{-1}\int_{-\pi}^{\pi}\tilde{\pi}\left(e^{-i\lambda}\right)^{-1}\Sigma\tilde{\pi}\left(e^{i\lambda}\right)^{\prime-1}e^{i\lambda(k-j)}d\lambda$. Now, for any $\epsilon>0$ fix k_0 such that $\sum_{|u_2|\geq k_0}^{M-1}\left\|\vartheta_{zM,zM-u_2}^{M\prime}\right\|\leq\sum_{|u_2|\geq k_0}^{M-1}\sum_{l=0}^{zM-1}\left\|\tilde{\pi}'_{l,l+M-zM}\right\|\left\|\Sigma_{l+M-zM}^{-1}\right\|\left\|\tilde{\pi}_{l+u_2,zM+M-l}\right\|\leq\epsilon$ where the representation of Lewis and Reinsel (1985, p.402) for $\vartheta_{zM,zM-u_2}^{M\prime}$ and summability of $\tilde{\pi}_{j,k}$ across j uniformly in k, established in the Auxiliary Appendix, is used. Then

$$\eta_M(\lambda)=\frac{1}{M}\sum_{z=1/M}^{1}\sum_{u_2=-k_0+1}^{k_0-1}k(z-\frac{u_2}{M})k(z)\left(\operatorname{vec}\vartheta_{zM,zM-u_2}^{M\prime}\right)e^{-i\lambda u_2}+\epsilon c_0$$

for some constant c_0. As $M\to\infty$, $k(z-\frac{u_2}{M})\to k(z)$ uniformly in $z\in[0,1]$ and $u_2\in[-k_0,k_0]$ and $\vartheta_{zM,zM-u_2}^{M}\to\vartheta_{-u_2}^{\infty}=\sum_{l=0}^{\infty}\tilde{\pi}'_l\Sigma^{-1}\tilde{\pi}_{l-u_2}$ uniformly in $z\in(0,1)$ by Lemma A.8. Note that $\vartheta_{u_2}^{\infty}$ does not depend on z any more. By definition $\vartheta_{-u}^{\infty}=(2\pi)^{-2}\int_{-\pi}^{\pi}f_{\Omega}^{-1}(\lambda)d\lambda e^{i\lambda u}$ and by Dominated Convergence

$$\eta_M(\lambda)\to(2\pi)^{-1}\operatorname{vec}\left(f_{\Omega}^{-1}\left(\lambda\right)\right)'\int_0^1k(z)^2dz\equiv\eta(\lambda)$$

uniformly in $\lambda\in[-\pi,\pi]$. Now consider

$$\begin{aligned}\left\|\int_{-\pi}^{\pi}\eta_M(\mu-\lambda)K_n(\lambda)d\lambda-\eta(\mu)\right\|&\leq\left\|\int_{-\pi}^{\pi}\left(\eta_M(\mu-\lambda)-\eta(\mu-\lambda)\right)K_n(\lambda)d\lambda\right\|\\&\quad+\left\|\int_{-\pi}^{\pi}\eta(\mu-\lambda)K_n(\lambda)d\lambda-\eta(\mu)\right\|\\&\leq\epsilon_1+\epsilon_2\end{aligned}$$

where $\left\|\int_{-\pi}^{\pi}\left(\eta_M(\mu-\lambda)-\eta(\mu-\lambda)\right)K_n(\lambda)d\lambda\right\| \leq \sup_\mu|\eta_M(\mu)-\eta(\mu)|\int_{-\pi}^{\pi}K_n(\lambda)d\lambda \leq \epsilon_1$ for M large enough by uniform convergence of $\eta_M(\lambda)$ and $\left\|\int_{-\pi}^{\pi}\eta(\mu-\lambda)K_n(\lambda)d\lambda-\eta(\mu)\right\| \leq \epsilon_2$ for n large enough by Theorem 2.11.1 of Brockwell and Davis (1991). It follows that

$$\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}2\pi\eta_M(\lambda)\left(\operatorname{vec}f_y(\lambda)\otimes f_{\varepsilon x'}(\mu)\right)K_n(\lambda-\mu)d\mu d\lambda$$
$$\to 2^{-1}\int_{-1}^{1}k(z)^2dz\int_{-\pi}^{\pi}\operatorname{vec}\left(f_\Omega^{-1}(\lambda)'\right)'\left(\operatorname{vec}f_y(\lambda)\otimes f_{\varepsilon x'}(\lambda)\right)d\lambda.$$

The terms involving $K_{pp}(\Gamma^{\varepsilon y}_{t_1-s_1+j_1}\otimes\Gamma^{yx}_{t_1-s_1-j_2})$ and $\mathcal{K}_4^1$ in (A.33) go to zero by the same arguments as in Parzen (1957, p. 341) because

$$\left\|\frac{1}{n}\sum_{t_1,s_1=1}^{n-m}\sum_{j_1,j_2=1}^{M}\prod_{l=1}^{2}k\left(j_l/M\right)\left(\operatorname{vec}\vartheta^{M\prime}_{j_1,j_2}\right)'K_{pp}(\Gamma^{\varepsilon y}_{t_1-s_1+j_1}\otimes\Gamma^{yx}_{t_1-s_1-j_2})\right\|$$
$$\leq\frac{1}{n}\sum_{t_1,s_1=1}^{n-m}\sum_{j_1,j_2=1}^{M}\left\|\vartheta^M_{j_1,j_2}\right\|\left\|K_{pp}(\Gamma^{\varepsilon y}_{t_1-s_1+j_1}\otimes\Gamma^{yx}_{t_1-s_1-j_2})\right\|=O(1)$$

and the cumulant term is of lower order. Next turn to

$$\vartheta^M_{j_1,j_2}E\left[v_{t_1,j_2}\tilde{\ell}'\check{w}_{s_2,j_4}\right]\vartheta^M_{j_3,j_4}E\left[v_{t_2,j_3}\tilde{\ell}'\check{w}_{s_1,j_1}\right]$$
$$=\vartheta^M_{j_1,j_2}(\tilde{\ell}'\Gamma^{\varepsilon x}_{t_1-s_2}\Gamma^{yy}_{t_1-s_2-j_2+j_4}+\left(\tilde{\ell}'\Gamma^{xy}_{s_2-t_1+j_2}\right)'\left(\Gamma^{\varepsilon y}_{t_1-s_2+j_4}\right)'+\mathcal{K}_4^2)'\vartheta^M_{j_3,j_4}$$
$$\times(\tilde{\ell}'\Gamma^{\varepsilon x}_{t_2-s_1}\Gamma^{yy}_{t_2-s_1-j_3+j_1}+\left(\tilde{\ell}'\Gamma^{xy}_{s_2-t_1+j_2}\right)'\left(\Gamma^{\varepsilon y}_{t_1-s_2+j_4}\right)'+\mathcal{K}_4^2)$$

which follows from Lemma A.6ii) where for a typical term in this product one obtains

$$\sum_{j_1,j_2,j_3,j_4}\prod_{l=1}^{4}k\left(j_l/M\right)\vartheta^M_{j_1,j_2}\tilde{\ell}'\sum_{h_1=-n+m+1}^{n-m-1}\left[(1-\tfrac{|h_1|}{n})\Gamma^{\varepsilon x}_{h_1}\Gamma^{yy}_{h_1-j_2+j_4}\right]\vartheta^M_{j_3,j_4}\sum_{h_2=-n+m+1}^{n-m-1}(1-\tfrac{|h_2|}{n})\Gamma^{\varepsilon y}_{h_2+j_1}\tilde{\ell}'\Gamma^{xy}_{h_2+j_3}$$

and changing variables $k_2=h_2+j_1$, $u_1=j_1-j_2$, $u_2=j_4-j_2$ and $u_3=j_4-j_3$ leads to

$$\left\|\sum_{u_1,u_2,u_3,j_4}\prod_{l=1}^{4}k\left(j_l/M\right)\vartheta^M_{u_1-u_2+j_4,-u_2+j_4}\tilde{\ell}'\sum_{h_1}\left[(1-\tfrac{|h_1|}{n})\Gamma^{\varepsilon x}_{h_1}\Gamma^{yy}_{h_1+u_2}\right]\vartheta^M_{j_4-u_3,j_4}\sum_{k_2}(1-\tfrac{|k_2-j_1|}{n})\Gamma^{\varepsilon y}_{k_2}\tilde{\ell}'\Gamma^{xy}_{k_2+u_1-u_2+u_3}\right\|$$
$$\leq\sum_{u_1,u_2,u_3,j_4}\left\{\prod_{l=1}^{4}\left|k\left(j_l/M\right)\right|\left\|\vartheta^M_{u_1-u_2+j_4,-u_2+j_4}\right\|\left\|\tilde{\ell}'\sum_{h_1}(1-\tfrac{|h_1|}{n})\Gamma^{\varepsilon x}_{h_1}\Gamma^{yy}_{h_1+u_2}\right\|\right.$$
$$\left.\times\left\|\vartheta^M_{j_4-u_3,j_4}\right\|\left\|\sum_{k_2}(1-\tfrac{|k_2-j_1|}{n})\Gamma^{\varepsilon y}_{k_2}\tilde{\ell}'\Gamma^{xy}_{k_2+u_1-u_2+u_3}\right\|\right\}=O(M).$$

Similar arguments show that the remaining terms of $E\left[\tilde{\ell}' d_5 d_5' \tilde{\ell}\right]$ are all $O(M/n)$. Finally, note that $f_\Omega(\lambda) = 2\pi f_\varepsilon(\lambda) f_y(\lambda)$ such that $f_y(\lambda) f_\Omega^{-1}(\lambda) = (2\pi)^{-1} f_\varepsilon^{-1}(\lambda) I_p$. From

$$(\operatorname{vec} f_\Omega^{-1}(\lambda)')' (\operatorname{vec} f_y(\lambda)) = \operatorname{tr} f_\Omega^{-1}(\lambda) f_y(\lambda) = (2\pi)^{-1} p f_\varepsilon^{-1}(\lambda)$$

it follows that $(\operatorname{vec} f_\Omega^{-1}(\lambda)')' [\operatorname{vec} f_y(\lambda) \otimes f_{\varepsilon x}(\lambda)'] = (2\pi)^{-1} p f_\varepsilon^{-1}(\lambda) f_{\varepsilon x}(\lambda)'$.

