RMSE Reduction for GMM Estimators of Linear Time Series Models

Guido M. Kuersteiner*
Preliminary and Incomplete

January 27, 2000

Abstract

In this paper we analyze GMM estimators for time series models as advocated by Hayashi and Sims, and Hansen and Singleton. It is well known that these estimators achieve efficiency bounds if the number of lagged observations in the instrument set goes to infinity.

A new version of the GMM estimator based on kernel weighted moment conditions is proposed. Higher order asymptotic expansions are used to obtain optimal rates of expansions for the number of instruments to minimize the asymptotic MSE of the estimator.

Estimates of optimal bandwidth parameters are then used to construct a fully feasible GMM estimator where the number of lagged instruments are endogenously determined by the data.

Expressions for the asymptotic bias of kernel weighted GMM estimators are obtained. It is shown that standard GMM procedures have larger asymptotic biases than kernel weighted GMM. A bias correction for the estimator is proposed. It is shown that the bias corrected version achieves a faster rate of convergence of the higher order terms of the MSE than the uncorrected estimator.

Key Words: time series, feasible GMM, number of instruments

*Massachusetts Institute of Technology, Dept. of Economics, 50 Memorial Drive E52-251B, Cambridge, MA 02142. Email: gkuerste@mit.edu, http://web.mit.edu/gkuerste/www/. I wish to thank Xiaohong Chen, Roland Gallant, Jinyong Hahn, Jerry Hausman, Whitney Newey and Ken West for helpful comments.

1. Introduction

In recent years GMM estimators have become one of the main tools in estimating economic models based on first order conditions for optimal behavior of economic agents. Hansen (1982) established the asymptotic properties of a large class of GMM estimators. It was subsequently shown by Chamberlain (1987), Hansen (1985) and Newey (1988) that GMM estimators based on conditional moment restrictions can be constructed to achieve semiparametric efficiency bounds.

In independent sampling situations feasible versions of such estimators were implemented by Newey (1990). In a time series context examples of such estimators are Hayashi and Sims (1983), Stoica, Soderstrum and Friedlander (1985), Hansen and Singleton (1991,1996) and Hansen, Heaton and Ogaki (1996). To this date no analysis of the allowed expansion rate for the number of instruments has been provided. In this paper a data dependent selection rule for the number of instruments is obtained and a fully feasible version of GMM estimators for linear time series models is proposed. The number of lagged instruments is chosen in a way similar to a bandwidth selection procedure for nonparametric density estimation.

While for some time series estimators the number of instruments needed to achieve the efficiency lower bound is small this is not the case in general. Calculations based on asymptotic covariance matrices in Hansen and Singleton (1991) indicate that the number of instruments needed to achieve the lower bounds can be large in some cases. In particular the calculations in Stoica, Soderstrum and Friedlander (1985) for the ARMA(1,1) model indicate that when the moving average coefficient is close to the unit circle the asymptotic efficiency of the parameter estimates approaches the bound slowly with the number of instruments increasing.

This indicates that estimators which allow the number of instruments to grow rapidly with the sample size are empirically important and can lead to overall faster rates of convergence of the higher order terms contributing to the MSE of the estimator. A feasible version of an estimator where the number of instruments grows at the same rate as the sample was recently developed in Kuersteiner (1997) for a special problem. In general however much slower expansion rates for the instrument set are required. This fact was shown by Newey (1990) and Donald and Newey (1997) in a cross section context.

Here a GMM procedure based on kernel weighted moment conditions is proposed. The analysis of the higher order asymptotic terms reveals that bias terms dominate the asymptotic Mean Squared Error (MSE). The idea behind using the kernel weighted version of the GMM estimator is to dampen the importance of these bias terms and thus allowing a larger number of instruments to be included.

While the automatic choice of bandwidth parameters has a relatively long tradition in the nonparametric literature for density estimation, its equivalent in the semiparametric literature is relatively recent. Two approaches have been proposed. Andrews (1999) looks at a moment selection procedure based on minimizing an information criterion over a set of possible moment restrictions.

Alternatively, Linton (1995) analyses the optimal choice of bandwidth parameters based on minimizing the asymptotic MSE of the estimator. He applies this technique to nonparametric kernel estimates of the partially linear regression model. Xiao and Phillips (1996) apply similar ideas to determine the optimal bandwidth in the estimation of the residual spectral density in a Whittle likelihood based regression set up. More recently Linton (1997) extended his procedure to the determination of the optimal bandwidth choice in a efficient semiparametric instrumental variables estimator. While his approach is based on kernel estimates of the optimal instruments, Donald and Newey (1997) use similar arguments to determine the optimal number of base functions in polynomial approximations to the optimal instrument. The idea behind these estimators is to analyze higher order asymptotic expansions of the estimators around their true parameter values. While the first order asymptotic terms typically do not depend on the estimation of infinite dimensional nuisance parameters as shown in Andrews (1994) and Newey (1994) this is not the case for higher order terms of the expansions.

For fully parametric models the higher order terms go to zero with the rate $O_p(n^{-1})$ where n is the sample size. For semiparametric models the rate of convergence typically depends on the way the infinite dimensional nuisance parameters are estimated. Donald and Newey (1997) show that the optimal rate of convergence of the approximate MSE is $O(n^{-\frac{2s}{2s+d}})$ for LIML and JIVE estimators and $O(n^{-\frac{2s}{2s+2d}})$ for 2SLS where s is the degree of differentiability of the nonlinear mean function and d is the dimension of the regressor space. These results conform with the results of Xiao and Phillips (1996) who find an asymptotic rate of convergence of the MSE of $O(n^{-\frac{2s}{2s+1}})$ where s is the degree of differentiability of the innovation spectral density.

In this paper we will obtain expansions similar to the ones of Donald and Newey (1997) for the case of GMM estimators for models with lagged dependent right hand side variables. This set up is important for the analysis of intertemporal optimization models which are characterized by first order conditions of maximization. One particular area of application are asset pricing problems.

Expressions for the asymptotic MSE are obtained. It turns out that the rate of convergence of the higher order terms in the mean squared error are $O(n^{-\frac{2s}{2s+2}})$ which corresponds to the 2SLS case of Donald and Newey (1997). 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. Full implementation of the procedure requires the specification of estimators for the constants in the expression for the optimal bandwidth parameter. It is established that a plug-in estimator for the optimal bandwidth leads to a GMM estimator that is fully feasible and achieves the same asymptotic distribution as the infeasible optimal estimator. Moreover, it is shown that the asymptotic bias is lower if suitable kernel weights are applied to the moment conditions. A semiparametric correction of the asymptotic bias term is proposed. It turns out that a bias corrected version of the GMM estimator achieves a faster optimal rate of convergence of the higher order terms. In this sense the MSE of the bias corrected GMM estimator is an order of magnitude smaller than the MSE of the uncorrected GMM estimator.

The paper is organized as follows. Section 2 presents the time series models and introduces notation. Section 3 contains the analysis of higher order asymptotic MSE terms and derives the optimal number of instruments. Section 4 discusses implementation of the procedure, in particular consistent estimation of the constants in the optimal bandwidth formula. Section 5 analyzes the asymptotic bias of the kernel weighted GMM estimator and introduces the bias corrected GMM estimator.

2. Linear Time Series Models

We consider the linear time series framework of Hansen and Singleton (1996). Let $y_t \in \mathbb{R}^p$ be a strictly stationary stochastic process with $Ey_t^2 < \infty$. By the Wold representation theorem there exists an infinite moving average representation

$$y_t = \mu + C(L)u_t \tag{2.1}$$

where $\mu \in \mathbb{R}^p$ and u_t is a strictly stationary white noise process with $Eu_t = 0$ and $Eu_t u_t' = \sum$. We define the information set of the observer as the σ-filed $\mathcal{F}_t$ generated by current and lagged values of y_t such that $\mathcal{F}_t = \sigma(y_t, y_{t-1}, ...)$. It is assumed that economic theory provides restrictions of the form

$$\Delta(L, \beta)y_t = \varepsilon_t + \alpha_0 \tag{2.2}$$

where $\varepsilon_t = \Phi(L)u_t$ and $\Phi(L) = \Phi_0 + \Phi_1 L + ... + \Phi_{m-1}L^{m-1}$ is a $1 \times p$ vector of lag polynomials of order $m-1$ with $m > 0$ such that ε_t is strictly stationary with $E\varepsilon_t = 0$ and follows an MA(m-1) process. We denote its autocovariance function by $\gamma_j^\varepsilon = E\varepsilon_t\varepsilon_{t-j}$ with $\gamma_j^\varepsilon = 0$ for $|j| \geq m$. The coefficients γ_j^ε can be expressed in terms of Φ_i as $\gamma_j^\varepsilon = \sum_{i=0}^{m-1} \Phi_i \Sigma \Phi_{i-j}'$ where $\Phi_i = 0$ for $i < 0$.

The economic model (2.2) implies moment restrictions of the form

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

These moment restrictions are the basis for the formulation of GMM estimators exploiting orthogonality between ε_{t+m} and elements of the random variables generating $\mathcal{F}_t$. 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 (2.2). A classical example are Asset Pricing models.

The parameter vector of interest is β. To simplify the exposition we assume that the $1 \times p$ vector $\Delta(L,\beta)$ contains finite order lag polynomials of known functional form up to the unknown parameter vector β. Here, it is assumed that $\beta \in \mathbb{R}^d$. In particular assume $\Delta(L,\beta) = \delta_0(\beta) - \delta_1(\beta)L - ... - \delta_r(\beta)L^r$. Identification of the structural parameters β follows from the following Assumption.

Assumption A. *The map $\delta(\beta) = (\delta_0(\beta), ...\delta_r(\beta)) : \Theta \longmapsto \Xi$ is a homeomorphism where $\Xi = \{\xi \in \mathbb{R}^r \times \mathbb{R}^p \,|b - \xi_1 z - ... - \xi_r z^r \neq 0, |z| \leq 1\}$. Without loss of generality it is assumed that $\delta(\beta) : \Xi \longmapsto \Xi$ and that $\delta_i(\beta)$ is the i-th coordinate projection, i.e. $\beta = vec\delta(\beta)$. A normalization restriction $\beta_1 = 1$ is imposed where β_1 is the first element of β.*

The spectral density matrix of y_t is proportional to $\left|C(e^{i\lambda})\right|^2$ where the norm of a complex matrix A is defined as $|A|^2 = trAA^*$ with A^* the complex conjugate transpose of A. The following more formal restrictions are imposed on u_t and $C(L)$.

Assumption B. *Let $u_t \in \mathbb{R}^p$ be a strictly stationary, white noise vector process with $Eu_t = 0$, $Eu_t u_s' = \Sigma\{s = t\}$ where $\{\}$ is the indicator function. Moreover, $E(u_t u_s'|\mathcal{F}_t) = Eu_t u_s'$ for $t \geq s$. Let u_t^i be the i-th element of u_t and $cum_{i_1,...,i_k}(t_1, ..., t_{k-1})$ the k-th order cross cumulant of $u_{t+t_1}^{i_1}, ..., u_t^{i_k}$ defined in (A.1) in the Appendix. Assume that*

$$\sum_{t_1=-\infty}^{\infty} \cdots \sum_{t_{k-1}=-\infty}^{\infty} |cum_{i_1,...,i_k}(t_1, ..., t_{k-1})| < \infty \text{ for } k \leq 8.$$

Assumption (B) implies that the economic model (2.2) is correctly specified only up to second order properties of the data. The assumption of a white noise innovation sequence is also a natural choice because of the Wold representation theorem. The presence of white noise innovations implies that $y_t, y_{t-1}, ...$ are the only available instruments and GMM estimators using this class of instruments achieve the GMM efficiency bound for all GMM estimators based on orthogonality conditions between ε_{t+m} and $\mathcal{F}_t$.

The conditional homoskedasticity condition $E(u_t u_s'|\mathcal{F}_t) = Eu_t u_s'$ is restrictive as it rules out time changing variances. Relaxing this restriction results in more complicated GMM weight matrices as was shown in Kuersteiner (1997, 1998). In principle this higher order moment restriction could be used in addition to the conditions (2.3). The resulting estimator is however nonlinear and will not be considered here.

The summability assumption for the cumulants limits the temporal dependence of the innovation process. Andrews (1991) shows for $k = 4$ that the summability condition on the cumulants is implied by a strong mixing assumption for u_t.