For xi) and xii) let $a_{j_2}^M = \sum_{j_1=1}^M \Gamma_{j_1}^{xy} \vartheta_{j_1,j_2}^M$. Note that

$$\textstyle\sum_{j_1=1}^M \Gamma_{j_1}^{xy} k\left(j_1/M\right) \vartheta_{j_1,j_2}^M - a_{j_2}^M = M^{-q} k_q \sum_{j_1=1}^M |j_1|^q \Gamma_{j_1}^{xy} \vartheta_{j_1,j_2}^M + o\left(M^{-q}\right)$$

uniformly in j_2 such that the presence of the weight function can be neglected. As shown in the auxiliary appendix the largest term of $E\left[d_7 d_7'\right]$ is

$$\begin{aligned}
E\left[d_7\right] &= \frac{1}{Mn} \sum_{j_2,\ldots,j_4=1}^{M} \sum_{t_1=1,t_2=r_2}^{n-m} k\left(j_4/M\right)\left(I \otimes a_{j_2}^M\right) \\
&\times \left(\left(\operatorname{vec}\left(\textstyle\sum_{l_1=-m+1}^{m-1} \gamma_{l_1}^{\varepsilon} \Gamma_{t_2-t_1+j_4-j_3-l_1}^{yy}\right)\right)' \otimes \Gamma_{t_1-t_2+j_2}^{\varepsilon y}\right) \operatorname{vec} \vartheta_{j_3,j_4}^M \\
&= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \frac{1}{M} \sum_{j_3,j_4=1}^{M} k\left(j_4/M\right)\left(I \otimes f_M^a\left(\lambda\right)\right)\left(\operatorname{vec}\left(f_\Omega(\lambda)\right)' \otimes f_{\varepsilon y}(\mu)\right) \operatorname{vec} \vartheta_{j_3,j_4}^M e^{i\lambda(j_4-j_3)} K_n(\mu-\lambda) d\lambda d\mu
\end{aligned}$$

where $f_M^a\left(\lambda\right) = \sum_{j_2=1}^M a_{j_2}^M e^{i\lambda j_2}$ and Definition A.3 was used to substitute for $f_\Omega(\lambda)$. Then by the same arguments as in the proof of Lemma A.12x), set $u_1 = j_4$, $u_2 = j_4 - j_3$ and $z = u_1/M$ and define

$$\eta_M(\lambda) = \frac{1}{M} \sum_{z=1/M}^{1} \sum_{u_2=-M+1}^{M-1} k(z)\left(I \otimes f_M^a\left(\lambda\right)\right)\left(\operatorname{vec}\left(f_\Omega(\lambda)\right)' \otimes f_{\varepsilon y}(\mu)\right)\left(\operatorname{vec} \vartheta_{zM-u_2,zM}^{M\prime}\right) e^{i\lambda u_2}.$$

Now, $f_M^a\left(\lambda\right) \leq \sum_{j_2=1}^M \left\|a_{j_2}^M\right\| < \infty$ by Lemma 2.7iv) in the Auxiliary Appendix such that $f_M^a\left(\lambda\right) \to f^a\left(\lambda\right) = \sum_{j_1,j_2=1}^\infty \Gamma_{j_1}^{xy} \vartheta_{j_1,j_2} e^{-i\lambda j_2}$ uniformly in λ by Folland (1984, p.240) and $\sum_{u_2=-M+1}^{M-1} \vartheta_{zM-u_2,zM}^{M\prime} e^{i\lambda u_2} \to (2\pi)^{-1} f_\Omega^{-1}\left(\lambda\right)'$. Then, $\sqrt{n}/M E\left[d_7\right]$ converges to

$$\begin{aligned}
&2^{-1}\int_{-1}^{1} k(z)dz \int_{-\pi}^{\pi}\left(\operatorname{vec}\left(f_\Omega(\lambda)\right)' \otimes f^a\left(\lambda\right) f_{\varepsilon y}(\lambda)\right) \operatorname{vec}\left(f_\Omega^{-1}\left(\lambda\right)'\right) d\lambda \\
&= \left(\int_{-1}^{1} k(z)dz\right) \frac{p}{2} \int_{-\pi}^{\pi} f^a\left(\lambda\right) f_{\varepsilon y}(\lambda) d\lambda
\end{aligned}$$

because $f^a\left(\lambda\right) f_{\varepsilon y}(\lambda)$ is a vector of dimension $d \times 1$ and $\operatorname{vec}\left(f_\Omega(\lambda)\right)' \operatorname{vec}\left(f_\Omega^{-1}\left(\lambda\right)'\right) = \operatorname{tr} f_\Omega(\lambda)' f_\Omega^{-1}\left(\lambda\right)' = p$. Note that $\int_{-\pi}^{\pi} f^a\left(\lambda\right) f_{\varepsilon y}(\lambda) d\lambda = \sum_{j_1,j_2=1}^\infty \Gamma_{j_1}^{xy} \vartheta_{j_1,j_2} \Gamma_{-j_2}^{\varepsilon y}$. In the auxiliary appendix it is shown that $E\left[d_5 d_7'\right] = E\left[d_5\right] E\left[d_7'\right] + o\left(M/\sqrt{n}\right)$. ■

A.5. Main Results

Proof of Proposition 3.1. Let $b_{n,M}$ be defined in Lemma A.11. For $E[b_{n,M}]$ consider $E[d_i]$ and $E[H_i D d_j]$. First, $E[d_i] = 0$ for $i \leq 3$. The terms d_4, d_6, d_8, d_9 are of lower order by the proof of Lemma A.11. The terms $H_i D^{-1} d_j$ are all of lower order. To see this note that $H_i = o_p(M/n^{1/2})$ for $i \geq 2$, $d_j = O_p(M/\sqrt{n})$ for $j \geq 4$ such that for $i \geq 2$ and $j \geq 4$, $H_i D^{-1} d_j = o_p\left(M/n^{1/2}\right)$. Also, for $j < 4$, $E[d_j] = 0$ such that $E\left[H_1 D^{-1} d_j\right] = 0$ for $j < 4$. Finally, $H_1 D^{-1} d_j = o_p(M/\sqrt{n})$ for $j \geq 4$ because $H_1 = o(1)$ as $M \to \infty$. The largest order term is therefore $E[d_5 + d_7]$. By the proof of Lemma A.12x-xi) it follows that $E[d_5 + d_7] = pM/\sqrt{n}\left(\mathcal{A}_1 \int_{-1}^{1} k(x)^2 dx + \mathcal{A}_2 \int_{1}^{1} k(x) dx\right) + o(M/\sqrt{n})$. ■

Proof of Theorem 3.3. First focus on smooth kernels. Consider the terms in the expansion $D^{-1}\sum_{i=0}^{9} d_i - D^{-1}\sum_{i=1}^{4}\sum_{j=0}^{9} H_i D^{-1} d_j$ of the estimator which depend on M and n and are largest in probability. From the results in Equations (A.8) to (A.30) it follows that the largest such terms are H_1, d_0, d_2, d_3, d_5 and d_7. Of those terms examine cross products of the form $E\left[d_i d_j'\right]$, $E[d_i d_0'] D^{-1} H_i$ and $H_i D^{-1} E[d_0 d_0'] D^{-1} H_j$. Letting $\mathcal{B}_1^{(q)} = k_q^{-1} \lim_{M\to\infty} (H_{13} + H_{14})/M^{-q}$, the largest terms vanishing at rate M^{-q} as $M \to \infty$ are $E[d_0 d_3'] = -M^{-q} k_q \mathcal{B}_1^{(q)} + o(M^{-q})$ as shown in Lemma A.12vii) and $-E[d_0 d_0'] D^{-1}(H_{13} + H_{14}) = M^{-q} k_q \mathcal{B}_1^{(q)} + o(M^{-q})$ by Lemmas A.12i) and A.12viii) . The two terms cancel because they are of opposite sign.

Now define $\mathcal{B}_0^{(q)} = k_q^{-2} \lim_{M\to\infty} P_M'(I - W_M)\Omega_M^{-1}(I - W_M)P_M/M^{-q2}$. Terms of order M^{-2q} include $E[d_0 d_2'] = M^{-2q} k_q^2 \mathcal{B}_0^{(q)} + o(M^{-2q})$ by Lemma A.12vi) and $-E[d_0 d_0'] D^{-1} H_{12}' = -M^{-2q} k_q^2 \mathcal{B}_0^{(q)} + o(M^{-2q})$ by Lemma A.12ii). Since $E[d_0 d_2']$ and $-E[d_0 d_0'] D^{-1} H_{12}'$ are of opposite sign these terms cancel. We are left with $E\left[(d_3 - (H_{13} + H_{14})D^{-1} d_0)(d_3 - (H_{13} + H_{14})D^{-1} d_0)'\right] = O(M^{-2q})$ by Lemmas A.12vii-ix).

The term that grows with M and is largest in order is $d_5 + d_7$ where $E[(d_5 + d_7)(d_5 + d_7)'] = O(M^2/n)$ by Lemma A.12x-xii). Then $\varphi_n(M, \ell, k(.)) = O(M^2/n) + O(M^{-2q})$.

Next turn to the case of the truncated kernel. Now H_{11}, d_0, d_1, d_5 and d_7 are largest in probability. From Lemmas A.12i) and A.12iv) it follows that $E[d_0 d_0'] D^{-1} H_{11} = E[d_1 d_0']$ such that these terms cancel out. The largest terms remaining are therefore $E[d_1 d_1'] = -H_{11} + o(\sigma_{1M})$, d_5 and d_7. The largest term growing with M is $E[(d_5 + d_7)(d_5 + d_7)'] = O(M^2/n)$ as before.

To summarize, the dominating terms in $\varphi_n(M, \ell, k(.))$ are $A_n = E[(d_5 + d_7)(d_5 + d_7)']$ and

$$B_n = E\left[(d_3 - (H_{13} + H_{14})D^{-1} d_0)(d_3 - (H_{13} + H_{14})D^{-1} d_0)'\right] + E\left[d_1 d_1'\right]$$

such that $\varphi_n(M, \ell, k(.)) = A_n + B_n + o\left(\rho_{n,M}\right)$. For all $n \geq 1$ it holds that $A_n \geq 0$ and $B_n \geq 0$.

From Lemma A.12x-xii) it follows that for $\mathcal{A}_0 = \mathcal{A}_1 \int_{-1}^{1} k(x)^2 dx + \mathcal{A}_2 \int_{-1}^{1} k(x) dx$,

$$E\left[\ell' D^{-1}(d_5 + d_7)(d_5 + d_7)' D^{-1} \ell\right] = (Mp)^2/n \ell' D^{-1} \mathcal{A}_0' \mathcal{A}_0 D^{-1} \ell + o(M^2/n).$$

From Lemma A.12vii) it follows that $M^{2q}E\left[d_0 d_3'\right] = -k_q \mathcal{B}_1^{(q)} + o(1)$ and from Lemma A.12i) it follows that $E\left[d_0 d_0'\right] = D + o(1)$ such that

$$M^{2q}(H_{13}+H_{14})D^{-1}E\left[d_0 d_0'\right]D^{-1}(H_{13}+H_{14})' = k_q^2 \mathcal{B}_1^{(q)} D^{-1}\mathcal{B}_1^{(q)\prime} + o(1).$$

This implies that

$$(H_{13}+H_{14})D^{-1}E\left[d_0 d_3'\right] - (H_{13}+H_{14})D^{-1}E\left[d_0 d_0'\right]D^{-1}(H_{13}+H_{14})' = o(M^{-2q})$$

or in other words $B_n = E\left[d_3 d_3'\right] - (H_{13}+H_{14})D^{-1}E\left[d_0 d_0'\right]D^{-1}(H_{13}+H_{14}) + o(M^{-2q})$. Here $E\left[d_3 d_3'\right] = M^{-2q}k_q^2 \mathcal{B}_2^{(q)} + o(M^{-2q})$ as shown in Lemma A.12ix) where $\mathcal{B}_2^{(q)}$ is defined in (3.13). This implies that for smooth kernels $B_n = M^{-2q}k_q^2\left(\mathcal{B}_2^{(q)} - \mathcal{B}_1^{(q)}D^{-1}\mathcal{B}_1^{(q)\prime}\right) + o\left(M^{-2q}\right)$ since $E\left[d_1 d_1'\right] = -H_{11} + o\left(\sigma_{1M}\right)$ is of lower order. For the truncated kernel on the other hand, d_3, H_{13} and H_{14} vanish such that $B_n = -H_{11} + o\left(\sigma_{1M}\right)$. ■

Proof of Theorem 4.1. Use the short hand notation $\varphi_n(M) = \frac{(Mp)^2}{n}\mathcal{A} + \ell' D^{-1}\sigma_{1M}D^{-1}\ell$ and $\hat{\varphi}_n(M) = \frac{(Mp)^2}{n}\widehat{\mathcal{A}} + \ell'\hat{D}^{-1}_{k_{\max},\hat{h}}\hat{\sigma}_{1M}\hat{D}^{-1}_{k_{\max},\hat{h}}\ell$ where $\hat{\sigma}_{1M} = \left(\hat{D}_{k_{\max},\hat{h}} - \hat{D}_{M,\hat{h}}\right)$. Consider

$$\begin{aligned}\left|\frac{\hat{\varphi}_n(M) - \varphi_n(M)}{\varphi_n(M)}\right| \leq{}& \left|\frac{\ell' D^{-1}\left(\hat{\sigma}_{1M} - \sigma_{1M}\right)D^{-1}\ell}{\varphi_n(M)}\right| + \left|\frac{\mathcal{A} - \widehat{\mathcal{A}}}{\mathcal{A}}\right| \\ &+2\left|\frac{\ell'\left(D^{-1} - \hat{D}^{-1}_{k_{\max},\hat{h}}\right)\sigma_{1M}D^{-1}\ell}{\ell' D^{-1}\sigma_{1M}D^{-1}\ell}\right| + o_p\left(M^3 h_{\max}\left(\log n/n\right)^{1/2} + o_p\left(M^{-2}\right)\right)\end{aligned}$$

because $\varphi_n(M) \geq M^2/np^2\mathcal{A}$ and $\ell' D^{-1}\left(D - D_M\right)D^{-1}\ell \geq 0$. By Lemmas 4.7, 4.8 and 4.9 in the Auxiliary Appendix it follows that $\widehat{\mathcal{A}} - \mathcal{A} = O_p\left(h_{\max}\left(\log n/n\right)^{1/2}\right)$. By Lemma 4.6ii) of the Auxiliary Appendix it follows that

$$\varphi_n(M)^{-1}\left(\ell' D^{-1}\left(\hat{\sigma}_{1M} - \sigma_{1M}\right)D^{-1}\ell\right) = M^3 O_p(h_{\max}\left(\log n/n\right)^{1/2}) + o_p\left(M^{-2}\right) \tag{A.34}$$

and $D^{-1} - \hat{D}^{-1}_{k_{\max},\hat{h}} = O_p\left(h_{\max}\sqrt{\log n/n}\right)$ by Lemma 4.7 in the Auxiliary Appendix. This shows that

$$\hat{\varphi}_n(M) = \varphi_n(M)\left(1 + M^3 O_p(h_{\max}\left(\log n/n\right)^{1/2} + o_p\left(M^{-2}\right)\right) = \varphi_n(M)\left(1 + o_p(1)\right) \tag{A.35}$$

uniformly in $M \leq M_{\max}$. Let $\tilde{\varphi}_n(M) = \varphi_n(M) + g(M) - \ell' D^{-1}\sigma_{1M}D^{-1}\ell$ where $g(M) - \ell' D^{-1}\sigma_{1M}D^{-1}\ell = o\left(\nu^M\right)$ because $\ell' D^{-1}\sigma_{1M}D^{-1}\ell/g(M) - 1 = o\left(1\right)$ by assumption. It follows that $\varphi_n(M) - \tilde{\varphi}_n(M) = o(g(M))$. Let $\tilde{M}_1^*$ minimize $\tilde{\varphi}_n(M)$. By the same arguments as in Hannan and Deistler (1988, p.333) it now follows that $\tilde{M}_1^*/M_1^* - 1 \to 0$, $\hat{M}_1^*/\tilde{M}_1^* - 1 = o_p(1)$ and $\tilde{\varphi}_n(\hat{M}_1^*)/\tilde{\varphi}_n(\tilde{M}_1^*) = 1 + o_p(1)$. ■

Table 1a

corr(u_t,v_t) = .1, ϕ = .1

θ	Estimator	Median Bias (n=128)	Dec Range (n=128)	MSE (n=128)	MAE (n=128)	Size (n=128)	Median # Inst (n=128)	Median Bias (n=512)	Dec Range (n=512)	MSE (n=512)	MAE (n=512)	Size (n=512)	Median # Inst (n=512)
-0.9	OLS	0.11	0.32	0.03	0.13	0.16		0.11	0.16	0.02	0.11	0.46	
	GMM-1	0.15	3.93	10.79	1.48	0.05	1.0	0.07	3.65	4.88	1.31	0.05	1.0
	GMM-25	0.11	0.51	0.05	0.18	0.08	25.0	0.10	0.53	0.05	0.18	0.07	25.0
	GMM-Tuk-Han	0.11	1.38	0.41	0.46	0.03	7.5	0.14	1.09	0.21	0.36	0.02	9.9
	GMM-BR	0.12	1.87	0.79	0.62	0.06	3.9	0.13	1.51	0.48	0.51	0.04	5.2
	GMM-Trunc	0.23	2.50	2.69	0.92	0.07	1.0	0.26	2.01	2.06	0.80	0.08	2.0
	CUE-1	0.09	9.42	49.69	3.74	0.01	1.0	-0.01	9.29	46.08	3.58	0.00	1.0
	CUE-25	0.02	9.82	45.67	3.56	0.82	25.0	0.01	8.88	42.35	3.41	0.83	25.0
	WWA	0.03	2.12	109.83	1.33	0.12		0.02	2.04	9.71	0.98	0.13	
-0.5	OLS	0.10	0.26	0.02	0.12	0.20		0.11	0.13	0.01	0.11	0.57	
	GMM-1	0.14	3.06	24.65	1.33	0.04	1.0	0.06	3.07	4.73	1.12	0.03	1.0
	GMM-25	0.10	0.42	0.04	0.15	0.07	25.0	0.10	0.42	0.04	0.16	0.07	25.0
	GMM-Tuk-Han	0.11	1.44	0.65	0.51	0.02	4.5	0.10	1.26	0.47	0.44	0.01	5.0
	GMM-BR	0.10	1.91	1.87	0.68	0.03	2.4	0.10	1.64	0.84	0.58	0.01	2.6
	GMM-Trunc	0.16	2.13	17.58	0.89	0.04	1.0	0.14	2.07	2.27	0.72	0.04	2.0
	CUE-1	0.10	7.44	35.61	2.97	0.01	1.0	0.02	7.64	37.63	3.09	0.00	1.0
	CUE-25	0.04	7.43	38.80	3.10	0.81	25.0	0.07	6.81	35.27	2.91	0.80	25.0
	WWA	0.06	2.11	3.57	0.80	0.10		0.09	1.91	326.76	1.34	0.09	
0	OLS	0.10	0.23	0.02	0.11	0.19		0.10	0.11	0.01	0.10	0.62	
	GMM-1	0.14	2.54	2.64	0.95	0.02	1.0	0.05	2.48	15.19	1.06	0.01	1.0
	GMM-25	0.09	0.36	0.03	0.13	0.06	25.0	0.10	0.37	0.03	0.14	0.06	25.0
	GMM-Tuk-Han	0.09	1.71	1.57	0.64	0.03	2.1	0.09	1.88	13.23	0.78	0.01	1.6
	GMM-BR	0.10	1.80	1.64	0.67	0.02	1.1	0.09	1.85	13.31	0.79	0.01	0.8
	GMM-Trunc	0.11	1.99	1.53	0.69	0.03	1.0	0.12	2.02	1.69	0.71	0.02	1.0
	CUE-1	0.15	6.16	33.15	2.77	0.00	1.0	0.00	6.76	31.22	2.70	0.00	1.0
	CUE-25	0.05	6.23	31.32	2.70	0.82	25.0	0.07	5.97	29.29	2.62	0.82	25.0
	WWA	0.04	2.03	1.38	0.69	0.08		0.09	1.79	10.09	0.76	0.05	
0.5	OLS	0.09	0.24	0.02	0.11	0.15		0.09	0.12	0.01	0.10	0.49	
	GMM-1	0.13	3.37	7.07	1.25	0.03	1.0	0.06	3.29	3.54	1.11	0.03	1.0
	GMM-25	0.09	0.37	0.03	0.14	0.04	25.0	0.09	0.40	0.03	0.15	0.06	25.0
	GMM-Tuk-Han	0.10	1.50	0.60	0.51	0.03	4.8	0.10	1.24	0.36	0.41	0.01	5.4
	GMM-BR	0.10	1.97	3.76	0.73	0.03	2.5	0.09	1.62	1.13	0.58	0.02	2.8
	GMM-Trunc	0.10	2.06	1.19	0.68	0.04	1.0	0.13	1.74	1.53	0.65	0.03	2.0
	CUE-1	0.11	8.15	36.92	3.20	0.01	1.0	0.01	7.63	38.53	3.18	0.00	1.0
	CUE-25	0.03	7.25	31.00	2.81	0.80	25.0	0.00	7.27	36.55	3.01	0.83	25.0
	WWA	0.04	3.19	8.0E+06	95.11	0.20		0.07	1.91	76.40	1.48	0.12	
0.9	OLS	0.08	0.29	0.02	0.11	0.11		0.09	0.15	0.01	0.09	0.33	
	GMM-1	0.15	4.08	10.14	1.58	0.06	1.0	0.11	3.93	5.11	1.37	0.05	1.0
	GMM-25	0.10	0.45	0.04	0.16	0.04	25.0	0.10	0.47	0.04	0.17	0.05	25.0
	GMM-Tuk-Han	0.11	1.24	0.40	0.42	0.03	8.3	0.12	0.95	0.17	0.32	0.02	11.2
	GMM-BR	0.15	1.78	0.83	0.61	0.06	4.3	0.11	1.30	0.37	0.44	0.04	5.8
	GMM-Trunc	0.19	2.22	2.84	0.86	0.06	2.0	0.21	1.68	1.43	0.67	0.06	4.0
	CUE-1	0.12	11.10	52.35	4.03	0.01	1.0	-0.01	9.87	52.45	3.91	0.00	1.0
	CUE-25	0.08	9.45	45.79	3.59	0.80	25.0	0.05	9.41	48.00	3.65	0.83	25.0
	WWA	0.02	46.82	1.1E+07	141.13	0.53		0.01	18.57	10325.00	21.12	0.47	