Assumption C (s). *Let u_t satisfy Assumption (B) and let $y_t = \mu_y + \sum_{k=0}^{\infty} C_k u_{t-k}$ where C_k are real matrices of dimension $p \times p$ such that $\sum_k |k|^s \|C_k\| < \infty$ for some $s > 0$.*

The following definitions will be used throughout the paper and are given next. Let y_t satisfy Assumption (C(s)). Partition $y_t = [y_t^1, y_t^{2\prime}]$ where y_t^1 is the first element of y_t. Then define $x_t = [y_t^{2\prime}, y_{t-1}', ..., y_{t-r}']$. Let $\mu_y = Ey_t$ and $\mu_x = Ex_t$. Define $w_{t,i} = (x_{t+m} - \mu_x)\left(y_{t-i} - \mu_y\right)'$, $\Gamma_i^{xy} = Ew_{t,i}$ and $\Gamma_{-i}^{yx} = Ew_{t,i}'$ and let $\bar{w}_{t,i} = w_{t,i} - \Gamma_i^{xy}$. Next define $w_{t,j-i}^y = \left(y_{t-i} - \mu_y\right)\left(y_{t-j} - \mu_y\right)'$ with $Ew_{t,j-i}^y = \Gamma_{j-i}^{yy}$. Let $\varepsilon_t = \Phi(L)u_t$ and define $v_{t,i} = \varepsilon_{t+m}(y_{t-i} - \mu_y)$. Also define $E\varepsilon_{t+m}x_s = \Gamma_{t-s}^{\varepsilon x}$ and $E\varepsilon_{t+m}y_s = \Gamma_{t-s}^{\varepsilon y}$. We define the following second order spectral densities

$$f_{ab}(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \Gamma_j^{ab} e^{-i\lambda j}$$

and

$$f_{ab}^{(q)}(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} |j|^q \Gamma_j^{ab} e^{-i\lambda j} \text{ where } a, b \in \{x, y, \varepsilon\}.$$

The shorter notation f_a is used for f_{aa}. A forth order spectrum of particular interest is $f_\Omega(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sum_{l=-m+1}^{m-1} \gamma_l^{\varepsilon} \Gamma_{j-l}^{yy} e^{-i\lambda j}$ which can be represented as $f_\Omega(\lambda) = 2\pi f_\varepsilon(\lambda) f_y(\lambda)$.

Assumption D. *There exists an $\epsilon > 0$ such that the spectral density $f_\varepsilon(\lambda) > \epsilon$ uniformly in $\lambda \in [-\pi, \pi]$.*

Remark 1. *Assumption (D) is an invertibility condition for the innovation process ε_t. It guarantees that $1/f_\varepsilon(\lambda)$ has the same smoothness properties as $f_\varepsilon(\lambda)$. In particular the Fourier expansion of $f_\varepsilon^{-1}(\lambda)$ has coefficients $\zeta_j = \int f_\varepsilon^{-1}(\lambda) e^{i\lambda j} d\lambda$ such that $\sum_{j=-\infty}^{\infty} |j|^q \left|\zeta_j\right|$ for all $q < \infty$.*

One of the main advantages of using moment conditions (2.3) as a basis for estimating the parameters is that no additional restrictions need to be imposed on the $C(L)$ polynomial. No such restrictions will be assumed for the theoretical part of this paper. Nevertheless it is useful to be able to relate structural and reduced form. In order to completely relate the model (2.2) to the generating process (2.1) we define additional $p-1 \times p$ matrices $\Psi(L)$ and $\Lambda(L)$ such that $\Psi(L)y_t = \Lambda(L)\varepsilon_t + \alpha_1$. The matrices $\Phi(L)$ and $\Lambda(L)$ satisfy

$$\begin{bmatrix} \Delta(L, \beta) \\ \Psi(L) \end{bmatrix}^{-1} \begin{bmatrix} \Phi(L) \\ \Lambda(L) \end{bmatrix} = C(L) \text{ and } \begin{bmatrix} \Delta(1, \beta) \\ \Psi(1) \end{bmatrix}^{-1} \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = \mu.$$

Infeasible efficient GMM estimation for β is based on exploiting all the implications of the moment restriction (2.3). In our context this is equivalent to choosing instrument vectors $z_{t,M} = (y_t', y_{t-1}', ..., y_{t-M+1}')'$ and letting the number of instruments go to infinity. We define an infeasible estimator of β as a reference point to which we compare feasible versions of the estimator. The infeasible estimator of β is based on a nonrandom projection matrix $D_M^{-1}P_M\Omega_M^{-1/2}$ and is given by

$$\hat{\beta}_M = D_M^{-1}P_M\Omega_M^{-1}\frac{1}{n}\sum_t (y_t^1 - \mu_y^1)(z_{t,M} - \imath_M \otimes \mu_y)$$

where $\imath_M$ is an $M \times 1$ vector of ones, $\Omega_M = \sum_{j=-m+1}^{m-1} \gamma_j^{\varepsilon} E z_{t,M} z_{t,M}^{'}$, $P_M = E(x_{t+m} - \mu_x)(z_{t,M} - \imath_M \otimes \mu_y)'$ and $D_M = P_M'\Omega_M^{-1}P_M$.

In order to characterize the limit of D_M and $P_M'\Omega_M^{-1}$ as $M \to \infty$ we introduce the sequence space l^2 of square summable sequences $s = \{s_i\}_{i=1}^{\infty}$ with elements $s_i \in \mathbb{R}^p$ such that $s \in l^2$ if $\sum_i \|s_i\| < \infty$. We define the operator Ω component-wise by its image for all $s \in l^2$ by $b_i = \lim_{m\to\infty} \sum_j^m \omega_{i,j} s_j$ where $\omega_{i,j} = \sum_{l=-m+1}^{m-1} \Gamma_{j-i+l}^{yy} \gamma_l^{\varepsilon}$ is the i,j th block of Ω_M. The operator Ω has a well defined and bounded inverse if it is selfadjoint, bounded and noncompact. These conditions are satisfied for covariance matrices under Assumptions (B) and (C). The Closed Graph theorem then implies boundedness of Ω^{-1}, i.e. $s\Omega^{-1} \in l^2$ for all $s \in l^2$. Denote by $\vartheta_{i,j}$ the i,j-th element of Ω^{-1}. In the same way let $P \in \bigotimes_{j=1}^d l^2$ be an element of the d dimensional product of sequence spaces l^2 in the sense that each column of P is an element of l^2. It then follows that the limiting operator $P\Omega^{-1}$ maps l^2 sequences into l^2 sequences.

Let $D = \lim_M P_M^{'}\Omega_M^{-1}P_M = P'\Omega^{-1}P$ and $d_n = \lim_M P_M'\Omega_M^{-1}\frac{1}{\sqrt{n}}\sum(z_{t,M} - \imath_M \otimes \mu)\varepsilon_t$. It can be shown that $D^{-1}d_n \xrightarrow{p} N(0, D^{-1})$ as $n \to \infty$ under the assumptions made about y_t. It is also true that $\sqrt{n}\hat{\beta}_M - D^{-1}d_n \xrightarrow{p} 0$ as $M \to \infty$. The last statement is no longer true, at least without specifying the rate at which M goes to infinity, once we replace $\hat{\beta}_M$ by a feasible counterpart.

A feasible version of $\hat{\beta}_M$ is obtained by replacing $D_M^{-1}P_M\Omega_M^{-1}$ by an estimated counterpart $\hat{D}_M^{-1}\hat{P}_M'\hat{\Omega}_M^{-1}$. The notation $\hat{\beta}_{n,M}$ is used for such a feasible estimator. We call an estimator fully feasible if M is a function of the data alone. A fully feasible estimator is denoted by $\hat{\beta}_{n,\hat{M}}$.

From the results in Hansen (1985) it follows that estimators for which M goes to infinity are achieving the efficiency lower bound for GMM as long as there are no additional restrictions placed on the lag polynomials $\Psi(L)$ and $\Lambda(L)$.

Once the infeasible estimator has been replaced by a feasible counterpart where $D_M^{-1}P_M'\Omega_M^{-1}$ is estimated from the data the choice of the number of included instruments becomes a more delicate matter. It is well known that introducing additional instruments often comes at the cost of substantial biases for parameter estimates.

A fully feasible procedure therefore requires a data dependent selection rule for the parameter M in a finite sample. We derive such a selection rule in the next section.

3. Higher Order Asymptotic Expansions

In this section we are concerned with the question of determining the optimal bandwidth parameter M^* in $\hat{\beta}_{n,M}$ while the fully feasible version of the estimator where M^* is replaced by an estimate is discussed in the next section.

The criterion used to determine the optimal bandwidth M^* is to minimize the Mean Squared Error (MSE) of terms in a Taylor Series expansion of $\hat{\beta}_{n,M}$ that depend on M and are of highest order in probability. Choosing an optimal value for M^* is based on exploiting the trade off between adding more instruments resulting in higher efficiency and the finite sample biases introduced by additional instruments.

In this paper a generalized class of GMM estimators based on kernel weighted moment restrictions is introduced. Under the assumptions of this paper the conditioning set $\mathcal{F}_t$ is generated by lagged observations $y_t, y_{t-1}, ...$ leading to an infinite set of unconditional moment restrictions of the form $E\varepsilon_{t+m}y_{t-j} = 0$. A conventional GMM estimator is based on using the first M of these moment restrictions. More generally one can consider non-random weights $k(j/M) \in [-1, 1]$ such that

$$k(j/M)E\varepsilon_{t+m}y_{t-j} = 0.$$

This more general approach covers the standard procedure as a special case where $k(j/M) = \{|j/M| \leq 1\}$ is the truncated kernel. One reason for allowing more general kernel functions is discussed in Section 5. It turns out that kernel weighting reduces the asymptotic bias of the GMM estimator. Other advantages are that certain kernel functions such as the Quadratic Spectral kernel allow to use all lagged observations in a sample as instruments. The Quadratic Spectral (QS) kernel is known to have certain optimality properties that apply to our context.

Optimal nuisance parameter selection based on minimizing asymptotic mean squared errors has been used in similar contexts by Xiao and Phillips (1998) and Donald and Newey (1997). The main new technical difficulty handled in this paper is to allow for lagged dependent right hand side variables. The MSE calculations presented here are therefore unconditional rather than conditional.

We first specify the formal requirements the kernel weight function $k(.)$ has to satisfy.

Assumption E. *The kernel function $k(.)$ satisfies $k : \mathbb{R} \mapsto [-1, 1]$, $k(0) = 1$, $k(x) = k(-x) \forall x \in \mathbb{R}$, $\int |k(x)| dx < \infty$, $k(.)$ is continuous at 0 and at all but a finite number of points.*

Assumption F. *The kernel function $k(.)$ satisfies Assumption (E) and for $q \in (0, \infty)$ there exists a constant k_q such that $k_q = \lim_{x \to 0}(1 - k(x))/ |x|^q$. Assume that there exists a largest q such that $k_q \in (0, \infty)$.*

Assumption (E) corresponds to the assumptions made in Andrews (1991). A number of well known kernels such as the Truncated, Quadratic Spectral, Bartlett, Parzen and Tukey Hanning kernels satisfy Assumption (E). Assumption (F) rules out the Truncated kernel. For the Quadratic Spectral, Parzen and Tukey Hanning kernels $q = 2$ and for the Bartlett kernel $q = 1$.

We define the matrix

$$k_M = diag(k(1/M), ..., k(n-1/M))'$$

having kernel weight $k(j/M)$ in the j-th diagonal element and zeros otherwise. Kernels for which $k(x) \neq 0$ for $|x| > 1$ are permitted to allow for potentially all lagged observations to be included as instruments. An instrument selection matrix $S_n(t) = diag(\{t > 1\}, ...\{t > n\})$ is introduced to exclude instruments for which there is no data in the sample. The vector of available instruments is denoted by $\bar{z}_{t,n} = S_n(t)(z_{t,n} - \imath_n \otimes \bar{y})$. The empirical analogue to the moment condition is then

$$g_{n,M}(\beta) = \frac{1}{n}\sum_{t=1}^{n-m}(\Delta(L,\beta)(y_{t+m} - \bar{y}))\bar{z}'_{t,n}K_M$$

with $\bar{y} = n^{-1}\sum_{t=1}^{n} y_t$, $K_M = (k_M \otimes I_p)$ and I_p the p-dimensional identity matrix. The $1 \times n$ vector $\bar{z}'_{t,n}K_M$ is the vector of kernel weighted instruments. Note that for the truncated kernel $k_M = \imath_M$ such that

$$K_M = \begin{bmatrix} I_{Mp} & 0 \\ 0 & 0 \end{bmatrix}.$$

In other words in this case $\bar{z}'_{t,n}(k_M \otimes I_p) = (\bar{z}'_{t,M}, 0)$. Given the definition of the instrument vector $\bar{z}_{t,n}$ one has to estimate an n dimensional covariance matrix Ω^{-1}. As will become transparent from the higher order expansions this is not as problematic as it first seems. The weights in K_M serve as kernel weights for the estimated elements of Ω^{-1}. We define $\hat{\Omega}_n(j) = \frac{1}{n}\sum_{t=\max(1+j,1)}^{\min(n,n+j)} \bar{z}_{t,n}\bar{z}'_{t-j,n}$. The optimal weight matrix is then given by

$$\hat{\Omega}_n = \sum_{j=-m+1}^{m-1} \hat{\gamma}^{\varepsilon}(j)\hat{\Omega}_n(j)$$

where $\hat{\gamma}^{\varepsilon}(j) = \frac{1}{n}\sum_{t=r+k-m+1}^{n-m} \hat{\varepsilon}_t\hat{\varepsilon}_{t-j}$ and $\hat{\varepsilon}_t = \Delta(L, \tilde{\beta}_{n,M})(y_{t+m} - \bar{y})$ for some consistent first stage estimator $\tilde{\beta}_{n,M}$. Note that $\hat{\Omega}_n$ is symmetric but not necessarily positive definite. This is unimportant as long as the estimator $\tilde{\beta}_{n,M}$ is known in closed form which is the case for linear models.