Table 1b corr(u_t,v_t) = .5, ϕ = .1

θ	Estimator	Median Bias	Dec Range	MSE	MAE	Size	Median # Inst	Median Bias	Dec Range	MSE	MAE	Size	Median # Inst
		n=128						n=512					
-0.9	OLS	0.54	0.29	0.30	0.54	1.00		0.54	0.14	0.30	0.54	1.00	
	GMM-1	0.66	3.39	9.90	1.54	0.11	1.0	0.62	3.38	5.67	1.35	0.09	1.0
	GMM-25	0.54	0.47	0.32	0.54	0.80	25.0	0.54	0.46	0.33	0.54	0.82	25.0
	GMM-Tuk-Han	0.55	1.19	0.52	0.61	0.31	9.2	0.57	0.86	0.43	0.58	0.38	12.7
	GMM-BR	0.62	1.48	1.10	0.75	0.28	4.8	0.58	1.18	0.55	0.64	0.30	6.6
	GMM-Trunc	0.64	1.78	4.82	1.01	0.38	3.0	0.66	1.49	1.40	0.89	0.47	6.0
	CUE-1	0.65	9.03	43.07	3.53	0.05	1.0	0.61	7.73	33.73	3.11	0.05	1.0
	CUE-25	0.52	8.70	38.64	3.33	0.82	25.0	0.44	7.99	40.02	3.28	0.83	25.0
	WWA	0.33	2.02	7.88	0.94	0.12		0.30	1.97	6.34	0.90	0.12	
-0.5	OLS	0.52	0.23	0.28	0.52	1.00		0.52	0.11	0.27	0.52	1.00	
	GMM-1	0.58	2.68	6.32	1.20	0.10	1.0	0.54	2.56	2.78	1.08	0.09	1.0
	GMM-25	0.52	0.37	0.29	0.51	0.82	25.0	0.52	0.37	0.29	0.52	0.87	25.0
	GMM-Tuk-Han	0.52	1.20	0.65	0.63	0.28	5.2	0.52	1.05	0.50	0.58	0.26	6.2
	GMM-BR	0.56	1.57	1.35	0.78	0.22	2.7	0.55	1.29	0.81	0.69	0.20	3.2
	GMM-Trunc	0.58	1.76	2.64	0.87	0.28	2.0	0.60	1.78	1.44	0.85	0.30	2.0
	CUE-1	0.58	6.72	29.84	2.77	0.06	1.0	0.52	7.23	37.43	3.03	0.06	1.0
	CUE-25	0.48	6.52	34.40	2.92	0.83	25.0	0.48	5.88	28.08	2.62	0.85	25.0
	WWA	0.42	1.79	5.01	0.83	0.17		0.45	1.66	39.72	1.00	0.14	
0	OLS	0.49	0.20	0.25	0.49	1.00		0.49	0.10	0.25	0.50	1.00	
	GMM-1	0.52	2.25	4.63	1.04	0.08	1.0	0.44	2.17	40.49	1.13	0.06	1.0
	GMM-25	0.49	0.32	0.25	0.49	0.82	25.0	0.50	0.32	0.26	0.50	0.81	25.0
	GMM-Tuk-Han	0.49	1.52	2.69	0.80	0.30	2.2	0.50	1.56	39.18	0.96	0.24	1.6
	GMM-BR	0.49	1.55	2.75	0.82	0.22	1.2	0.48	1.58	39.26	0.97	0.18	0.8
	GMM-Trunc	0.49	1.70	2.40	0.83	0.24	1.0	0.50	1.71	1.49	0.81	0.24	1.0
	CUE-1	0.50	5.26	31.27	2.61	0.06	1.0	0.39	5.79	29.07	2.58	0.04	1.0
	CUE-25	0.46	5.38	23.71	2.36	0.84	25.0	0.45	5.37	27.04	2.47	0.86	25.0
	WWA	0.44	1.82	1.23	0.75	0.17		0.48	1.56	7.07	0.82	0.16	
0.5	OLS	0.46	0.22	0.23	0.47	1.00		0.47	0.11	0.22	0.47	1.00	
	GMM-1	0.52	2.92	5.05	1.23	0.08	1.0	0.50	2.85	3.77	1.13	0.07	1.0
	GMM-25	0.47	0.34	0.24	0.47	0.76	25.0	0.47	0.37	0.25	0.48	0.81	25.0
	GMM-Tuk-Han	0.47	1.19	0.55	0.58	0.23	5.5	0.47	1.01	0.42	0.54	0.21	6.2
	GMM-BR	0.51	1.67	2.00	0.78	0.20	2.9	0.51	1.36	0.80	0.66	0.16	3.2
	GMM-Trunc	0.51	1.76	1.32	0.80	0.24	2.0	0.54	1.67	1.80	0.85	0.25	2.0
	CUE-1	0.51	7.26	41.99	3.27	0.05	1.0	0.48	6.49	37.14	3.04	0.04	1.0
	CUE-25	0.41	6.53	32.58	2.80	0.83	25.0	0.39	6.13	31.98	2.75	0.82	25.0
	WWA	0.32	2.19	460770.00	25.30	0.20	0.0	0.36	1.63	338.35	1.56	0.13	0.0
0.9	OLS	0.44	0.27	0.21	0.45	0.99		0.45	0.14	0.21	0.45	1.00	
	GMM-1	0.52	4.02	16.51	1.59	0.08	1.0	0.50	3.64	4.75	1.34	0.07	1.0
	GMM-25	0.46	0.44	0.24	0.46	0.64	25.0	0.45	0.46	0.25	0.46	0.69	25.0
	GMM-Tuk-Han	0.50	1.13	0.45	0.56	0.24	9.2	0.49	0.85	0.34	0.51	0.27	12.4
	GMM-BR	0.53	1.51	0.81	0.70	0.22	4.8	0.51	1.15	0.50	0.59	0.22	6.5
	GMM-Trunc	0.55	1.89	11.44	1.00	0.30	4.0	0.58	1.47	1.75	0.86	0.38	5.0
	CUE-1	0.44	10.56	49.44	3.85	0.03	1.0	0.48	9.19	45.63	3.61	0.02	1.0
	CUE-25	0.44	8.81	45.13	3.52	0.81	25.0	0.37	8.48	44.65	3.41	0.81	25.0
	WWA	0.19	5.04	80234.00	21.81	0.27		0.21	2.03	238.46	1.99	0.12	

Table 1c corr(u_t,v_t) = .9, ϕ = .1

θ	Estimator	Median Bias	Dec Range	MSE	MAE	Size	Median # Inst	Median Bias	Dec Range	MSE	MAE	Size	Median # Inst
		n=128						n=512					
-0.9	OLS	0.97	0.21	0.95	0.97	1.00		0.97	0.11	0.95	0.97	1.00	
	GMM-1	1.05	2.28	6.82	1.41	0.27	1.0	1.01	2.25	5.88	1.28	0.26	1.0
	GMM-25	0.98	0.34	0.98	0.98	1.00	25.0	0.98	0.33	0.98	0.98	1.00	25.0
	GMM-Tuk-Han	0.97	0.73	1.00	0.96	0.86	12.8	0.95	0.58	0.94	0.94	0.91	16.8
	GMM-BR	0.97	0.94	1.09	0.97	0.81	6.7	0.92	0.80	0.94	0.92	0.86	8.8
	GMM-Trunc	1.03	0.91	1.63	1.11	0.75	5.0	1.02	0.71	1.26	1.05	0.82	7.0
	CUE-1	1.05	5.68	37.25	3.06	0.21	1.0	0.99	6.02	33.02	2.94	0.21	1.0
	CUE-25	1.12	4.31	22.50	2.45	0.86	25.0	1.06	4.30	21.57	2.41	0.85	25.0
	WWA	0.92	1.76	1.80	1.05	0.36		0.89	1.58	4.45	1.02	0.32	
-0.5	OLS	0.94	0.14	0.87	0.93	1.00		0.94	0.07	0.88	0.94	1.00	
	GMM-1	1.00	1.53	2.59	1.18	0.45	1.0	0.99	1.54	2.33	1.12	0.42	1.0
	GMM-25	0.93	0.22	0.88	0.94	0.99	25.0	0.93	0.22	0.89	0.94	1.00	25.0
	GMM-Tuk-Han	0.94	0.62	0.96	0.94	0.85	7.0	0.93	0.55	0.92	0.93	0.89	8.0
	GMM-BR	0.96	0.74	1.20	1.00	0.79	3.7	0.96	0.68	1.01	0.95	0.83	4.2
	GMM-Trunc	0.99	1.23	2.38	1.14	0.60	1.0	1.00	1.18	1.77	1.07	0.65	1.0
	CUE-1	1.02	3.97	20.94	2.28	0.36	1.0	1.01	3.99	22.77	2.33	0.35	1.0
	CUE-25	0.97	3.48	18.10	2.09	0.91	25.0	0.89	3.45	19.97	2.17	0.90	25.0
	WWA	0.87	1.16	2.25	0.97	0.61		0.86	1.08	1.17	0.91	0.60	
0	OLS	0.89	0.10	0.79	0.89	1.00		0.89	0.05	0.80	0.89	1.00	
	GMM-1	0.90	1.22	1.53	0.99	0.49	1.0	0.86	1.19	2.76	1.01	0.44	1.0
	GMM-25	0.89	0.16	0.79	0.89	0.97	25.0	0.89	0.17	0.80	0.89	0.97	25.0
	GMM-Tuk-Han	0.89	0.75	1.18	0.95	0.75	2.7	0.88	0.82	1.30	0.94	0.70	1.9
	GMM-BR	0.89	0.79	1.21	0.95	0.72	1.4	0.88	0.85	2.30	0.97	0.66	1.0
	GMM-Trunc	0.89	1.04	1.50	0.99	0.61	1.0	0.87	1.08	1.44	0.95	0.56	1.0
	CUE-1	0.90	2.64	13.32	1.75	0.45	1.0	0.81	3.10	20.85	2.02	0.42	1.0
	CUE-25	0.86	2.60	13.16	1.71	0.92	25.0	0.85	2.65	20.29	1.98	0.91	25.0
	WWA	0.85	1.01	1.14	0.92	0.76		0.88	0.85	1.81	0.95	0.77	
0.5	OLS	0.85	0.16	0.72	0.85	1.00		0.85	0.08	0.72	0.85	1.00	
	GMM-1	0.89	1.85	5.16	1.16	0.28	1.0	0.84	1.77	2.34	1.08	0.27	1.0
	GMM-25	0.85	0.25	0.73	0.85	0.98	25.0	0.85	0.27	0.74	0.85	1.00	25.0
	GMM-Tuk-Han	0.85	0.75	0.90	0.87	0.74	7.0	0.84	0.70	0.81	0.85	0.77	7.4
	GMM-BR	0.88	1.02	1.29	0.95	0.68	3.7	0.88	0.90	0.96	0.89	0.72	3.8
	GMM-Trunc	0.92	1.55	2.22	1.06	0.48	1.0	0.90	1.38	1.86	1.01	0.51	1.0
	CUE-1	0.87	4.40	22.30	2.34	0.23	1.0	0.79	5.07	31.10	2.67	0.22	1.0
	CUE-25	0.78	3.45	16.61	2.00	0.87	25.0	0.86	3.42	18.67	2.09	0.89	25.0
	WWA	0.68	1.27	1.52	0.85	0.39		0.73	1.11	7.64	0.90	0.40	
0.9	OLS	0.81	0.24	0.67	0.81	1.00		0.81	0.13	0.66	0.81	1.00	
	GMM-1	0.85	2.66	3.67	1.29	0.15	1.0	0.86	2.62	6.28	1.33	0.15	1.0
	GMM-25	0.82	0.37	0.70	0.82	0.99	25.0	0.83	0.41	0.71	0.83	1.00	25.0
	GMM-Tuk-Han	0.79	0.82	0.77	0.81	0.69	12.0	0.80	0.72	0.72	0.80	0.76	15.5
	GMM-BR	0.80	1.12	0.93	0.85	0.63	6.2	0.77	0.96	0.74	0.78	0.67	8.1
	GMM-Trunc	0.87	1.24	1.54	0.98	0.62	5.0	0.89	0.88	1.19	0.96	0.72	7.0
	CUE-1	0.84	6.86	33.29	3.01	0.11	1.0	0.82	6.64	35.78	3.15	0.10	1.0
	CUE-25	0.81	4.56	21.97	2.35	0.78	25.0	0.95	4.62	23.95	2.50	0.78	25.0
	WWA	0.57	1.64	1.41	0.82	0.14		0.61	1.55	5.35	0.86	0.13	