We now define the feasible GMM estimator for a given M such that $M \geq d$. Under Assumption (A) the structural parameters β are identified and $\beta_{n,M}$ has a closed form expression. Let Z_k

be the matrix of stacked instruments $Z_M = [\bar{z}_{1,M}, ..., \bar{z}_{n-m,M}]'$ and $X = [x_{m+1} - \bar{x}, ..., x_{n-m} - \bar{x}]$ the matrix of regressors. Also, Y is the stacked vector of the first demeaned element in y_t. Then define the $d \times pn$ matrix

$$\hat{P}_M = n^{-1} X' Z_n$$

with elements $\hat{\Gamma}_j^{xy} = \frac{1}{n} \sum_{t=j+1}^{n-m} (x_{t+m} - \bar{x})(y_{t-j} - \bar{y})$. The estimator $\beta_{n,M}$ can now be written as

$$\beta_{n,M} = \left[\hat{P}_M K_M \hat{\Omega}_n^{-1} K_M \hat{P}_M' \right]^{-1} \hat{P}_M K_M \hat{\Omega}_n^{-1} K_M \frac{Z_n' Y}{n}$$

where $\hat{P}_M K_M \hat{\Omega}_{n,M}^{-1} K_M \hat{P}_M' = \sum_{i,j} \hat{\Gamma}_i^{xy} k(i/M) \hat{\vartheta}_{i,j} k(j/M) \hat{\Gamma}_j^{yx}$ and $\hat{\vartheta}_{i,j} = \left[\hat{\Omega}_n^{-1} \right]_{i,j}$ is the i,j-th block of $\hat{\Omega}_n^{-1}$. We are considering sequences M_n for which $M_n \leq M_{n+1}$ and $M_n \to \infty$ such that $M_n/\sqrt{n} \to 0$. For notational convenience we usually write $M = M_n$. It then follows that $\left\| \hat{D}_M - D \right\| = o_p(1)$ and $\left\| \hat{d}_M - d_n \right\| = o_p(1)$ where $\hat{D}_M = \hat{P}_M K_M \hat{\Omega}_n^{-1} K_M \hat{P}_M'$ and $\hat{d}_M = \hat{P}_M K_M \hat{\Omega}_n^{-1} K_M \frac{Z_M' \varepsilon}{\sqrt{n}}$.

The bandwidth parameter M is chosen such that the MSE of a weighted sum of the estimators $\beta_{n,M}$ is minimized. We approximate the MSE by first expanding $\beta_{n,M}$ in terms of its elements and then obtaining the MSE for the terms in the expansion that are largest in probability and depend both on M and n. For this purpose a second order Taylor approximation of $\hat{D}_M^{-1}$ around D^{-1} leads to

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

The expansion is valid as long as $M/\sqrt{n} \to 0$. The size of the mean squared error of the estimator is given in the next lemma. Define the mean squared error of $\beta_{n,M}$ as

$$\varphi_n(M, \ell, k(.)) = n\ell' E D^{1/2}(\beta_{n,M} - \beta)(\beta_{n,M} - \beta) D^{1/2\prime}\ell - 1$$

where the normalization $D^{1/2}$ is used to standardize the asymptotic variance. The vector $\ell \in \mathbb{R}^d$ is a vector of weights given to the elements in β. It is usually assumed that $\ell'\ell = 1$ although that is not crucial to the results.

Lemma 3.1. *Suppose Assumptions (A), (B) and (C(s)) hold with $s \geq q$ and $k()$ satisfies Assumptions (E) and (F). Then for any $\ell \in \mathbb{R}^d$ with $\ell'\ell = 1$ the MSE is $\varphi_n(M, \ell, k(.)) = O(M^2/n) + O(M^{-2q})$. The optimal rate of expansion for the set of instruments is $M = O(n^{1/(2+2q)})$. If the truncated kernel $k(x) = \{|x| \leq 1\}$ is used then $\varphi_n(M, \ell, k(.)) = O(M^2/n) + o(M^{-2q})$.*

This result is similar to the result for the 2SLS estimator obtained in Donald and Newey (1997). The source of the $O(M^{-2q})$ bias terms is however different in our context. This is due to the fact that we are weighting the moment restrictions with a weight function $k(x)$ which introduces an additional variance term of order M^{-2q}. The second part of the Lemma shows that using the truncation kernel, i.e. using a standard GMM procedure with a certain number of instruments results in variance terms of lower order than the ones found in Donald and Newey (1997).

The reason why the variance terms are of lower order in the truncated case lies in the stationarity assumption made in the time series context. Since the correlation between instruments and regressors has to decay at a faster than polynomial rate as instruments with longer and longer lags are used, the importance of omitting these far distant instruments is of lower than polynomial order.

The optimal rate of expansion $n^{1/(2+2q)}$ for the bandwidth parameter is slower than the optimal rate encountered in other contexts of automated bandwidth selection, in particular for density estimation. The reason for the slower rate of convergence lies in the presence of asymptotic bias terms of order $O(M/\sqrt{n})$ which dominate the usually present variance terms of order $O(M/n)$.

An immediate corollary resulting from Lemma (3.1) is that the feasible estimator has the same asymptotic distribution as the optimal infeasible estimator as long as $M^2/n \to 0$.

Corollary 3.2. *Suppose Assumptions (A), (B) and (C(s)) hold with $s \geq q$ and $k()$ satisfies Assumptions (E) and (F). If $M \to \infty$ and $M^2/n \to 0$ as $n \to \infty$ then $\sqrt{n}\hat{\beta}_{n,M} - D^{-1}d_n = o_p(1)$.*

The corollary shows that the number of instruments included for estimation can grow at most at rate $o(\sqrt{n})$ in order to achieve the same asymptotic distribution as the infeasible optimal estimator $D^{-1}d_n$. The optimal rate of expansion for M is much slower than $o(\sqrt{n})$. The next proposition gives an expression for the asymptotic MSE using the largest in probability terms depending on M and n. For this purpose we define 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).

Proposition 3.3. *Suppose Assumptions (A), (B) and (C(s)) hold with $s \geq q$ and $k()$ satisfies Assumptions (E) and (F). If $M \to \infty$ and $M^{2q+2}/n \to \kappa$ then for any $\ell \in \mathbb{R}^d$ with $\ell'\ell = 1$*

$$\lim_n n/M^2 \varphi_n(M, \ell, k(.)) = \mathcal{A}\left(\int_{-\infty}^{\infty} k^2(x)dx\right)^2 + k_q^2 \mathcal{B}^{(q)}/\kappa$$

with the constants $\mathcal{A} = \mathcal{A}_1 D^{-1/2}\ell\ell' D^{-1/2}\mathcal{A}_1'$ and $\mathcal{B}^{(q)} = 1/2\ell' D^{-1/2}(\mathcal{B}_1^{(q)} - \mathcal{B}_2^{(q)} D^{-1}\mathcal{B}_2^{(q)\prime})D^{-1/2}\ell$ where $\mathcal{A}_1, \mathcal{B}_1^{(q)}$ and $\mathcal{B}_2^{(q)}$ are defined as

$$\mathcal{A}_1 = \int_{-\pi}^{\pi} \left(vecf^{\Omega}(\lambda)^{-1}\right)' \left[(vecf^{yy}(\lambda)') \otimes f^{\varepsilon x}(\lambda)' + K_{pp}\left(f^{\varepsilon y}(\lambda)' \otimes f^{yx}(\lambda)\right)\right] d\lambda, \quad (3.1)$$

$$\mathcal{B}_1^{(q)} = \int_{-\pi}^{\pi} f_{xy}^{(q)}(\lambda) f_{\Omega}^{-1}(\lambda) f_{yx}(\lambda) d\lambda + \int_{-\pi}^{\pi} f_{xy}(\lambda) f_{\Omega}^{-1}(\lambda) f_{yx}^{(q)}(\lambda) d\lambda. \tag{3.2}$$

$$\mathcal{B}_2^{(q)} = \int_{-\pi}^{\pi} \left[f_{xy}^{(q)}(\lambda) f_{\Omega}^{-1}(\lambda) f_{yx}^{(q)}(\lambda) + f_{\Omega}(\lambda) / \left(f_{\varepsilon}^{(q)}(\lambda) \right)^2 + 2 f_{xy}^{(q)}(\lambda) / f_{\varepsilon}^{(q)}(\lambda) \right] d\lambda, \tag{3.3}$$

The Mean Squared error displays the usual trade off between higher accuracy due to more included terms represented by $M^{-2q} k_q \mathcal{B}^{(q)}$ and distortions introduced by estimating more unknown parameters manifesting itself in $n^{-1} M^2 \mathcal{A} \int k^2(x) dx$. It turns out that the leading contributor to the latter term is the bias from $\left(\hat{\Gamma}_i^{xy} - \Gamma_i^{xy} \right) k(i/M) \vartheta_{i,j} k(j_2/M) v_{t,j}$ which would have zero expectation if $\hat{\Gamma}_i^{xy}$ were nonrandom.

Proposition (3.3) thus gives an analytical explanation of the empirical fact observed when applying GMM procedures in the time series context. Typically, inclusion of a small number of lags leads to significant changes in the parameter estimates. These changes are in fact due to the presence of the Bias term $n^{-1} M^2 \mathcal{A} \int k^2(x) dx$.

The properties of the more standard, non smoothed GMM estimator can be obtained as a corollary to Proposition (3.3). In fact, in this case $k(x) = \{|x| \leq 1\}$ such that $\int k^2(x) dx = 2$ and $k_q = 0$.

Corollary 3.4. *Suppose Assumptions (A), (B) and (C(s)) hold with $s \geq q$ and $k(x) = \{|x| \leq 1\}$. If $M \to \infty$ and $M^{2q+2}/n \to \kappa$ then for any $\ell \in \mathbb{R}^d$ with $\ell' \ell = 1$*

$$\lim_n n/M^2 \varphi_n(M, \ell, k(.)) = 4\mathcal{A}.$$

In other words inclusion of more lags carries no first order benefits of polynomial order and the MSE behaves asymptotically like $n^{-1} M^2$. A more refined analysis of the lower order variance term could be used to determine the optimal choice of M. This problem is however beyond the scope of this paper. For the remaining discussion we therefore exclude the truncated kernel.

We use Proposition (3.3) to determine the optimal number of lagged instruments in the sense of minimizing the approximate (asymptotic) MSE of $\beta_{n,M}$. From well known arguments we deduce that the optimal lag length choice, M^*, is given by

$$M^* = n^{1/(2q+2)} \left(\frac{q k_q^2 \mathcal{B}^{(q)}}{\mathcal{A} \left(\int k(x)^2 dx \right)^2} \right)^{\frac{1}{2+2q}}.$$

Using M^* directly does not result in a feasible procedure because the constants $\mathcal{A}$ and $\mathcal{B}^{(q)}$ are unknown. In the next section estimators for the constants $\mathcal{A}$ and $\mathcal{B}^{(q)}$ are discussed.

4. Fully Feasible GMM

In this section we derive the missing results that are needed to obtain a fully feasible procedure. In particular one needs to replace the unknown optimal bandwidth parameter M^* by an estimate $\hat{M}^*$. Moreover, it needs to be shown that using the estimate $\hat{M}^*$ instead of the optimal value M^* in forming $\hat{\beta}_{n,M}$ does not introduce additional distortions.

In order to have a fully feasible procedure we need a consistent first stage estimator. We define a feasible first stage GMM estimator $\tilde{\beta}_{n,M}$ as the solution to minimizing $\bar{g}(\beta)'\bar{g}(\beta)$ with

$$\bar{g}(\beta) = n^{-1} \sum_{t=M+1}^{n} (\Delta(L,\beta)(y_{t+m} - \bar{y})\tilde{z}_{t,M}. \tag{4.1}$$

The instrument $\tilde{z}_{t,M} = (y_t' - \bar{y}, y_{t-1}' - \bar{y}, ..., y_{t-M+1}' - \bar{y})'$ is a Mp dimensional vector of lagged observations where M indicates the highest lag. As long as M is fixed and finite $\bar{g}(\theta) \xrightarrow{p} \bar{P}_M$. Classical results show that $\tilde{\beta}_{n,M}$ is consistent and asymptotically normal with $\sqrt{n}(\tilde{\beta}_{n,M} - \beta) \xrightarrow{d} N(0, \left(\bar{P}_M'\bar{P}_M\right)^{-1} \bar{P}_M'\Omega_M^{-1}\bar{P}_M \left(\bar{P}_M'\bar{P}_M\right)^{-1})$. Typically one chooses a low number of instruments. The consistent first stage estimate $\tilde{\beta}_{n,M}$ now can be used to obtain consistent estimates of the residuals ε_t which in turn are needed both to construct the optimal weight matrix $\hat{\Omega}_n^{-1}$ and the constants $\mathcal{A}_1, \mathcal{B}_1^{(q)}$ and $\mathcal{B}_2^{(q)}$. Estimation of $\hat{\Omega}_n^{-1}$ was dealt with in the previous chapter and we turn to the estimation of the coefficients $\mathcal{A}_1, \mathcal{B}_1^{(q)}$ and $\mathcal{B}_2^{(q)}$. The following analysis shows that estimation of $\mathcal{A}_1$ and $\mathcal{B}_1^{(q)}$ can be done nuisance parameter free in the sense that consistent estimates of $\mathcal{A}_1$ and $\mathcal{B}_1^{(q)}$ do not depend on additional unknown parameters. Unfortunately the same is not true for $\mathcal{B}_2^{(q)}$ in which case we have to rely on either an approximating parametric model for $C(L)$ or additional bandwidth parameters. In this paper we choose the former approach.

We now first consider the simpler estimation problem for the constants $\mathcal{A}_1$ and $\mathcal{B}_1^{(q)}$. For this purpose define the $(p-1) \times p$ matrix $E_2 = [0, I_{p-1}]$ and the vector valued function $a(\lambda) = \left[e^{-i\lambda}, e^{-i2\lambda}, ..., e^{-i\lambda r}\right]'$ such that $f_{xy}(\lambda) = \left[f_y(\lambda)E_2', (a(\lambda) \otimes f_y(\lambda))'\right]$. We also note that $f_\Omega(\lambda) = 2\pi f_\varepsilon(\lambda) f_y(\lambda)$ such that to

$$f_{xy}(\lambda) f_\Omega^{-1}(\lambda) = (2\pi)^{-1} \begin{bmatrix} I_{p-1} \\ a(\lambda) \otimes I_p \end{bmatrix} f_\varepsilon^{-1}(\lambda).$$

where we define the Fourier expansion of $f_\varepsilon^{-1}(\lambda)$ as $f_\varepsilon^{-1}(\lambda) = (2\pi)^{-1} \sum_j \zeta_j e^{-i\lambda j}$. While $f_\varepsilon^{-1}(\lambda)$ could be estimated nonparametrically from the autocovariances of the $\hat{\varepsilon}_t$ this would not be taking full account of the structure of the model. A better procedure is to exploit the fact that ε_t has a MA(q) representation under the maintained model assumptions.

We assume that consistent estimates of the MA(q) representation of ε_t have been obtained. Using consistent estimates of the parameter β this can be done by using a nonlinear least squares or pseudo maximum likelihood procedure as described in chapter 8 of Brockwell and Davis (1987). Denoting the consistent MA(q) parameters by $\tilde{\theta}_j$ the coefficients ζ_j can be obtained from $\zeta_k = (2\pi)^{-2}\sigma^4 \sum_{j=0}^{\infty} e_1' B^j e_1 e_1' B^{j+k} e_1$ where e_1 is the first unit vector in $\mathbb{R}^m$ and

$$B = \begin{bmatrix} \tilde{\theta}_1 & \tilde{\theta}_2 & \cdots & \tilde{\theta}_q \\ 1 & 0 & & 0 \\ & \ddots & & \vdots \\ 0 & & 1 & 0 \end{bmatrix}.$$

In order to estimate the constant $\mathcal{B}_1^{(q)}$ we need to evaluate $\int f_{xy}^{(q)}(\lambda) f_\Omega^{-1}(\lambda) f_{yx}(\lambda) d\lambda$ which can be written as

$$\int f_{xy}^{(q)}(\lambda) f_\Omega^{-1}(\lambda) f_{yx}(\lambda) d\lambda = \int f_{xy}^{(q)}(\lambda) f_\varepsilon^{-1}(\lambda) \left[\begin{array}{cc} E_2' & (a(\lambda) \otimes I_p)' \end{array} \right] d\lambda.$$

The k,l-th block from this matrix is then given by $\sum_j \zeta_{j+k-l} |j|^q \Gamma_j^{yy}$ if $k,l \neq 0$. The autocovariance matrix Γ_j^{yy} in the previous expression is replaced by $\Gamma_j^{yy} E_2'$ or $E_2 \Gamma_j^{yy}$ if $k = 0$ or $l = 0$ and by $E_2 \Gamma_j^{yy} E_2'$ if both $k = 0$ and $l = 0$. The covariance matrix of y_t can be expressed in terms of the coefficients of the underlying data generating process as $\Gamma_j^{yy} = \Sigma_{l=0}^{\infty} C_l C_{l+j}'$. Note however that C_l need not be estimated here. In fact it is sufficient to substitute $\hat{\Gamma}_j^{yy} = \frac{1}{n}\sum_t y_t y_{t-j}'$ as an estimator into expressions of the form $\sum_j \zeta_{j+k-l} |j|^q \Gamma_j^{yy}$. We denote the estimate of $\mathcal{B}_1^{(q)}$ by $\widehat{\mathcal{B}_1^{(q)}}$. Similarly, we note that $D = \int f_{xy}(\lambda) f_\Omega^{-1}(\lambda) f_{yx}(\lambda) d\lambda$ can be represented as $\int f_\varepsilon^{-1}(\lambda) f_{xy}(\lambda) \left[\begin{array}{cc} E_2' & (a(\lambda) \otimes I_p)' \end{array} \right] d\lambda$ which then can be estimated from quantities of the form $\sum_j \zeta_{j+k-l} \Gamma_j^{yy}$ by the same arguments as before. Such an estimator of D is denoted by $\hat{D}$.

To express the constant $\mathcal{A}$ we use the same definitions as before. From $(vecf_\Omega^{-1})' vecf_y = pf_\varepsilon^{-1}$ it follows that $(vecf^\Omega(\lambda)^{-1})' \left[vecf^{yy}(\lambda)' \otimes f^{\varepsilon x}(\lambda)' \right] = pf_\varepsilon^{-1}(\lambda) f_{\varepsilon x}(\lambda)'$. Furthermore, using the properties of the commutation matrix gives $(vecf_\Omega^{-1})' K_{pp} = (vecf_\Omega^{-1'})'$ leading to

$$(vecf^\Omega(\lambda)^{-1})' \left[K_{pp}(f^{\varepsilon y}(\lambda) \otimes f^{yx}(\lambda)) \right] = f_{\varepsilon x}(\lambda)' f_\varepsilon^{-1}(\lambda).$$

Again we make use of the fact that $f_{\varepsilon x}(\lambda)' = f_{\varepsilon y}(\lambda) \left[\begin{array}{cc} E_2' & (a(\lambda) \otimes I_p)' \end{array} \right]$. The spectral density $f_{\varepsilon y}(\lambda) = (2\pi)^{-1} \sum_k \sum_{i=0 \vee k}^{m-1} \Phi_i C_{i-k} e^{-i\lambda k}$ can be expressed in terms of the coefficients of the underlying DGP. Consistent estimation of $f_{\varepsilon y}(\lambda)$ is difficult because even though the parameters C_i could be inferred from the approximate model for $C(L)$ it is not possible to estimate Φ_i without estimating the errors u_t which in turn requires full specification of the structural model. Nonparametric density estimation on the other hand entails a bandwidth selection problem similar to the one encountered for the estimation of β.

Fortunately, we are not directly interested in $f_{\varepsilon y}(\lambda)$ but rather in $\int f_{\varepsilon x}(\lambda)' f_\varepsilon^{-1}(\lambda) d\lambda$ which is $\left[\sum_k \zeta_k \Gamma_k^{\varepsilon y} E_2', \sum_k \zeta_k \Gamma_{k-1}^{\varepsilon y}, ..., \sum_k \zeta_k \Gamma_{k-r}^{\varepsilon y} \right]$. This quantity can be estimated consistently by replacing ζ_k with $(2\pi)^{-2} \sigma^4 \sum_{k=0}^{\infty} e_1' B^j e_1 e_1' B^{j+k} e_1$ and $\Gamma_k^{\varepsilon y}$ by $\hat{\Gamma}_k^{\varepsilon y} = n^{-1} \sum_{t=k+1}^{n-m} \hat{\varepsilon}_{t+m} y_{t-k}$. Using these estimates one then estimates $\widehat{\mathcal{A}}$ by

$$\widehat{\mathcal{A}} = (p+1)^2 \left[\sum_{k=-n+1}^{n} \hat{\zeta}_k \hat{\Gamma}_k^{\varepsilon y} E_2', ..., \sum_{k=-n+1}^{n} \hat{\zeta}_k \hat{\Gamma}_{k-r}^{\varepsilon y} \right] \hat{D}^{-1/2} \ell \ell' \hat{D}^{-1/2} \left[\sum_{k=-n+1}^{n} \hat{\zeta}_k E_2 \hat{\Gamma}_k^{\varepsilon y}, ..., \sum_{k=-n+1}^{n} \hat{\zeta}_k \hat{\Gamma}_{k-r}^{\varepsilon y} \right]' \tag{4.2}$$

The intuition why quantities of the form $\sum_k \hat{\zeta}_k \hat{\Gamma}_k^{\varepsilon y}$ are consistent comes from the fact that ζ_k satisfies summability restrictions by Assumption (D) and can be estimated uniformly consistently. It thus acts like a kernel smoothing operation on the estimated covariance terms $\hat{\Gamma}_k^{\varepsilon y}$.

Unfortunately, the parameter $\mathcal{B}_2^{(q)}$ is harder to estimate mainly because of the presence of the term $\int f_{xy}^{(q)}(\lambda) f_\Omega^{-1}(\lambda) f_{yx}^{(q)}(\lambda) d\lambda$ which involves unknown Γ_j^{yy} matrices in all three spectral densities and does not lend itself to the same simplifications as before. One possible estimation strategy is nonparametric kernel density estimation of all the spectral densities involved.

An alternative to estimating $\hat{\Gamma}_j^{yy}$ is Andrews' (1991) approach of fitting a, possibly misspecified, parametric model $\tilde{C}(L)$ to $C(L)$ and using the parametric dependence of $\mathcal{B}_2^{(q)}$ on $C(L)$ to obtain a feasible $\hat{M}^*$. Analogue to the results in Andrews the misspecification in $C(L)$ does not affect the asymptotic distribution of $\beta_{n,\hat{M}^*}$ but it results in suboptimal higher order asymptotic properties.

For simplicity we choose a VAR(κ) model as approximating process for $C(L)$ such that

$$y_t = A_1 y_{t-1} + ... + A_\kappa y_{t-\kappa} + \varepsilon_t. \tag{4.3}$$

The choice of κ is guided mainly by practical considerations. A natural candidate is to set $\kappa = r$ where r is the number of lagged variables in the equation of interest. However, if the number of variables p in the system is large then κ should be chosen small, i.e. close to one. Alternatively, consistent model selection criteria could be used to select an optimal κ.

In order to calculate the impulse response coefficients associated with (4.3) define the matrices

$$A = \begin{bmatrix} A_1 & A_2 & \cdots & A_\kappa \\ I & 0 & \cdots & 0 \\ & \ddots & & \vdots \\ & & I & 0 \end{bmatrix} \text{ and } E_1 = \begin{bmatrix} I \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

with dimensions $\kappa p \times \kappa p$ and $\kappa p \times p$. The j-th impulse coefficient of the approximating model is given by $\tilde{C}_j = E_1' A^j E_1$. For any $\epsilon > 0$ there exists a $T_\epsilon < \infty$ such that $\left\| E_1' (\sum_j^T A^j - (I-A)^{-1}) E_1 \right\| < \epsilon$. The autocovariance function Γ_j^{yy} is then approximated by $\tilde{\Gamma}_j^{yy} = \sum_{l=0}^{T_\epsilon} \tilde{C}_l \tilde{C}_{l+j}$. The spectral densities of the approximate model are denoted by $\tilde{f}_{xy}^{(q)}(\lambda)$ and $\tilde{f}_\Omega^{-1}(\lambda)$. We define $\widetilde{\mathcal{B}^{(q)}}$ as

$$\widetilde{\mathcal{B}^{(q)}} = \int_{-\pi}^{\pi} \left[\tilde{f}_{xy}^{(q)}(\lambda) \tilde{f}_\Omega^{-1}(\lambda) \tilde{f}_{yx}^{(q)}(\lambda) + \tilde{f}_\Omega(\lambda) / \left(f_\varepsilon^{(q)}(\lambda) \right)^2 + 2 \tilde{f}_{xy}^{(q)}(\lambda) / f_\varepsilon^{(q)}(\lambda) \right] d\lambda.$$

Substituting estimates $\widehat{\tilde{C}}_j$ for $\tilde{C}_j$ in $\widetilde{\mathcal{B}_2^{(q)}}$ leads to an estimate $\widehat{\widetilde{\mathcal{B}_2^{(q)}}}$. We assume that $\widehat{\widetilde{\mathcal{B}_2^{(q)}}}$ is estimated such that it is $\sqrt{n}$-consistent for $\widetilde{\mathcal{B}^{(q)}}$.