Table 2a

corr(u_t,v_t) = .1, ϕ = .3

θ	Estimator	n=128 Median Bias	Dec Range	MSE	MAE	Size	Median # Inst	n=512 Median Bias	Dec Range	MSE	MAE	Size	Median # Inst
-0.9	OLS	0.12	0.33	0.03	0.14	0.22		0.12	0.16	0.02	0.12	0.55	
	GMM-1	0.12	3.17	6.76	1.20	0.06	1.0	0.01	2.00	1.10	0.66	0.07	1.0
	GMM-25	0.11	0.53	0.05	0.19	0.10	25.0	0.10	0.52	0.05	0.18	0.10	25.0
	GMM-Tuk-Han	0.14	1.27	0.37	0.44	0.04	6.9	0.08	1.02	0.18	0.33	0.04	9.1
	GMM-BR	0.15	1.68	0.93	0.59	0.07	3.6	0.07	1.24	0.30	0.41	0.05	4.7
	GMM-Trunc	0.19	2.11	4.49	0.84	0.07	1.0	0.13	1.49	0.58	0.51	0.09	2.0
	CUE-1	0.10	6.86	38.98	3.13	0.02	1.0	-0.03	3.04	16.91	1.61	0.04	1.0
	CUE-25	0.02	8.03	40.89	3.29	0.82	25.0	-0.01	6.52	36.18	2.88	0.79	25.0
	WWA	0.06	2.13	31.36	1.06	0.12		0.01	1.59	105.45	1.16	0.13	
-0.5	OLS	0.10	0.27	0.02	0.12	0.23		0.11	0.14	0.01	0.11	0.61	
	GMM-1	0.11	2.55	3.76	0.93	0.05	1.0	0.03	1.61	0.72	0.54	0.03	1.0
	GMM-25	0.10	0.43	0.04	0.16	0.09	25.0	0.09	0.43	0.04	0.15	0.11	25.0
	GMM-Tuk-Han	0.08	1.25	0.50	0.43	0.03	4.6	0.06	0.97	0.19	0.32	0.02	5.9
	GMM-BR	0.08	1.58	1.14	0.58	0.04	2.4	0.07	1.16	0.30	0.37	0.03	3.0
	GMM-Trunc	0.10	1.75	1.09	0.61	0.04	2.0	0.08	1.19	0.47	0.41	0.04	2.0
	CUE-1	0.10	5.64	27.43	2.51	0.01	1.0	-0.01	2.41	11.62	1.27	0.02	1.0
	CUE-25	0.09	6.60	32.94	2.82	0.81	25.0	0.02	5.20	24.23	2.28	0.77	25.0
	WWA	0.06	1.87	6.36	0.80	0.08		0.05	1.29	0.58	0.45	0.06	
0	OLS	0.09	0.21	0.02	0.10	0.18		0.09	0.11	0.01	0.09	0.58	
	GMM-1	0.10	2.23	2.07	0.80	0.02	1.0	0.01	1.24	1.18	0.46	0.01	1.0
	GMM-25	0.09	0.34	0.03	0.13	0.05	25.0	0.08	0.34	0.03	0.13	0.08	25.0
	GMM-Tuk-Han	0.07	1.28	0.76	0.47	0.03	4.1	0.04	0.90	0.16	0.29	0.01	4.6
	GMM-BR	0.08	1.43	1.06	0.55	0.02	2.1	0.04	1.03	0.93	0.37	0.01	2.4
	GMM-Trunc	0.09	1.48	0.98	0.54	0.03	2.0	0.04	1.01	0.22	0.33	0.03	2.0
	CUE-1	0.09	4.34	22.64	2.07	0.00	1.0	-0.01	1.86	6.66	0.97	0.01	1.0
	CUE-25	0.06	5.78	27.37	2.47	0.81	25.0	0.02	4.44	19.31	1.92	0.77	25.0
	WWA	0.07	1.74	4.34	0.67	0.07		0.03	1.17	0.30	0.37	0.04	
0.5	OLS	0.08	0.21	0.01	0.09	0.10		0.08	0.11	0.01	0.08	0.36	
	GMM-1	0.09	2.57	4.19	0.97	0.03	1.0	0.00	1.36	0.87	0.51	0.02	1.0
	GMM-25	0.08	0.33	0.02	0.12	0.03	25.0	0.07	0.34	0.02	0.12	0.04	25.0
	GMM-Tuk-Han	0.06	1.28	0.52	0.44	0.01	5.5	0.05	0.83	0.16	0.27	0.00	7.0
	GMM-BR	0.07	1.57	1.66	0.59	0.02	2.9	0.05	0.93	0.24	0.32	0.01	3.6
	GMM-Trunc	0.08	1.63	2.34	0.61	0.02	2.0	0.05	0.97	0.26	0.32	0.01	3.0
	CUE-1	0.07	6.17	33.43	2.68	0.01	1.0	-0.02	2.36	11.04	1.29	0.01	1.0
	CUE-25	0.05	6.71	36.85	2.87	0.77	25.0	0.02	5.40	29.16	2.44	0.77	25.0
	WWA	0.01	2.35	2193.70	4.62	0.21		0.01	1.05	2.36	0.44	0.08	
0.9	OLS	0.06	0.24	0.01	0.09	0.05		0.07	0.13	0.01	0.07	0.19	
	GMM-1	0.12	3.28	9.49	1.25	0.04	1.0	-0.02	1.71	1.43	0.63	0.03	1.0
	GMM-25	0.07	0.40	0.03	0.14	0.03	25.0	0.07	0.39	0.03	0.14	0.03	25.0
	GMM-Tuk-Han	0.07	1.11	0.25	0.36	0.02	9.1	0.06	0.69	0.09	0.23	0.00	12.7
	GMM-BR	0.07	1.48	0.59	0.49	0.04	4.7	0.07	0.88	0.15	0.29	0.02	6.7
	GMM-Trunc	0.12	1.64	1.30	0.62	0.04	3.0	0.09	0.85	0.24	0.31	0.02	6.0
	CUE-1	0.11	7.76	43.48	3.36	0.01	1.0	-0.06	3.22	19.67	1.74	0.02	1.0
	CUE-25	0.08	9.01	48.93	3.60	0.78	25.0	0.02	7.53	36.77	2.96	0.76	25.0
	WWA	-0.01	39.90	87377.00	36.23	0.54		0.00	14.08	4641.30	12.86	0.44	