Assumption G. $\sqrt{n}(\widehat{\widetilde{\mathcal{B}_2^{(q)}}} - \widetilde{\mathcal{B}^{(q)}}) = O_p(1)$.

It is then established in the following lemma that the estimates for $\mathcal{B}^{(q)}/\mathcal{A}$ formed by $\widehat{\widetilde{\mathcal{B}^{(q)}}}/\widehat{\mathcal{A}}$ where

$$\widehat{\widetilde{\mathcal{B}^{(q)}}} = \ell' \hat{D}^{-1/2}(\widehat{\widetilde{\mathcal{B}_1^{(q)}}} - \widehat{\mathcal{B}_2^{(q)}} \hat{D}^{-1} \widehat{\mathcal{B}_2^{(q)}}') \hat{D}^{-1/2}\ell \tag{4.4}$$

are well enough behaved to be used in a plug in procedure.

Lemma 4.1. *Let $\widehat{\mathcal{A}}$ be defined in (4.2) and $\widehat{\widetilde{\mathcal{B}^{(q)}}}$ be defined in (4.4) and satisfy (G). Then $\sqrt{n}(\widehat{\mathcal{A}}-\mathcal{A}) = o_p(1)$ and $\sqrt{n}(\widehat{\widetilde{\mathcal{B}^{(q)}}}/\widehat{\mathcal{A}} - \widetilde{\mathcal{B}^{(q)}}/\mathcal{A}) = o_p(1)$ where $\widetilde{\mathcal{B}^{(q)}} = \ell' D^{-1/2}(\widetilde{\mathcal{B}_1^{(q)}} - \mathcal{B}_2^{(q)} D^{-1} \mathcal{B}_2^{(q)\prime}) D^{-1/2}\ell$.*

While $\widehat{\mathcal{A}}$ is positive by construction the same is not true for $\widehat{\widetilde{\mathcal{B}^{(q)}}}$. In practice it is therefore necessary to truncate the resulting bandwidth parameter at lag length one negative bandwidth. Ultimately one is interested in the properties of a fully automated estimator $\beta_{n,\hat{M}^*}$ where the data determined optimal Bandwidth $\hat{M}^*$ is plugged into the kernel function. In order to analyze this estimator we need an additional Lipschitz condition for the class of permitted kernels.

Assumption H. *The kernel $k(.)$ satisfies $|k(x)| \leq C_1 |x|^{-b}$ for some $b > 1 + 1/c_2$ and some $C_1 < \infty$ for $c_2 \in (0, \infty)$ and $|k(x) - k(y)| \leq C |x - y|$ $\forall x, y \in \mathbb{R}$ for some $C < \infty$.*

Assumption (H) corresponds to the assumptions made in Andrews (1991). Using the previous results we are now in a position to state the main theorem of this paper which establishes that an automated bandwidth selection procedure can be used to pick the number of instruments based on sample information alone.

Theorem 4.2. *Suppose Assumptions (A), (B) and (C(s)) hold with $s \geq q$ and $k()$ satisfies Assumptions (E), (F) and (H). If $\hat{M}^* = \left(nqk_q\widehat{\widetilde{\mathcal{B}^{(q)}}}/2\widehat{\mathcal{A}}\left(\int k(x)^2dx\right)^2\right)^{\frac{1}{2+q}}$ then $\sqrt{n}\hat{\beta}_{n,\hat{M}^*} - D^{-1}d_n = o_p(1)$ and*

$$\lim_n n/\tilde{M}^2\varphi_n(\hat{M}^*, \ell, k(.)) = \lim_n n/\tilde{M}^2\varphi_n(\tilde{M}, \ell, k(.))$$

where $\tilde{M} = \left(nqk_q\widetilde{\mathcal{B}^{(q)}}/2\mathcal{A}\left(\int k(x)^2dx\right)^2\right)^{\frac{1}{2+q}}$.

Theorem (4.2) shows that using the feasible bandwidth estimator $\hat{M}^*$ results in estimates $\hat{\beta}_{n,M^*}$ that have asymptotic Mean Squared Errors that are equivalent to asymptotic Mean Squared Errors of estimators where a nonrandom pseudo-optimal bandwidth sequence $\tilde{M}$ is used. More importantly, the theorem establishes that the fully feasible estimator attains the same limiting distribution as the infeasible optimal GMM estimator $D^{-1}d_n$.

5. Bias Reduction and Bias Correction

In this section we analyze the asymptotic bias of $\beta_{n,M}$ as a function of the sample size n and the bandwidth parameter M. An approximation to the bias is obtained by again considering terms that are largest in probability and depend on n and M.

Theorem 5.1. *Suppose Assumptions (A), (B) and (C(s)) hold with $s \geq q$ and $k()$ satisfies Assumptions (E). If $M \to \infty$ and then*

$$\lim_{n\to\infty} n/ME(\beta_{n,M} - \beta) = D^{-1}\mathcal{A}_1' \int k^2(x)dx.$$

A simple consequence of this result is that for many standard kernels the asymptotic bias of the kernel weighted GMM estimator is lower than the bias for the standard GMM estimator based on the truncated kernel.

Corollary 5.2. *Suppose Assumptions (A), (B) and (C(s)) hold with $s \geq q$ and $k()$ satisfies Assumptions (E). If $M \to \infty$ and $\int k^2(x)dx \leq 2$ then*

$$\lim_{n\to\infty} \left\| n/ME(\beta_{n,M} - \beta) \right\| \leq \lim_{n\to\infty} \left\| n/ME(\beta^T_{n,M} - \beta) \right\|$$

where $\beta^T_{n,M}$ is the GMM estimator based on the truncated kernel.

In practice any one of the following well known kernels could be used: the Bartlett $k_B(x) = (1 - |x|)\{|x| \leq 1\}$, the Parzen $k_P(x) = (1 - 6x^2 + 6|x|^3)\{|x| \leq 1/2\} + (2(1 - |x|^3)\{1/2 \leq |x| \leq 1\}$, the Tukey-Hanning $k_T(x) = (1 + \cos(\pi x))/2\{|x| \leq 1\}$ and the Quadratic Spectral $k_S(x) = 25(12\pi^2x^2)^{-1}(5\sin(6\pi x/5)/(6\pi x) - \cos(6\pi x/5))$.

The asymptotic bias for different kernel weighted GMM estimators depends on the constant $\int k(x)^2dx$. These values were published in Andrews (1991) and are 2/3 for the Bartlett, .539285 for the Parzen, 3/4 for the Tukey-Hanning and 1 for the Quadratic Spectral kernel. It thus follows that using any of these standard kernels reduces the asymptotic bias of the estimator.

Another important issue is whether the bias term can be corrected for. The benefits of such a correction are analyzed first. It turns out that correcting for the bias term increases the optimal rate of expansion for the bandwidth parameter and consequently accelerates the speed of convergence to the asymptotic normal limit distribution.

Using the result in Theorem (5.1) the following bias corrected estimator is proposed

$$\beta^*_{n,M} = \beta_{n,M} - \frac{M}{n}\hat{D}^{-1}\widehat{\mathcal{A}}_1' \int k^2(x)dx.$$

The bias term $\widehat{\mathcal{A}}_1$ can be estimated by the methods described in the previous section. The quality of the estimator $\widehat{\mathcal{A}}_1$ determines the impact of the correction on the convergence rate of the corrected estimator. If $\widehat{\mathcal{A}}_1 - \mathcal{A}_1$ is only $o_p(1)$ then the convergence rate of $\beta^*_{n,M}$ is essentially the same as the one for $\beta_{n,M}$. If $\widehat{\mathcal{A}}_1 - \mathcal{A}_1 = O_p(n^{-\delta})$ for $\delta \in (0, 1/2]$ then the convergence rate of the estimator is improved. The mean squared error of the bias corrected estimator is defined as

$$\varphi^*_n(M, \ell, k(.)) = nD^{1/2}\ell' E(\beta^*_{n,M} - \beta)(\beta^*_{n,M} - \beta)'\ell D^{1/2} - 1$$

and we obtain the following result.

Theorem 5.3. *Suppose Assumptions (A), (B) and (C(s)) hold with $s \geq q$ and $k()$ satisfies Assumptions (E) and (F). If $\widehat{\mathcal{A}}_1 - \mathcal{A}_1 = O_p(n^{-1/2})$ then for any $\ell \in \mathbb{R}^d$ with $\ell'\ell = 1$ the MSE is $\varphi^*_n(M, \ell, k(.)) = O(M/n) + O(M^{-2q})$. The optimal rate of expansion for the set of instruments is $M = O(n^{1/(1+2q)})$.*

It follows from Theorem (5.3) that for $M \to \infty$ and $M^{2q+1}/n \to \varkappa$ the rate of convergence of the higher order terms in the estimator is now $n^{-2q/(1+2q)}$ as opposed to the previous rate of $n^{2q/(2q+2)}$. Bias correction in other words improves the MSE by an order of magnitude. The result critically depends on the ability to estimate $\mathcal{A}_1$ with a parametric rate of convergence.

6. Conclusions

We have analyzed the higher order asymptotic properties of GMM estimators for time series models. This extends the literature on optimal bandwidth choice in semiparametric procedures to the case of dependent processes. Using expressions for the asymptotic Mean Squared Error a selection rule for the optimal number of lagged instruments is derived. It is shown that plugging an estimated version of the optimal rule into the estimator leads to a fully feasible GMM procedure.

A new version of the GMM estimator for linear time series models was proposed where the moment conditions are weighted by a kernel function. The asymptotic expansions suggest that the dominating terms of the MSE are bias terms stemming from estimated correlations between instruments and regressors. Kernel weighting of the moment restrictions reduces the importance of these bias terms. It is shown that correcting the estimator for the highest order bias term leads to an overall increase in the optimal rate at which higher order terms vanish asymptotically. In this sense the proposed procedure reduces the asymptotic MSE of the estimator by an order of magnitude.

A. Proofs

We first recall a few well established results on higher order cross cumulants to introduce notation. A reference for this material is Brillinger (1981).

Definition A.1. *Let $u_t \in \mathbb{R}^p$ be a strictly stationary vector process with elements u_t^i such that $Eu_t^i = 0$ and $E\left(u_t^i\right)^k < \infty$. Let $\xi = (\xi_1, ..., \xi_k) \in \mathbb{R}^k$ and $u = (u_{t_1}^{i_1}, ..., u_{t_k}^{i_k})$ then $\phi_{i_1,...i_k,t_1,...,t_k}(\xi) = Ee^{i\xi' u}$ is the joint moment generating function with corresponding cumulant generating function $\ln \phi_{i_1,...i_i,t_1,...,t_k}(\xi)$. The joint k-th order cumulant function is*

$$cum^*_{i_1,...,i_k}(t_1, ..., t_k) = \frac{\partial^{v_1+...+v_k}}{\partial \xi_1^{v_1} \cdots \partial \xi_k^{v_k}}|_{\xi=0} \ln \phi_{i_1,...i_k,t_1,...,t_k}(\xi)$$

where v_i are nonnegative integers $v_1 + ... + v_k = k$. Alternatively the notation $cum^*(u_{t_1}^{i_1}, ..., u_{t_k}^{i_k})$ is used where more convenient. By stationarity it is enough to define $cum_{i_1,...,i_k}(t_1, ..., t_{k-1}) = cum^*_{i_1,...,i_k}(t_1, ..., t_{k-1}, 0)$

Assumption I. *Let $cum_{i_1,...,i_k}(t_1, ..., t_k)$ be defined as in Definition (A.1) such that*

$$\sum_{t_1=-\infty}^{\infty} \cdots \sum_{t_{k-1}=-\infty}^{\infty} |cum_{i_1,...,i_k}(t_1, ..., t_{k-1})| < \infty.$$

Definition A.2. *Let Assumption (I) hold. Then the k-th order cross cumulant spectrum of $u_t^{i_1}, ..., u_t^{i_k}$ is defined as*

$$f_{i_1,...i_k}(\lambda_1, ..., \lambda_{k-1}) = (2\pi)^{-k+1} \sum_{t_1=-\infty}^{\infty} \cdots \sum_{t_{k-1}=-\infty}^{\infty} cum_{i_1,...,i_k}(t_1, ..., t_{k-1}) \exp\left\{-i \sum_{j=i}^{k-1} \lambda_j\right\}$$

for $\infty < \lambda_j < \infty$.

Lemma A.3. *Assume y_t satisfies Assumption (C). Let c_k^i be the i-th row vector of C_k such that $y_t^i = \mu_y^i + \sum_{k=0}^{\infty} c_k^i u_{t-k}$. Define the $1 \times p$ vector polynomial $c^i(L) = \sum_{j=0}^{\infty} c_k^i L^k$ with j-th element $c^{i,j}(L) = \sum_{k=0}^{\infty} c_k^{i,j} L^k$. The cross cumulant spectrum of $(y_t^{i_1}, ..., y_t^{i_k})$ is given by*

$$f_{y^{i_1},...,y^{i_k}}(\lambda_1, ..., \lambda_{k-1}) = (2\pi)^{-k+1} \sum_{j_1=1}^{p} \cdots \sum_{j_k=1}^{p} c^{i_1,j_1}(e^{i\lambda_1}) \cdots c^{i_k,j_k}(e^{-i\sum_{j=i}^{k-1} \lambda_j}) f_{i_1,...i_k}(\lambda_1, ..., \lambda_{k-1}),$$

the cross cumulant is

$$\begin{aligned} cum(y_{t_1}^{i_1}, ..., y_{t_k}^{i_k}) &= cum(y_{t_1-t_k}^{i_1}, ..., y_{t_{k-1}-t_k}^{i_{k-1}}, y_0^{i_k}) \\ &= \sum_{j_1=1,..,j_k=1}^{p} \sum_{l_1=0,...,l_k=0}^{\infty} c_{l_1}^{i_1,j_1} c_{l_2}^{i_2,j_2} \cdots c_{l_k}^{i_k,j_k} cum_{i_1,...,i_k}(l_1 + t_1 - t_k, ..., l_k + t_{k-1} - t_k) \end{aligned}$$

and satisfies $\sum_{t_1=-\infty}^{\infty} \cdots \sum_{t_{k-1}=-\infty}^{\infty} \left| cum(y_{t_1-t_k}^{i_1}, ..., y_{t_{k-1}-t_k}^{i_{k-1}}, y_0^{i_k}) \right| < \infty$.

Proof. The first part follows directly from Brillinger (1981, Theorem 2.8.1). The cumulant $cum(y_{t_1}^{i_1}, ..., y_{t_k}^{i_k})$ is obtained from

$$cum(y_{t_1}^{i_1}, ..., y_{t_k}^{i_k}) = \int \cdots \int f_{y^{i_1},...,y^{i_k}}(\lambda_1, ..., \lambda_{k-1}) e^{-i \sum_j^{k-1} \lambda_j t_j} d\lambda_1 ... d\lambda_{k-1}.$$

For the summability of the cumulant note that

$$\sum_{t_1=-\infty}^{\infty} \cdots \sum_{t_{k-1}=-\infty}^{\infty} |cum_{i_1,...,i_k}(l_1 + t_1 - t_k, ..., l_k + t_{k-1} - t_k)| < \infty$$

uniformly in $l_1, ... l_k$ by Assumption (I). The result then follows from the absolute summability of $c_{l_j}^{i_1,j_1}$ for $j = 1, ..., n$. ■

Using these definitions we prove some results for higher moments involving matrices.

Lemma A.4. *Let W, X, Y, Z be random vectors with elements w_i, x_i, y_i, z_i such that $Ew_i = ... = Ez_i = 0$ and $E|x_i|^4 < \infty, ..., E|z_i|^4 < \infty$. Let A and B be fixed coefficient matrices of dimensions such that the matrix product $W'AXY'BZ$ is a well defined scalar. Then*

$$\begin{aligned} EW'AXY'BZ &= (vecA')' E(X \otimes W) E(Z' \otimes Y') vecB' + tr(EAXZ')(EB'YW') \\ &\quad + tr(EAXY')(EBZW') + \mathcal{K}_4 \end{aligned}$$

where $\mathcal{K}_4 = \sum_{j_1,...,j_4} ... \sum a_{j_1,j_2} b_{j_3,j_4} cum(w_{j_1}, x_{j_2}, y_{j_3}, z_{j_4})$.

Proof. The scalar expression $W'AXY'BZ$ can be written equivalently as $(vecA)'(X \otimes W)(Z' \otimes Y')vecB = trEAXZ'B'YW' = trAXY'BZW'$. The result then follows from $E(w, x, y, z) = E(wx)E(yz) + E(xy)E(wz) + E(xz)E(wy) + cum(w, x, y, z)$. ■

Lemma A.5. *Let X, Y be random vectors, W, Z random matrices with all elements having zero mean and A, B fixed coefficient matrices such that the matrix product $WAXY'BZ$ is well defined. Then*

$$\begin{aligned} EtrWAXY'BZ &= (vecB')' E(Y \otimes Z) E(X' \otimes W') vecA + tr(EAXY')(EBZW') \\ &\quad + tr(B' \otimes I) E(Y' \otimes W)(I \otimes A) Evec(X) vec(Z')' + \mathcal{K}_4 \end{aligned}$$

where $\mathcal{K}_4 = \sum_k \sum_{j_1,...,j_4} ... \sum a_{j_1,j_2} b_{j_3,j_4} cum(w_{j_1,k}, x_{j_2}, y_{j_3}, z_{j_4,k})$.

Proof. Note that $trWAXY'BZ = tr(B' \otimes I)(Y' \otimes W)(I \otimes A) vec(X) vec(Z')$ and use the same reasoning as before. ■

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

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

ii) $E(v_{t,i}\ell' w_{s,j}) = (\ell'\Gamma^{\varepsilon x}_{t-s})\Gamma^{yy}_{t-s+j-i} + \Gamma^{\varepsilon y}_{t-s+j}(\ell'\Gamma^{xy}_{s-t+i}) + \mathcal{K}^2_4$ *where* $\mathcal{K}^2_4$ *is a* $p \times p$ *matrix with typical element* (a, b)

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

iii) $E(v_{t,i}v'_{s,j}) = \gamma^{\varepsilon}_{t-s}\Gamma^{yy}_{t-i+j-s} + \mathcal{K}^3_4$ *where* $\mathcal{K}^3_4$ *is a* $p \times p$ *matrix with typical element* (a, b)

$$\left[\mathcal{K}^3_4\right]_{a,b} = \begin{cases} 0 & i, j \geq 0 \\ cum(\varepsilon_{t+m}, \varepsilon_{s+m}, y^a_{t-i}, y^b_{s-j}) & otherwise \end{cases},$$

iv) $E(w'_{t,i}\ell\ell' w_{s,j}) = \Gamma^{yx}_{-i}\ell\ell'\Gamma^{xy}_{j} + \Gamma^{yx}_{t-i-s}\ell\ell'\Gamma^{xy}_{s-t+j} + \ell'\Gamma^{xx}_{s-t}\ell\Gamma^{yy}_{t-i+j-s} + \mathcal{K}^4_4$ *where* $\mathcal{K}^4_4$ *is a* $p \times p$ *matrix with typical element* (a, b)

$$\left[\mathcal{K}^4_4\right]_{a,b} = cum(\ell' x_s, \ell' x_t, y^a_{t-i}, y^b_{s-j}),$$

v) $Etr(w_{t,i}w'_{s,j}) = tr\Gamma^{xy}_{i}\Gamma^{yx}_{-j} + \gamma^{yy}_{t-i-s+j}\gamma^{xx}_{t-s} + tr\Gamma^{yx}_{t-i-s}\Gamma^{xy}_{t-s+j} + \mathcal{K}^5_4$

$$\mathcal{K}^5_4 = \sum_l \sum_m cum(x^l_s, x^l_t, y^m_{t-i}, y^m_{s-j}),$$

where $\gamma^{yy}_{t-s} = E(y_s - \mu_y)'(y_t - \mu_y)$ *and* $\gamma^{xx}_{t-s} = E(y_s - \mu_y)'(y_t - \mu_y)$.
vi) $E(w_{t,i}w'_{s,j}) = \Gamma^{xy}_{i}\Gamma^{yx}_{-j} + \gamma^{yy}_{t-i+j-s}\Gamma^{xx}_{t-s} + \Gamma^{xy}_{t-i-s}\Gamma^{yx}_{t-s+j} + \mathcal{K}^6_4$ *where* $\mathcal{K}^6_4$ *is a* $p \times p$ *matrix with typical element* (a, b)

$$\left[\mathcal{K}^6_4(t, s, i, j)\right]_{a,b} = \sum_l cum(x^a_s, x^b_t, y^l_{t-i}, y^l_{s-j}),$$

vii) $E(v_{t,i}vec(w'_{s,j})') = \left(\Gamma^{\varepsilon x}_{t+m-s}\right)' \otimes \Gamma^{yy}_{t-i+j-s} + \Gamma^{yx}_{t-i-s} \otimes (\Gamma^{\varepsilon y}_{t-s+j}) + \mathcal{K}^7_4(t, s, i, j)$ *where* $\mathcal{K}^7_4$ *is a* $p \times p$ *matrix with typical element* (a, b)

$$\left[\mathcal{K}^7_4(t, s, i, j)\right]_{a,b} = cum(\varepsilon_t, y^a_{t-i}, y^{b \bmod p+1}_{s-j}, x^{[b/r]+1}_s).$$

Proof. These results are easily shown by applying $E(wxyz) = EwxEyz + EwyExz + EwzExy + cum$ to each element of the respective random matrix or vector and expressing the result in matrix notation. ■

Lemma A.7. *If* $v_{t,i} = \varepsilon_{t+m}(y_{t-i} - \mu_y)$ *and* $w_{t,i} = (x_{t+m} - \mu_x)\left(y_{t-i} - \mu_y\right)'$ *and* A *is a* $p \times p$ *matrix of constants then*

$$\begin{aligned}
& Etr((w^y_{s_1,i,j} - \Gamma^{yy}_{i-j})Av_{t_1,i}v^{'}_{t_2,i}A'(w^y_{s_2,i,j} - \Gamma^{yy}_{i-j})') \\
= \; & (vecA')' \left\{ \left((\Gamma^{yy}_{t_2-i+j-s_2} \otimes \Gamma^{\varepsilon y}_{t_2-s_2}) + K_{pp}(\Gamma^{\varepsilon y}_{t_2-s_2+j} \otimes \Gamma^{yy}_{s_2+i-t_2-j}) + \mathcal{K}^1_4\right)\right. \\
& \times \left.\left((\Gamma^{yy}_{t_1-i+j-s_1} \otimes \Gamma^{\varepsilon y}_{t_1-s_1}) + (\Gamma^{\varepsilon y}_{t_1-s_1+j} \otimes \Gamma^{yy}_{s_1+i-t_1-j}) + \mathcal{K}^1_4\right)\right\} vecA' \\
& + tr(A(\gamma^{\varepsilon}_{t_1-t_2}\Gamma^{yy}_{t_1-i+j-t_2} + \mathcal{K}^3_4))(A(\gamma^{yy}_{s_1-i+j-s_2}\Gamma^{yy}_{s_1-s_2} + \Gamma^{yy}_{s_1-i-s_2}\Gamma^{yy}_{s_1-s_2+j} + \mathcal{K}^6_4)) \\
& + tr\left\{[A' \otimes I]\left[(vec(\Gamma^{yy}_{t_2-i+j-s_1}) \otimes \Gamma^{\varepsilon y'}_{t_2-s_1}) + K_p(\Gamma^{\varepsilon y}_{t_2-s_1+j} \otimes \Gamma^{yy}_{t_2-i-s_1+j}) + \mathcal{K}^1_4\right]\right. \\
& \times [I \otimes A]\left.\left[\left(\Gamma^{\varepsilon x}_{t_1-s_2}\right)' \otimes \Gamma^{yy}_{t_1-i+j-s_2} + \Gamma^{yy}_{t_1-i-s_2} \otimes (\Gamma^{\varepsilon y}_{t_1-s_2+j}) + \mathcal{K}^7_4\right]'\right\} + \mathcal{K}_8
\end{aligned}$$

with $\mathcal{K}_8 = \sum_k \sum_{j_1,\dots,j_4} \dots \sum a_{j_1,j_2}b_{j_3,j_4}cum(y^k_{s_1-i}y^{j_1}_{s_1-j}, \varepsilon_{t_1+m}y^{j_2}_{t_1-i}, \varepsilon_{t_2+m}y^{j_3}_{t_2-i}, y^{j_4}_{s_2-i}y^k_{s_2-j})$. *Note that* $\mathcal{K}_8$ *is a forth order cumulant of the random variables* $y^k_{s_1-i}y^{j_1}_{s_1-j}$ *and* $\varepsilon_{t_1+m}y^{j_2}_{t_1-i}$ *which can be expressed as a sum of products of cumulants of the underlying processes* ε_t *and* y_t.