Table 2b corr(u_t,v_t) = .5, ϕ = .3

θ	Estimator	Median Bias (n=128)	Dec Range (n=128)	MSE (n=128)	MAE (n=128)	Size (n=128)	Median # Inst (n=128)	Median Bias (n=512)	Dec Range (n=512)	MSE (n=512)	MAE (n=512)	Size (n=512)	Median # Inst (n=512)
-0.9	OLS	0.58	0.29	0.34	0.58	1.00		0.58	0.14	0.34	0.58	1.00	
	GMM-1	0.50	2.94	8.51	1.22	0.12	1.0	0.16	1.92	33.95	0.83	0.10	1.0
	GMM-25	0.57	0.46	0.36	0.57	0.84	25.0	0.54	0.45	0.33	0.54	0.86	25.0
	GMM-Tuk-Han	0.56	1.01	0.48	0.60	0.37	9.1	0.47	0.91	0.32	0.49	0.39	12.0
	GMM-BR	0.58	1.30	0.87	0.69	0.31	4.7	0.44	1.09	0.37	0.50	0.30	6.2
	GMM-Trunc	0.64	1.51	2.07	0.85	0.43	4.0	0.50	1.37	33.72	0.81	0.47	5.0
	CUE-1	0.44	7.03	37.29	3.03	0.08	1.0	0.02	3.06	13.03	1.47	0.10	1.0
	CUE-25	0.49	8.42	40.73	3.31	0.82	25.0	0.25	6.53	26.56	2.54	0.78	25.0
	WWA	0.30	1.95	26.57	1.05	0.14		0.16	1.40	42.09	0.76	0.10	
-0.5	OLS	0.52	0.23	0.28	0.52	1.00		0.53	0.12	0.28	0.53	1.00	
	GMM-1	0.43	2.30	10.78	1.00	0.11	1.0	0.13	1.47	0.65	0.52	0.08	1.0
	GMM-25	0.52	0.37	0.29	0.51	0.86	25.0	0.48	0.37	0.26	0.49	0.88	25.0
	GMM-Tuk-Han	0.48	1.02	0.54	0.56	0.30	5.1	0.35	0.92	0.25	0.41	0.28	6.2
	GMM-BR	0.47	1.41	0.93	0.66	0.24	2.7	0.33	1.07	0.32	0.45	0.21	3.3
	GMM-Trunc	0.49	1.50	0.93	0.69	0.30	2.0	0.34	1.27	0.45	0.50	0.29	2.0
	CUE-1	0.37	5.35	24.39	2.36	0.08	1.0	0.00	2.43	12.79	1.33	0.09	1.0
	CUE-25	0.46	6.35	32.41	2.76	0.83	25.0	0.23	4.87	25.99	2.28	0.80	25.0
	WWA	0.37	1.63	3.02	0.78	0.16		0.19	1.20	0.48	0.45	0.10	
0	OLS	0.45	0.19	0.21	0.45	1.00		0.46	0.10	0.21	0.46	1.00	
	GMM-1	0.35	1.99	1.65	0.79	0.05	1.0	0.10	1.19	0.77	0.46	0.05	1.0
	GMM-25	0.45	0.30	0.22	0.45	0.81	25.0	0.42	0.31	0.20	0.43	0.83	25.0
	GMM-Tuk-Han	0.40	1.13	0.71	0.57	0.29	4.1	0.26	0.91	0.20	0.37	0.24	4.3
	GMM-BR	0.39	1.40	1.10	0.65	0.21	2.1	0.20	1.09	0.50	0.42	0.18	2.3
	GMM-Trunc	0.41	1.42	0.75	0.61	0.25	2.0	0.25	1.01	0.24	0.39	0.24	2.0
	CUE-1	0.30	4.19	19.67	1.99	0.06	1.0	0.01	1.81	7.53	0.97	0.05	1.0
	CUE-25	0.37	5.00	25.85	2.33	0.83	25.0	0.20	4.06	19.64	1.90	0.77	25.0
	WWA	0.34	1.49	13.27	0.75	0.14		0.18	1.09	0.29	0.38	0.09	
0.5	OLS	0.38	0.20	0.16	0.39	1.00		0.39	0.10	0.15	0.39	1.00	
	GMM-1	0.32	2.49	5.47	0.99	0.05	1.0	0.09	1.33	0.83	0.50	0.02	1.0
	GMM-25	0.39	0.31	0.16	0.39	0.67	25.0	0.36	0.32	0.15	0.37	0.70	25.0
	GMM-Tuk-Han	0.34	1.11	0.43	0.48	0.15	5.6	0.24	0.82	0.19	0.34	0.14	6.4
	GMM-BR	0.37	1.44	0.78	0.61	0.13	2.9	0.24	0.99	0.27	0.39	0.10	3.4
	GMM-Trunc	0.38	1.59	3.79	0.73	0.16	2.0	0.26	1.07	0.35	0.43	0.17	2.0
	CUE-1	0.25	5.49	31.35	2.62	0.04	1.0	0.00	2.30	10.70	1.24	0.03	1.0
	CUE-25	0.34	6.24	33.80	2.78	0.79	25.0	0.17	5.12	24.73	2.24	0.74	25.0
	WWA	0.17	1.84	1919.90	4.46	0.17		0.09	0.96	0.66	0.38	0.07	
0.9	OLS	0.33	0.24	0.12	0.33	0.92		0.33	0.13	0.11	0.33	1.00	
	GMM-1	0.30	3.21	3.91	1.17	0.05	1.0	0.07	1.69	1.30	0.62	0.02	1.0
	GMM-25	0.34	0.40	0.14	0.34	0.40	25.0	0.31	0.39	0.13	0.33	0.41	25.0
	GMM-Tuk-Han	0.33	1.05	0.33	0.44	0.13	8.8	0.28	0.77	0.16	0.33	0.11	11.2
	GMM-BR	0.36	1.52	0.71	0.59	0.13	4.5	0.27	1.02	0.23	0.39	0.09	5.8
	GMM-Trunc	0.39	1.80	1.74	0.77	0.17	3.0	0.33	1.22	0.60	0.50	0.18	3.0
	CUE-1	0.24	6.67	34.73	2.97	0.02	1.0	-0.01	3.02	16.60	1.55	0.03	1.0
	CUE-25	0.34	8.53	40.44	3.33	0.78	25.0	0.13	7.15	36.81	2.87	0.73	25.0
	WWA	0.03	5.33	9371.70	13.37	0.28		0.02	1.14	219.10	1.17	0.10	

Table 2c $corr(u_t,v_t) = .9$, $\phi = .3$

θ	Estimator	Median Bias	Dec Range	MSE	MAE	Size	Median # Inst	Median Bias	Dec Range	MSE	MAE	Size	Median # Inst
		n=128						n=512					
-0.9	OLS	1.04	0.17	1.08	1.04	1.00		1.04	0.09	1.09	1.04	1.00	
	GMM-1	0.74	2.37	3.53	1.19	0.37	1.0	0.24	1.69	1.10	0.64	0.25	1.0
	GMM-25	1.04	0.29	1.08	1.03	1.00	25.0	0.98	0.28	0.97	0.98	1.00	25.0
	GMM-Tuk-Han	0.98	0.63	1.00	0.96	0.91	14.3	0.82	0.65	0.70	0.79	0.87	16.0
	GMM-BR	0.96	0.75	1.04	0.95	0.88	7.5	0.78	0.73	0.65	0.75	0.83	8.3
	GMM-Trunc	1.01	1.05	1.85	1.06	0.78	5.0	0.83	1.34	0.89	0.80	0.68	4.0
	CUE-1	0.50	7.04	37.37	2.96	0.24	1.0	-0.03	2.81	15.92	1.52	0.13	1.0
	CUE-25	0.96	4.63	25.05	2.46	0.87	25.0	0.35	4.74	25.03	2.22	0.77	25.0
	WWA	0.79	1.67	4.51	1.04	0.50		0.40	1.28	0.67	0.58	0.36	
-0.5	OLS	0.94	0.12	0.89	0.94	1.00		0.94	0.06	0.89	0.94	1.00	
	GMM-1	0.67	1.73	2.75	0.97	0.44	1.0	0.23	1.28	1.65	0.55	0.27	1.0
	GMM-25	0.93	0.20	0.88	0.93	0.99	25.0	0.88	0.21	0.79	0.88	1.00	25.0
	GMM-Tuk-Han	0.86	0.64	0.79	0.84	0.85	7.0	0.59	0.72	0.42	0.59	0.71	6.2
	GMM-BR	0.86	0.82	0.89	0.85	0.79	3.7	0.51	0.87	0.38	0.54	0.61	3.2
	GMM-Trunc	0.79	1.39	2.41	0.95	0.59	1.0	0.29	1.39	0.76	0.54	0.39	1.0
	CUE-1	0.43	5.53	30.20	2.50	0.30	1.0	-0.04	2.22	11.58	1.26	0.12	1.0
	CUE-25	0.84	3.98	24.05	2.29	0.89	25.0	0.29	4.18	24.50	2.06	0.77	25.0
	WWA	0.69	1.18	2.65	0.84	0.64		0.35	0.94	0.38	0.46	0.46	
0	OLS	0.82	0.11	0.67	0.82	1.00		0.82	0.06	0.67	0.82	1.00	
	GMM-1	0.56	1.45	1.11	0.75	0.30	1.0	0.17	1.09	0.63	0.44	0.15	1.0
	GMM-25	0.81	0.18	0.66	0.81	0.94	25.0	0.76	0.19	0.60	0.77	0.95	25.0
	GMM-Tuk-Han	0.73	0.75	0.74	0.75	0.69	4.2	0.44	0.81	0.30	0.47	0.50	3.6
	GMM-BR	0.70	1.00	0.83	0.76	0.62	2.2	0.36	1.00	0.47	0.48	0.43	1.9
	GMM-Trunc	0.64	1.23	0.89	0.74	0.46	1.0	0.25	1.08	0.42	0.43	0.27	1.0
	CUE-1	0.36	4.34	22.46	2.05	0.26	1.0	-0.02	1.76	13.34	1.19	0.09	1.0
	CUE-25	0.68	3.09	19.20	1.94	0.90	25.0	0.28	3.89	20.37	1.81	0.78	25.0
	WWA	0.60	1.12	38.37	0.94	0.57		0.30	0.88	0.28	0.40	0.42	
0.5	OLS	0.70	0.17	0.49	0.70	1.00		0.70	0.09	0.49	0.70	1.00	
	GMM-1	0.52	1.90	2.70	0.91	0.12	1.0	0.19	1.31	3.71	0.57	0.08	1.0
	GMM-25	0.69	0.27	0.49	0.70	0.97	25.0	0.65	0.29	0.45	0.66	0.99	25.0
	GMM-Tuk-Han	0.63	0.84	0.57	0.66	0.56	6.4	0.45	0.73	0.28	0.46	0.43	6.3
	GMM-BR	0.62	1.20	1.52	0.75	0.51	3.3	0.43	0.98	0.36	0.50	0.42	3.3
	GMM-Trunc	0.64	1.58	1.47	0.82	0.34	1.0	0.28	1.34	3.31	0.58	0.24	1.0
	CUE-1	0.38	4.99	24.24	2.19	0.11	1.0	0.03	2.04	12.79	1.23	0.05	1.0
	CUE-25	0.49	4.51	19.67	2.14	0.83	25.0	0.29	4.56	26.57	2.23	0.75	25.0
	WWA	0.37	1.19	57.07	0.91	0.17		0.19	0.83	0.92	0.37	0.11	
0.9	OLS	0.60	0.24	0.37	0.60	1.00		0.60	0.13	0.36	0.60	1.00	
	GMM-1	0.45	2.53	3.24	1.05	0.07	1.0	0.17	1.71	2.31	0.67	0.05	1.0
	GMM-25	0.60	0.37	0.39	0.61	0.97	25.0	0.57	0.39	0.36	0.58	0.97	25.0
	GMM-Tuk-Han	0.55	0.86	0.46	0.59	0.41	11.2	0.46	0.69	0.28	0.46	0.39	13.9
	GMM-BR	0.58	1.18	0.63	0.65	0.45	5.8	0.43	0.91	0.32	0.48	0.39	7.2
	GMM-Trunc	0.60	1.40	1.65	0.81	0.41	4.0	0.51	1.24	0.68	0.60	0.43	4.0
	CUE-1	0.35	5.52	27.97	2.54	0.06	1.0	0.07	2.60	11.65	1.34	0.04	1.0
	CUE-25	0.45	5.28	30.52	2.73	0.71	25.0	0.25	5.12	25.85	2.38	0.68	25.0
	WWA	0.18	1.55	10.25	0.77	0.06		0.09	0.99	0.39	0.37	0.04	