Proof. Apply Lemmas (A.5) and (A.6). ■

Lemma A.8. *Let* $\tilde{\Gamma}^{yy}_{i-j} = \frac{1}{n}\sum^n_{t=\min(i,j)+1} w^y_{t,i,j}$. *Let* $\vartheta_{i,j}$ *be a* $p \times p$ *matrix of fixed coefficients.. Then* $E\left\|(\tilde{\Gamma}^{yy}_{i-l} - \Gamma^{yy}_{i-l})\vartheta_{i,j}\sum_t \frac{v_{t,j}}{\sqrt{n}}\right\|^2 = O(n^{-1})$. *If* $\sum_i \|\vartheta_{i,j}\| < \infty$ *then*

$$\sum_i E\left\|(\tilde{\Gamma}^{yy}_{i-l} - \Gamma^{yy}_{i-l})\vartheta_{i,j}\sum_t v_{t,j}\right\|^2 < \infty.$$

Proof. We have

$$E\left\|(\tilde{\Gamma}^{yy}_{i-l} - \Gamma^{yy}_{i-l})\vartheta_{i,j}\sum_t \frac{v_{t,j}}{\sqrt{n}}\right\|^2 = n^{-3}\sum_{s_1,s_2,t_1,t_2} Etr((w^y_{s_1,i,j} - \Gamma^{yy}_{i-j})\vartheta_{i,j}v_{t_1,i}v^{'}_{t_2,i}\vartheta'_{i,j}(w^y_{s_2,i,j} - \Gamma^{yy}_{i-j})').$$

We need to consider all the terms appearing in Lemma (A.7). Terms involving second moment are summable over s_1, s_2, t_1, t_2 when standardized by n^{-2} by summability of Γ^{yy}_j and $\Gamma^{\varepsilon y}_j$ together with the Toeplitz Lemma. Forth order cumulant terms are summable over s and t such that $\sum_s \sum_t \|\mathcal{K}^1_4(t, s, i, j)\| < \infty$ which implies that terms involving 4th order cumulants are of lower order by Lemma (A.3). The 8-th order cumulant $\mathcal{K}_8$ is analyzed by considering the following matrix

$$X = \begin{bmatrix} y^k_{s_1-i} & y^{j_1}_{s_1-j} \\ \varepsilon_{t_1+m} & y^{j_2}_{t_1-i} \\ \varepsilon_{t_2+m} & y^{j_3}_{t_2-i} \\ y^{j_4}_{s_2-i} & y^k_{s_2-j} \end{bmatrix}.$$

with typical element $X_{i,j}$. Then from Brillinger (1981), Theorem 2.3.2,

$$cum(\prod_{j=1}^2 X_{1,j}, \prod_{j=1}^2 X_{2,j}, \prod_{j=1}^2 X_{3,j}, \prod_{j=1}^2 X_{4,j}) = \sum_{\upsilon} \prod_{\upsilon_s \in \upsilon} cum(X_{i,j}, i, j \in \upsilon_s)$$

where $cum(X_{i,j}, i, j \in \upsilon_s)$ is the joint cumulant of all the $X_{i,j}$ with indices $i, j \in \upsilon_s$ and the sum is over all indecomposable partitions υ of the table

$$\begin{array}{cc} (1,1) & (1,2) \\ (2,1) & (2,2) \\ (3,1) & (3,2) \\ (4,1) & (4,2) \end{array}.$$

A definition of indecomposable partitions is given in Brillinger (1981), p.20. By indecomposability of partitions all cumulants have at least two elements from different rows. From Lemma (A.3) it then follows that $\sum_{s_1,s_2,t_1,t_2} |\mathcal{K}_8| < \infty$. Finally to show

$$\sum_i E \left\| (\tilde{\Gamma}^{yy}_{i-l} - \Gamma^{yy}_{i-l})\vartheta_{i,j} \sum_t v_{t,j} \right\|^2 < \infty$$

we note that $E \left\| (\tilde{\Gamma}^{yy}_{i-l} - \Gamma^{yy}_{i-l})\vartheta_{i,j} \sum_t v_{t,j} \right\|^2 < C \|\vartheta_{i,j}\|^2$ for some constant $0 \leq C < \infty$ by the first result. ■

Lemma A.9. *Let H_{11} be defined as*

$$H_{11} = -\sum_{i=n+1}^{\infty} \sum_{j=n+1}^{\infty} \Gamma^{xy}_i \vartheta_{i,j} \Gamma^{yx}_{-j} + 2 \sum_{i=1}^{n} \sum_{j=n+1}^{\infty} \Gamma^{xy}_i \vartheta_{i,j} \Gamma^{yx}_{-j}$$

Then $H_{11} = o(n^{-2s})$ where s is such that $\sum |j|^s \left\| \Gamma^{yy}_j \right\| < \infty$.

Proof. Use the inequality $\|H_{11}\| \leq \left(\sum_{j=n+1}^{\infty} \left\| \Gamma^{xy}_{m+i} \right\| \right)^2 + 2 \sum_{j=k+1}^{\infty} \sum_{i=1}^{\infty} \|\Gamma^{xy}_i \vartheta_{i,j}\| \left\| \Gamma^{xy}_{m+j} \right\|$. From Lemma 5.2. in Kuersteiner (1998) it follows that $\sum_j |j|^s \sum_{i=1}^{\infty} \|\Gamma^{xy}_i \vartheta_{i,j}\| < \infty$ such that the second term is bounded by $\sum_{j=n+1}^{\infty} n^{-s} \left\| \Gamma^{xy}_{m+j} \right\| \leq n^{-2s} \sum_{j=k+1}^{\infty} |j|^s \left\| \Gamma^{xy}_{m+j} \right\| \to 0$ as $n \to \infty$ leading to $\|H_{11}\| = o(n^{-2s})$. ■

Lemma A.10. *Let* $H_{12} = \sum_{i=1}^{n}\sum_{j=1}^{n}\Gamma_i^{xy}(1-k(i/M))\vartheta_{i,j}(1-k(j/M))\Gamma_{-j}^{yx}$ *then* $H_{12} = O(M^{-2q})$.

Proof. We write

$$H_{12} = M^{-2q}\sum_{i=1}^{n}\sum_{j=1}^{n}\Gamma_i^{xy}\,|i|^q\,\frac{1-k(i/M)}{|i/M|^q}\vartheta_{i,j}\frac{1-k(j/M)}{|j/M|^q}\,|j|^q\,\Gamma_{-j}^{yx}.$$

By the Dominated Convergence Theorem

$$\begin{aligned}
&\sum_{j_1,j_2=1}^{n}\Gamma_{j_1}^{xy}\,|j_1|^q\,\frac{1-k(j_1/M)}{|j_1/M|^q}\vartheta_{j_1,j_2}\frac{1-k(j_2/M)}{|j_2/M|^q}\,|j_2|^q\,\Gamma_{-j_2}^{yx}\\
\to\quad & k_q^2\sum_{j_1,j_2=0}^{\infty}\Gamma_{j_1}^{xy}\,|j_1|^q\,\vartheta_{j_1,j_2}\,|j_2|^q\,\Gamma_{-j_2}^{yx} = C
\end{aligned}$$

such that $H_{12} = M^{-2q}k_q^2C + o(M^{-2q}) = O(M^{-2q})$. ■

Lemma A.11. *Let* $H_{13} = \sum_{i=1}^{n}\sum_{j=1}^{n}\Gamma_i^{xy}k(i/M)\vartheta_{i,j}(1-k(j/M))\Gamma_{-j}^{yx}$ *and* $H_{14} = H_{13}'$. *Then* $H_{13}+H_{14} = O(M^{-q})$ *and* $H_{13}+H_{14} = M^{-q}k_q2^{-1}\int\left[f_{xy}^{(q)}(\lambda)f_\Omega^{-1}(\lambda)f_{yx}(\lambda) + f_{xy}(\lambda)f_\Omega^{-1}(\lambda)f_{yx}^{(q)}(\lambda)\right]d\lambda + o(M^{-q})$.

Proof.

$$\begin{aligned}
H_{13}+H_{14} &= M^{-q}\sum_{i=1}^{n}\sum_{j=1}^{n}\Gamma_i^{xy}k(i/M)\vartheta_{i,j}\frac{(1-k(j/M))}{|j/M|^q}\,|j|^q\,\Gamma_{-j}^{yx}\\
&\quad +M^{-q}\sum_{i=1}^{n}\sum_{j=1}^{n}\Gamma_i^{xy}\,|i|^q\,\frac{(1-k(i/M))}{|j/M|^q}\vartheta_{i,j}k(i/M)\Gamma_{-j}^{yx}\\
&= M^{-q}k_q\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\Gamma_i^{xy}\vartheta_{i,j}\,|j|^q\,\Gamma_{-j}^{yx} + \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\Gamma_i^{xy}\,|i|^q\,\vartheta_{i,j}\Gamma_{-j}^{yx} + o(M^{-q})\\
&= M^{-q}k_q2^{-1}\int\left[f_{xy}^{(q)}(\lambda)f_\Omega^{-1}(\lambda)f_{yx}(\lambda) + f_{xy}(\lambda)f_\Omega^{-1}(\lambda)f_{yx}^{(q)}(\lambda)\right]d\lambda + o(M^{-q}),
\end{aligned}$$

■

Lemma A.12. *Let* H_{211} *be defined by*

$$H_{211} = \sum_{i=0}^{k}\sum_{j=0}^{k}\left(\hat{\Gamma}_i^{xy}-\tilde{\Gamma}_i^{xy}\right)k(i/M)\vartheta_{i,j}k(j/M)\left(\hat{\Gamma}_j^{xy}-\tilde{\Gamma}_j^{xy}\right)'$$

the $H_{211} = O_p(n^{-1})$.

Proof. First note that

$$\|H_{211}\| \leq \sum_{j_1,j_2=0}^{n} \left\|\hat{\Gamma}_{j_1}^{xy} - \tilde{\Gamma}_{j_1}^{xy}\right\| \|\vartheta_{j_1,j_2}\| \left\|\hat{\Gamma}_{j_2} - \tilde{\Gamma}_{j_2}\right\|.$$

From $\hat{\Gamma}_{m+j} - \tilde{\Gamma}_{m+j} = (\bar{x} - \mu_x)(\bar{y} - \mu_y)'$ one obtains

$$\begin{aligned} E\left\|\hat{\Gamma}_{m+j} - \tilde{\Gamma}_{m+j}\right\|^2 &= n^{-4} \sum_{t_1,t_2,t_3,t_4}^{n} Etr(x_{t_1} - \mu_x)(y_{t_2} - \mu_y)'(y_{t_3} - \mu_y)(x_{t_4} - \mu_x)' \\ &= n^{-4} \sum_{t_1,t_2,t_3,t_4}^{n} tr(\Gamma_{t_1-t_2}^{xy}\Gamma_{t_3-t_4}^{yx} + \gamma_{t_2-t_3}\Gamma_{t_4-t_1}^{xx} + \Gamma_{t_1-t_3}^{xy}\Gamma_{t_2-t_4}^{yx} + \mathcal{K}_4) = O(n^{-2}) \end{aligned}$$

such that H_{211} is bounded in expectation by $n^{-2}c_1 \sum_{j_1,j_2=0}^{n} \|\vartheta_{j_1,j_2}\| + o(n^{-1}) = O(n^{-1})$.

Lemma A.13. *Let H_{212} be defined by*

$$H_{212} = \sum_{i=0}^{k}\sum_{j=0}^{k} \left(\hat{\Gamma}_i - \tilde{\Gamma}_i\right) k(i/M)\vartheta_{i,j}k(j/M)\tilde{\Gamma}_j' + \sum_{i=0}^{k}\sum_{j=0}^{k} \tilde{\Gamma}_i k(i/M)\vartheta_{i,j}k(i/M)\left(\hat{\Gamma}_j - \tilde{\Gamma}_j\right)'.$$

Then $H_{212} = O_p(n^{-1})$.

Proof. First note that

$$\|H_{212}\| \leq \sum_{j_1,j_2=0}^{n} \left\|\hat{\Gamma}_{j_1}^{xy} - \tilde{\Gamma}_{j_1}^{xy}\right\| \|\vartheta_{j_1,j_2}\| \left\|\tilde{\Gamma}_{j_2}^{yx}\right\| + \left\|\tilde{\Gamma}_{j_1}^{xy}\right\| \|\vartheta_{j_1,j_2}\| \left\|\hat{\Gamma}_{j_2}^{yx} - \tilde{\Gamma}_{j_2}^{yx}\right\|.$$

Now

$$\begin{aligned} E\left\|\tilde{\Gamma}_{m+j}\right\|^2 &= n^{-2} \sum_{t=i+1}^{n-m}\sum_{s=i+1}^{n-m} trE(x_{t+m} - \mu_x)(y_{t-j} - \mu_y)'(y_{s-j} - \mu_y)(x_{s+m} - \mu_x)' \\ &= n^{-2} \sum_{t=i+1}^{n-m}\sum_{s=i+1}^{n-m} tr(\Gamma_j^{xy}\Gamma_{-j}^{yx} + \Gamma_{s-t}^{xx}\gamma_{s-t}^{yy} + \Gamma_{t-s}^{xy}\Gamma_{t-s}^{yx} + K_4) \\ &= \left\|\Gamma_j^{xy}\right\|^2 + O(n^{-1}) \end{aligned}$$

where $\gamma_{s-t}^{yy} = E(y_{t-j} - \mu_y)'(y_{s-j} - \mu_y)$ and K_4 is a matrix containing fourth order cumulants of x_{t+m} and y_{t-j}. This together with the arguments in the proof of the previous lemma shows that $E\|H_{212}\| = n^{-1}\sum_{j_1,j_2=0}^{n} \|\vartheta_{j_1,j_2}\| \left\|\Gamma_{j_2}^{yx}\right\| = O(n^{-1})$. ■