Table 3a $corr(u_t,v_t) = .1, \phi = .5$

θ	Estimator	Median Bias	Dec Range	MSE	MAE	Size	Median # Inst	Median Bias	Dec Range	MSE	MAE	Size	Median # Inst
		n=128						n=512					
-0.9	OLS	0.11	0.32	0.03	0.13	0.25		0.11	0.16	0.02	0.11	0.56	
	GMM-1	0.06	1.55	0.78	0.53	0.08	1.0	0.00	0.65	0.07	0.20	0.09	1.0
	GMM-25	0.10	0.49	0.05	0.17	0.13	25.0	0.07	0.40	0.03	0.14	0.13	25.0
	GMM-Tuk-Han	0.07	0.94	0.17	0.31	0.06	6.5	0.02	0.59	0.05	0.18	0.09	8.7
	GMM-BR	0.08	1.16	0.28	0.38	0.10	3.4	0.01	0.60	0.06	0.19	0.09	4.5
	GMM-Trunc	0.08	1.27	0.49	0.44	0.09	1.0	0.02	0.59	0.06	0.19	0.09	3.0
	CUE-1	0.05	2.22	8.05	1.06	0.06	1.0	-0.01	0.68	0.08	0.22	0.10	1.0
	CUE-25	0.02	5.36	28.29	2.40	0.77	25.0	-0.03	1.32	3.05	0.60	0.53	25.0
	WWA	0.04	1.29	3.41	0.53	0.10		0.00	0.61	0.10	0.21	0.08	
-0.5	OLS	0.09	0.25	0.02	0.11	0.23		0.10	0.13	0.01	0.10	0.60	
	GMM-1	0.05	1.24	0.53	0.43	0.06	1.0	-0.01	0.51	0.04	0.16	0.07	1.0
	GMM-25	0.09	0.39	0.03	0.14	0.12	25.0	0.05	0.32	0.02	0.11	0.12	25.0
	GMM-Tuk-Han	0.05	0.81	0.17	0.27	0.06	6.5	0.01	0.46	0.03	0.15	0.08	8.5
	GMM-BR	0.05	0.91	0.36	0.32	0.07	3.4	0.01	0.49	0.04	0.15	0.08	4.4
	GMM-Trunc	0.07	0.94	0.23	0.32	0.07	3.0	0.01	0.48	0.04	0.15	0.09	5.0
	CUE-1	0.04	1.69	5.18	0.81	0.06	1.0	-0.01	0.54	0.05	0.17	0.07	1.0
	CUE-25	0.05	3.98	20.18	1.93	0.76	25.0	-0.02	1.06	3.33	0.51	0.52	25.0
	WWA	0.02	1.09	1.52	0.40	0.07		0.00	0.48	0.04	0.15	0.04	
0	OLS	0.08	0.19	0.01	0.09	0.16		0.08	0.10	0.01	0.08	0.51	
	GMM-1	0.02	0.95	0.27	0.33	0.03	1.0	-0.01	0.38	0.03	0.12	0.02	1.0
	GMM-25	0.07	0.30	0.02	0.11	0.07	25.0	0.05	0.26	0.01	0.09	0.07	25.0
	GMM-Tuk-Han	0.05	0.68	0.11	0.23	0.03	7.6	0.01	0.36	0.02	0.11	0.03	10.1
	GMM-BR	0.04	0.73	0.16	0.26	0.03	4.0	0.00	0.37	0.02	0.11	0.03	5.3
	GMM-Trunc	0.05	0.69	0.13	0.23	0.03	4.0	0.02	0.37	0.02	0.11	0.04	6.0
	CUE-1	0.01	1.24	4.33	0.64	0.03	1.0	-0.01	0.40	0.03	0.13	0.03	1.0
	CUE-25	0.06	3.39	15.53	1.63	0.73	25.0	0.00	0.79	2.46	0.38	0.45	25.0
	WWA	0.02	0.91	0.32	0.32	0.06		0.00	0.39	0.02	0.12	0.04	
0.5	OLS	0.06	0.17	0.01	0.07	0.05		0.06	0.09	0.00	0.06	0.23	
	GMM-1	0.01	0.98	0.40	0.35	0.01	1.0	-0.01	0.36	0.02	0.11	0.00	1.0
	GMM-25	0.05	0.27	0.01	0.10	0.02	25.0	0.04	0.23	0.01	0.08	0.02	25.0
	GMM-Tuk-Han	0.03	0.65	0.11	0.22	0.02	8.2	0.00	0.32	0.02	0.10	0.00	11.1
	GMM-BR	0.03	0.71	0.20	0.26	0.02	4.3	0.00	0.33	0.02	0.10	0.00	5.7
	GMM-Trunc	0.04	0.66	0.25	0.25	0.01	5.0	0.01	0.33	0.02	0.10	0.01	6.0
	CUE-1	0.00	1.32	3.83	0.65	0.02	1.0	-0.01	0.38	0.03	0.12	0.01	1.0
	CUE-25	0.04	4.51	27.87	2.16	0.67	25.0	0.00	0.80	3.31	0.48	0.42	25.0
	WWA	-0.01	1.07	539.74	2.85	0.19		-0.01	0.31	0.02	0.10	0.05	
0.9	OLS	0.04	0.19	0.01	0.07	0.01		0.04	0.10	0.00	0.05	0.07	
	GMM-1	0.01	1.13	0.48	0.41	0.01	1.0	-0.01	0.40	0.03	0.13	0.01	1.0
	GMM-25	0.04	0.30	0.02	0.10	0.01	25.0	0.03	0.25	0.01	0.08	0.02	25.0
	GMM-Tuk-Han	0.04	0.67	0.09	0.22	0.01	10.3	0.01	0.34	0.02	0.11	0.00	14.4
	GMM-BR	0.04	0.78	0.15	0.27	0.02	5.3	0.01	0.37	0.02	0.12	0.00	7.6
	GMM-Trunc	0.03	0.77	0.24	0.27	0.02	5.0	0.00	0.35	0.02	0.11	0.01	7.0
	CUE-1	0.00	1.72	5.90	0.86	0.02	1.0	-0.01	0.43	0.11	0.15	0.01	1.0
	CUE-25	0.05	5.84	30.28	2.47	0.69	25.0	-0.01	1.01	5.98	0.64	0.40	25.0
	WWA	-0.03	32.22	33962.00	25.04	0.53		0.00	4.71	5146.40	7.63	0.41	

Table 3b corr(u_t,v_t) = .5, ϕ = .5

θ	Estimator	n=128 Median Bias	Dec Range	MSE	MAE	Size	Median # Inst	n=512 Median Bias	Dec Range	MSE	MAE	Size	Median # Inst
-0.9	OLS	0.55	0.28	0.31	0.54	1.00		0.55	0.14	0.30	0.55	1.00	
	GMM-1	0.13	1.44	2.52	0.57	0.11	1.0	0.00	0.65	0.07	0.21	0.10	1.0
	GMM-25	0.50	0.43	0.28	0.50	0.84	25.0	0.35	0.37	0.14	0.35	0.72	25.0
	GMM-Tuk-Han	0.33	0.93	0.23	0.40	0.31	7.8	0.08	0.62	0.06	0.20	0.18	7.1
	GMM-BR	0.28	1.06	1.12	0.45	0.28	4.0	0.06	0.65	0.07	0.21	0.17	3.7
	GMM-Trunc	0.33	1.27	2.25	0.55	0.36	2.0	0.04	0.74	0.09	0.24	0.23	1.0
	CUE-1	0.01	2.16	13.04	1.19	0.11	1.0	-0.03	0.70	0.10	0.23	0.09	1.0
	CUE-25	0.20	5.11	26.21	2.32	0.77	25.0	-0.01	1.13	4.01	0.57	0.51	25.0
	WWA	0.11	1.20	2.04	0.51	0.12		0.02	0.58	0.06	0.18	0.08	
-0.5	OLS	0.47	0.22	0.23	0.47	1.00		0.47	0.12	0.22	0.47	1.00	
	GMM-1	0.10	1.22	0.66	0.42	0.10	1.0	0.00	0.51	0.05	0.17	0.08	1.0
	GMM-25	0.43	0.35	0.21	0.43	0.84	25.0	0.30	0.29	0.10	0.30	0.73	25.0
	GMM-Tuk-Han	0.25	0.76	0.16	0.33	0.26	5.8	0.06	0.49	0.04	0.16	0.15	5.7
	GMM-BR	0.22	0.90	0.27	0.36	0.22	3.0	0.04	0.51	0.04	0.16	0.14	3.0
	GMM-Trunc	0.24	0.99	0.29	0.39	0.29	3.0	0.06	0.53	0.05	0.18	0.19	3.0
	CUE-1	0.01	1.66	10.20	0.97	0.10	1.0	-0.02	0.56	0.06	0.18	0.07	1.0
	CUE-25	0.18	3.89	21.92	1.94	0.77	25.0	-0.02	0.91	1.51	0.41	0.48	25.0
	WWA	0.13	1.02	0.42	0.38	0.12		0.02	0.48	0.04	0.15	0.07	
0	OLS	0.37	0.18	0.15	0.38	1.00		0.38	0.09	0.14	0.38	1.00	
	GMM-1	0.05	0.93	0.27	0.32	0.05	1.0	-0.01	0.39	0.03	0.13	0.04	1.0
	GMM-25	0.35	0.27	0.13	0.35	0.80	25.0	0.24	0.23	0.07	0.24	0.68	25.0
	GMM-Tuk-Han	0.20	0.68	0.13	0.28	0.24	6.8	0.06	0.38	0.03	0.13	0.12	6.7
	GMM-BR	0.17	0.78	0.18	0.30	0.18	3.5	0.04	0.39	0.03	0.13	0.09	3.5
	GMM-Trunc	0.20	0.74	0.15	0.30	0.26	4.0	0.07	0.39	0.03	0.14	0.15	4.0
	CUE-1	0.00	1.28	5.26	0.66	0.05	1.0	-0.02	0.41	0.03	0.13	0.03	1.0
	CUE-25	0.14	3.26	18.96	1.73	0.74	25.0	-0.01	0.71	1.51	0.34	0.42	25.0
	WWA	0.09	0.87	0.27	0.31	0.10		0.01	0.39	0.02	0.12	0.06	
0.5	OLS	0.28	0.17	0.09	0.28	0.98		0.28	0.09	0.08	0.28	1.00	
	GMM-1	0.05	0.91	0.30	0.33	0.02	1.0	0.00	0.36	0.02	0.12	0.01	1.0
	GMM-25	0.26	0.26	0.08	0.27	0.49	25.0	0.18	0.22	0.04	0.18	0.32	25.0
	GMM-Tuk-Han	0.15	0.64	0.13	0.26	0.09	7.0	0.04	0.33	0.02	0.11	0.03	6.7
	GMM-BR	0.13	0.76	0.19	0.29	0.07	3.7	0.03	0.35	0.02	0.11	0.02	3.5
	GMM-Trunc	0.15	0.77	0.16	0.29	0.10	3.0	0.05	0.35	0.02	0.12	0.04	4.0
	CUE-1	0.00	1.28	5.77	0.75	0.02	1.0	-0.01	0.38	0.03	0.12	0.01	1.0
	CUE-25	0.12	4.41	25.05	2.07	0.66	25.0	0.00	0.77	3.22	0.44	0.35	25.0
	WWA	0.02	0.86	520.48	1.81	0.13		-0.01	0.31	0.02	0.10	0.05	
0.9	OLS	0.20	0.19	0.05	0.21	0.55		0.21	0.10	0.04	0.21	1.00	
	GMM-1	0.03	1.07	0.45	0.40	0.01	1.0	0.00	0.42	0.03	0.13	0.00	1.0
	GMM-25	0.20	0.32	0.06	0.21	0.13	25.0	0.13	0.25	0.03	0.14	0.09	25.0
	GMM-Tuk-Han	0.14	0.70	0.12	0.26	0.05	8.7	0.03	0.39	0.03	0.12	0.02	9.3
	GMM-BR	0.13	0.86	0.17	0.30	0.05	4.5	0.03	0.41	0.03	0.13	0.02	4.8
	GMM-Trunc	0.15	0.86	0.26	0.32	0.06	3.0	0.02	0.44	0.03	0.14	0.03	2.0
	CUE-1	0.00	1.53	9.39	0.98	0.02	1.0	-0.01	0.44	0.04	0.14	0.01	1.0
	CUE-25	0.10	6.36	36.53	2.70	0.67	25.0	0.01	0.92	5.88	0.61	0.34	25.0
	WWA	-0.04	6.39	7434.10	13.43	0.33		-0.01	0.35	121.68	0.75	0.11	

Table 3c

corr(u_t,v_t) = .9, ϕ = .5

θ	Estimator	Median Bias (n=128)	Dec Range (n=128)	MSE (n=128)	MAE (n=128)	Size (n=128)	Median # Inst (n=128)	Median Bias (n=512)	Dec Range (n=512)	MSE (n=512)	MAE (n=512)	Size (n=512)	Median # Inst (n=512)
-0.9	OLS	0.98	0.16	0.96	0.98	1.00		0.98	0.08	0.96	0.98	1.00	
	GMM-1	0.14	1.31	1.93	0.56	0.22	1.0	0.00	0.64	0.09	0.21	0.14	1.0
	GMM-25	0.91	0.26	0.83	0.91	1.00	25.0	0.63	0.25	0.40	0.63	1.00	25.0
	GMM-Tuk-Han	0.57	0.77	0.41	0.58	0.73	10.5	0.15	0.54	0.07	0.21	0.31	6.8
	GMM-BR	0.54	0.91	0.40	0.56	0.68	5.4	0.13	0.64	0.07	0.22	0.32	3.5
	GMM-Trunc	0.40	1.39	0.56	0.58	0.53	2.0	0.01	0.67	0.09	0.22	0.22	1.0
	CUE-1	-0.07	2.08	14.27	1.25	0.10	1.0	-0.04	0.70	0.15	0.24	0.07	1.0
	CUE-25	0.13	4.17	21.17	1.95	0.75	25.0	-0.05	0.81	0.89	0.31	0.36	25.0
	WWA	0.23	1.07	0.41	0.43	0.36		0.04	0.58	0.06	0.19	0.28	
-0.5	OLS	0.84	0.13	0.72	0.85	1.00		0.85	0.06	0.72	0.85	1.00	
	GMM-1	0.12	1.06	1.35	0.45	0.20	1.0	0.01	0.51	0.06	0.17	0.11	1.0
	GMM-25	0.78	0.22	0.62	0.78	0.99	25.0	0.54	0.19	0.30	0.54	1.00	25.0
	GMM-Tuk-Han	0.40	0.68	0.23	0.42	0.58	5.7	0.10	0.42	0.04	0.16	0.24	5.5
	GMM-BR	0.34	0.82	0.27	0.41	0.51	2.9	0.07	0.47	0.04	0.16	0.20	2.9
	GMM-Trunc	0.17	1.17	1.28	0.47	0.34	1.0	0.03	0.53	0.05	0.17	0.19	2.0
	CUE-1	-0.05	1.66	9.65	0.97	0.10	1.0	-0.03	0.57	0.09	0.19	0.07	1.0
	CUE-25	0.09	3.00	14.73	1.47	0.73	25.0	-0.04	0.65	0.28	0.24	0.34	25.0
	WWA	0.21	0.89	4.75	0.44	0.42		0.04	0.47	0.04	0.15	0.30	
0	OLS	0.68	0.13	0.46	0.68	1.00		0.68	0.06	0.46	0.68	1.00	
	GMM-1	0.09	0.84	0.33	0.32	0.11	1.0	0.00	0.39	0.04	0.13	0.04	1.0
	GMM-25	0.62	0.20	0.40	0.63	0.97	25.0	0.43	0.17	0.19	0.43	0.99	25.0
	GMM-Tuk-Han	0.35	0.61	0.19	0.37	0.50	5.9	0.11	0.37	0.03	0.15	0.22	6.4
	GMM-BR	0.29	0.75	0.24	0.37	0.44	3.0	0.08	0.38	0.03	0.14	0.17	3.3
	GMM-Trunc	0.24	0.82	0.21	0.35	0.33	2.0	0.08	0.38	0.03	0.14	0.17	3.0
	CUE-1	-0.02	1.22	6.08	0.70	0.07	1.0	-0.02	0.43	0.06	0.15	0.03	1.0
	CUE-25	0.14	2.44	15.40	1.46	0.72	25.0	-0.02	0.57	0.80	0.23	0.32	25.0
	WWA	0.15	0.71	0.20	0.29	0.25		0.02	0.37	0.03	0.12	0.20	
0.5	OLS	0.51	0.16	0.27	0.51	1.00		0.51	0.08	0.26	0.51	1.00	
	GMM-1	0.08	0.90	0.70	0.34	0.04	1.0	0.01	0.37	0.03	0.12	0.01	1.0
	GMM-25	0.47	0.25	0.24	0.48	0.96	25.0	0.32	0.20	0.12	0.33	0.91	25.0
	GMM-Tuk-Han	0.26	0.58	0.16	0.31	0.25	6.5	0.09	0.30	0.02	0.12	0.05	6.6
	GMM-BR	0.23	0.79	0.21	0.34	0.28	3.4	0.06	0.39	0.03	0.13	0.10	3.4
	GMM-Trunc	0.17	0.93	0.26	0.34	0.17	1.0	0.05	0.34	0.03	0.12	0.04	2.0
	CUE-1	0.00	1.16	4.64	0.61	0.03	1.0	-0.01	0.40	0.06	0.13	0.01	1.0
	CUE-25	0.17	4.19	25.46	2.03	0.65	25.0	0.00	0.68	1.88	0.34	0.27	25.0
	WWA	0.07	0.67	0.36	0.26	0.07		0.01	0.32	0.02	0.10	0.05	
0.9	OLS	0.37	0.20	0.15	0.38	1.00		0.37	0.10	0.14	0.37	1.00	
	GMM-1	0.08	1.11	1.62	0.43	0.02	1.0	0.02	0.45	0.04	0.14	0.01	1.0
	GMM-25	0.34	0.33	0.15	0.36	0.63	25.0	0.24	0.26	0.07	0.25	0.43	25.0
	GMM-Tuk-Han	0.21	0.65	0.13	0.28	0.14	9.9	0.08	0.39	0.03	0.14	0.05	9.8
	GMM-BR	0.21	0.91	0.23	0.35	0.24	5.1	0.07	0.50	0.05	0.17	0.12	5.1
	GMM-Trunc	0.21	1.01	0.25	0.37	0.16	3.0	0.04	0.52	0.05	0.16	0.06	2.0
	CUE-1	0.03	1.42	5.52	0.73	0.02	1.0	0.00	0.48	0.08	0.16	0.01	1.0
	CUE-25	0.10	5.68	27.07	2.32	0.58	25.0	0.01	0.73	4.98	0.54	0.28	25.0
	WWA	-0.01	0.76	0.74	0.31	0.05		0.00	0.33	0.02	0.10	0.01	

References

Anatolyev, S. (2005): "GMM, GEL, Serial Correlation and Asymptotics," *Econometrica*, 73, 983–1002.

Andrews, D. W. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59(3), 817–858.

Bierens, H. J. (1987): "Kernel Estimators of Regression Functions," in *Advances in Econometrics: Fifth World Congress*, ed. by T. Bewley, pp. 99–144. Cambridge University Press, New York.

Brockwell, P. J., and R. A. Davis (1991): *Time Series: Theory and Methods*. Springer Verlag-New York, Inc., second edn.

Canay, I. A. (2006): "Simultaneous Selection and Weighting of Moments in GMM using a Trapezoidal Kernel," *Working Paper Department of Economics, University of Wisconsin.*

Carrasco, M. (2011): "A regularization approach to the many instruments problem," *unpublished manuscript.*

Donald, S., G. W. Imbens, and W. K. Newey (2009): "Choosing Instrumental Variables in Conditional Moment Restriction Models," *Journal of Econometrics*, 151, 28–36.

Donald, S. G., and W. K. Newey (2001): "Choosing the Number of Instruments," *Econometrica*, 69(5), 1161–1191.

Folland, G. B. (1984): *Real Analysis - Modern Techniques and Their Applications*. John Wiley and Sons.

Hahn, J., J. Hausman, and G. Kuersteiner (2004): "Estimation with weak instruments: Accuracy of higher-order bias and MSE approximations," *Econometrics Journal*, 7, 272–306.

Hansen, L. P. (1985): "A Method for Calculating Bounds on the Asymptotic Covariance Matrices of Generalized Method of Moments Estimators," *Journal of Econometrics*, 30(1), 203–238.

Hansen, L. P., J. Heaton, and A. Yaron (1996): "Finite Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14, 262–280.

Hansen, L. P., J. C. Heaton, and M. Ogaki (1988): "Efficiency Bounds Implied by Multiperiod Conditional Moment Restrictions," *Journal of the American Statistical Association*, 83, 863–871.

Hansen, L. P., and K. J. Singleton (1991): "Computing Semiparametric Efficiency Bounds for Linear Time Series Models," in *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, ed. by W. A. Barnett, J. Powell, and G. Tauchen, pp. 387–412.

——— (1996): "Efficient Estimation of Linear Asset-Pricing Models With Moving Average Errors," *Journal of Business and Economic Statistics*, 14(1), 53–68.

Hayashi, F., and C. Sims (1983): "Nearly Efficient Estimation of Time Series Models with Predetermined, but not Exogenous Instruments," *Econometrica*, 51, 783–798.

Inoue, A. (2006): "A Bootstrap Approach to Moment Selection," *Econometrics Journal*, 9, 48–75.

Kuersteiner, G., and R. Okui (2010): "Constructing Optimal Instruments by First-Stage Prediction Averaging," *Econometrica*, 78, 697–718.

Kuersteiner, G. M. (2001): "Optimal Instrumental Variables Estimation for ARMA Models," *Journal of Econometrics*, 104(2), 359–405.

——— (2002a): "Efficient IV Estimation for Autoregressive Models with Conditional Heteroskedasticity," *Econometric Theory*, 18(3), 547–583.

——— (2002b): "Mean Squared Error Reduction for GMM Estimators of Linear Time Series Models," *unpublished manuscript*.

——— (2005): "Automatic Inference For Infinite Order Vector Autoregressions," *Econometric Theory*, pp. 85–115.

Magnus, J. R., and H. Neudecker (1979): "The Commutation Matrix: Some Properties and Applications," *Annals of Statistics*, 7(2), 381–394.

Nagar, A. L. (1959): "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations," *Econometrica*, 27(4), 575–595.

Newey, W. K., and R. J. Smith (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72, 219–255.

Newey, W. K., and K. D. West (1994): "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631–654.

Okui, R. (2005): "Instrumental Variable Estimation in the Presence of Many Moment Conditions," *Working Paper, Hong Kong University of Science and Technology.*

Parzen, E. (1957): "On Consistent Estimates of the Spectrum of a Stationary Time Series," *Annals of Mathematical Statistics*, 28(2), 329–348.

West, K. D., and D. Wilcox (1996): "A Comparison of Alternative Instrumental Variables Estimators of a Dynamic Linear Model," *Journal of Business and Economic Statistics*, 14, 281–293.

West, K. D., K.-F. Wong, and S. Anatolyev (2009): "Instrumental Variables Estimation of Heterskedastic Linear Models Using All Lags of Instruments," *Econometric Reviews*, 28, 441–467.