Lemma A.14. *Let $H_{221} = \sum_{i=0}^{n}\sum_{j=0}^{n} \left(\tilde{\Gamma}_i^{xy} - \Gamma_i^{xy}\right) k(i/M)\vartheta_{i,j}k(j/M)\left(\tilde{\Gamma}_j^{yx} - \Gamma_j^{yx}\right)'$ with $\tilde{\Gamma}_i^{xy} = n^{-1}\sum_{t=i+1}^{n-m}(x_{t+m} - \mu_x)(y_{t-i} - \mu_y)'$. Then $H_{221} = O_p(M/n)$.*

Proof. For H_{221} we consider

$$E\left\|\tilde{\Gamma}_i^{xy}-\Gamma_i^{xy}\right\|^2 = n^{-2}\sum_{t=i+1}^{n-m}\sum_{s=i+1}^{n-m} tr(\Gamma_{s-t}^{xx}\gamma_{s-t}^{yy}+\Gamma_{t-s}^{xy}\Gamma_{t-s}^{yx}+K_4) = O(n^{-1}) \qquad (A.1)$$

uniformly in i. Then

$$\begin{aligned} & E\left\|\sum_{j_1,j_2=0}^{n}\left(\tilde{\Gamma}_{j_1}^{xy}-\Gamma_{j_1}^{xy}\right)\vartheta_{j_1,j_2}\left(\tilde{\Gamma}_{-j_2}^{yx}-\Gamma_{-j_2}^{yx}\right)'|k(j_1/M)|\,|k(j_2/M)|\right\| \\ \leq & \sum_{j_1,j_2=0}^{n}\left(E\left\|\tilde{\Gamma}_{j_1}^{xy}-\Gamma_{j_1}^{xy}\right\|^2\right)^{1/2}\left(E\left\|\tilde{\Gamma}_{-j_2}^{yx}-\Gamma_{-j_2}^{yx}\right\|^2\right)^{1/2}\|\vartheta_{j_1,j_2}\|\,|k(j_2/M)| \\ \leq & \; Cn^{-1}M\sum_{j_1,j_2=0}^{\infty}\|\vartheta_{j_1,j_2}\|\,|k(j_2/M)|/M = O(M/n). \end{aligned}$$

where C is some constant. In the last equality we have used that $\sum_{,j_2}\|\vartheta_{j_1,j_2}\|\,|k(j_2/M)|/M \in l^1$ where $\{x_i\}_{i=1}^{\infty}\subset l^1$ if $\sum_i\|x_i\|<\infty$. This follows from $M^{-1}\sum_j^n|k(j/M)| \leq M^{-1}\sum_j^n \sup_{j\leq\xi\leq j+1}|k(\xi/M)| \leq \int|k(x)|\,dx$ and $\int|k(x)|\,dx \leq M^{-1}\sum_j^n\inf_{j\leq\xi\leq j+1}|k(\xi/M)| \leq M^{-1}\sum_j^n|k(j/M)|$ by the properties of the Riemann Integral (see Stroock, 1994). Thus $k(j/M)/M\in l^1\subset l^2$. ■

Lemma A.15. *Let $H_{222}=\sum_{i=0}^n\sum_{j=0}^n\left(\tilde{\Gamma}_i^{xy}-\Gamma_i^{xy}\right)k(i/M)\vartheta_{i,j}k(j/M)\tilde{\Gamma}_{-j}^{yx}$. Then $H_{222}=O_p(n^{-1/2}M)$.*

Proof. We use the same arguments as in the proof of the previous lemma.

$$\begin{aligned} E\left\|\sum_{j_1,j_2=0}^{\infty}\tilde{\Gamma}_{j_1}^{xy}\vartheta_{j_1,j_2}\left(\tilde{\Gamma}_{-j_2}^{yx}-\Gamma_{-j_2}^{yx}\right)'k(\frac{j_1}{M})k(\frac{j_2}{M})\right\| &\leq n^{-1/2}\sum_{j_1,j_2=0}^{\infty}\left(E\left\|\tilde{\Gamma}_{j_2}^{xy}\right\|^2\right)^{1/2}\|\vartheta_{i,j}\|\left|k(\frac{j_2}{M})\right| \\ &= O(n^{-1/2}M) \end{aligned}$$

■

Lemma A.16. *Let $H_{34}=\bar{P}_M\left[\Omega^{-1}\right]_n(\hat{\Omega}_n-\Omega_n)\left[\Omega^{-1}\right]_n\bar{P}_M'$ where $\bar{P}_M=(\Gamma_0^{xy},...,\Gamma_n^{xy})(k_M\otimes I_p)$. Then $H_{34}=o_p(n^{-1/2})$.*

Proof. First note that H_{34} can be written as

$$\sum_{i=1}^{n}\sum_{j=1}^{n}a_i^n(\hat{\omega}_{ij}-\omega_{ij})a_j^{\prime n}$$

with $a_i^n = \sum_{j=0}^{n} \Gamma_j k(j/M) \upsilon_{ji}$. Note that $|a_i^n| \leq \sum_{j=1}^{\infty} \|\Gamma_j \upsilon_{ji}\|$ such that $|a_i^n|$ is summable $\forall n$. Furthermore

$$\begin{aligned} \hat{\omega}_{ij} - \omega_{ij} &= \sum_{l=-m+1}^{m-1} \left(\hat{\gamma}_l^{\varepsilon} - \gamma_l^{\varepsilon}\right) \left(\hat{\Gamma}_{j+l-i}^{yy} - \Gamma_{j+l-i}^{yy}\right) \\ &+ \sum_{l=-m+1}^{m-1} \left(\hat{\gamma}_l^{\varepsilon} - \gamma_l^{\varepsilon}\right) \Gamma_{j+l-i}^{yy} \\ &+ \sum_{l=-m+1}^{m-1} \gamma_l^{\varepsilon} \left(\hat{\Gamma}_{j+l-i}^{yy} - \Gamma_{j+l-i}^{yy}\right) \end{aligned}$$

with

$$\begin{aligned} E\left\|\hat{\Gamma}_{j+l-i}^{yy} - \Gamma_{j+l-i}^{yy}\right\| &\leq \left(n^{-2} \sum_{t=i\vee j+l}^{n+(i\wedge j+l)} \sum_{s=i\vee j+l}^{n+(i\wedge j+l)'} tr\left(\gamma_{t-s}^{y} \Gamma_{t-s}^{yy} + \Gamma_{t-s}^{yy} \Gamma_{t-s}^{\prime yy} + K_4^y\right)\right)^{1/2} \\ &+ \left(E\left\|(\mu_y - \bar{y})(\mu_y - \bar{y})'\right\|^2\right)^{1/2} \end{aligned}$$

From $E\left\|(\mu_y - \bar{y})(\mu_y - \bar{y})'\right\|^2 = O(n^{-2})$ and the above inequality it follows that $E\left\|\hat{\Gamma}_{j+l-i}^{y} - \Gamma_{j+l-i}^{y}\right\| = O(n^{-1/2})$ uniformly in i such that

$$E\left\|\sum_{l=-m+1}^{m-1} \left(\hat{\gamma}_l^{\varepsilon} - \gamma_l^{\varepsilon}\right) \left(\hat{\Gamma}_{j+l-i}^{yy} - \Gamma_{j+l-i}^{yy}\right)\right\| = O(n^{-1})$$

and

$$E\left\|\sum_{l=-m+1}^{m-1} \gamma_l^{\varepsilon} \left(\hat{\Gamma}_{j+l-i}^{y} - \Gamma_{j+l-i}^{y}\right)\right\| = O(n^{-1/2}).$$

From these results

$$E\left\|\sum_{i=1}^{k} \sum_{j=1}^{k} a_i (\hat{\omega}_{ij} - \omega_{ij}) a_j'\right\| \leq \sum_{i=1}^{k} \|a_i^k\| \sum_{j=1}^{k} \left(E\left\|(\hat{\omega}_{ij} - \omega_{ij})\right\|^2\right)^{1/2} \left\|a_j^{k'}\right\| = O(n^{-1/2})$$

by the summability of a_i^k.

Lemma A.17. *Let* $H_{31} = (\hat{P}_M - \bar{P}_M) [\Omega^{-1}]_k (\hat{\Omega}_k - \Omega_k) [\Omega^{-1}]_k (\hat{P}_M - \bar{P}_M)'$, $H_{32} = (\hat{P}_M - \bar{P}_M) [\Omega^{-1}]_k (\hat{\Omega}_k - \Omega_k) [\Omega^{-1}]_k \bar{P}_M'$, $H_{33} = H_{32}'$. *Then* $H_{31} = o_p(n^{-1/2})$, $H_{32} = o_p(n^{-1/2})$ *and* $H_{33} = o_p(n^{-1/2})$.

Proof. Use the bound in Equation (A.1) to bound H_{31} and H_{32}. From the results of the previous Lemma it then follows that all these terms are $o_p(n^{-1/2})$. ■

Lemma A.18. *Let* $d_0 = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sum_{j_1,j_2=0}^{\infty} \Gamma_{j_1}^{xy} \vartheta_{j_1,j_2} v_{t,j_2}$. *Then* $\lim_n Ed_0 d_0' = D$.

Proof. Note that $Ed_0 = 0$ and

$$Ed_0 d_0' = \frac{1}{n} \sum_{t=1-m}^{n-m} \sum_{s=t}^{t+2m \vee n} \sum_{j_1,j_2,j_3,j_4=0}^{\infty} \Gamma_{j_1}^{yy} \vartheta_{j_1,j_2} \gamma_{t+m-s}^{\varepsilon} \Gamma_{t-s-j_2+j_3}^{yy} \vartheta_{j_3,j_4} \Gamma_{j_4}^{yy} \to P\Omega^{-1}P' \text{ as } n \to \infty.$$

■

Lemma A.19. *Let* $d_1 = -\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sum_{j_1=n+1}^{\infty} \sum_{j_2=n+1}^{\infty} \Gamma_{j_1}^{xy} \vartheta_{j_1,j_2} v_{t,j_2}$. *Then* $Ed_1 d_1' = o(n^{-4s})$.

Proof. Consider

$$Ed_1 d_1' = \frac{1}{n} \sum_{t=1-m}^{n-m} \sum_{s=t}^{t+2m \vee n} \sum_{j_1,j_2,j_3,j_4=n+1}^{\infty} \Gamma_{j_1}^{yy} \vartheta_{j_1,j_2} \gamma_{t+m-s}^{\varepsilon} \Gamma_{t-s-j_2+j_3}^{yy} \vartheta_{j_3,j_4} \Gamma_{j_4}^{yy} = o(n^{-4s})$$

where s is such that $\sum |j|^s \left\| \Gamma_j^{yy} \right\| < \infty$ and the result follows from the summability assumptions. ■

Lemma A.20. *Let* $d_2 = -\sum_{t=1,j_1=0}^{n} \sum_{j_2=n+1}^{\infty} \Gamma_{j_1}^{xy} \vartheta_{j_1,j_2} \frac{v_{t,j_2}}{\sqrt{n}}$ *and* $d_3 = -\sum_{t=1,j_2=0}^{n} \sum_{j_1=n+1}^{\infty} \Gamma_{j_1}^{xy} \vartheta_{j_1,j_2} \frac{v_{t,j_2}}{\sqrt{n}}$. *Then* $Ed_2 d_2' = o(n^{-2s})$ *and* $Ed_3 d_3' = o(n^{-2s})$.

Proof. For d_2 we have

$$Ed_2 d_2' = \frac{1}{n} \sum_{t=1-m}^{n-m} \sum_{s=t}^{t+2m \vee n} \sum_{j_2,j_3=n+1}^{\infty} \sum_{j_1,j_4=0}^{n} \Gamma_{j_1}^{yy} \vartheta_{j_1,j_2} \gamma_{t+m-s}^{\varepsilon} \Gamma_{t-s-j_2+j_3}^{yy} \vartheta_{j_3,j_4} \Gamma_{j_4}^{yy} = o(n^{-2s})$$

and similarly for d_3 since $\left\| \sum_{i=0}^{k} \Gamma_i \vartheta_{i,j} \right\|$ is summable over j by arguments given in Kuersteiner (1998). ■

Lemma A.21. *Let* $d_{41} = -\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sum_{j_1,j_2=0}^{n} \Gamma_{j_1}^{xy} (1 - k(j_1/M)) \vartheta_{j_1,j_2} (1 - k(j_2/M)) v_{t,j_2}$. *Then* $Ed_{41} d_{41}' = O(M^{-4q})$.

Proof. Consider

$$\begin{aligned} Ed_{41} d_{41}' &= \frac{1}{n} \sum_{t,s=1}^{n} \sum_{j_1,j_2,j_3,j_4=0}^{n} \Gamma_{j_1}^{xy} (1 - k(\frac{j_1}{M})) \vartheta_{j_1,j_2} (1 - k(\frac{j_2}{M})) E(v_{t,j_2} v_{s,j_3}') (1 - k(\frac{j_3}{M})) \vartheta_{j_3,j_4} (1 - k(\frac{j_4}{M})) \Gamma_{j_4}^{yx} \\ &= \frac{M^{-4q}}{n} \sum_{j_1,j_2,j_3,j_4=0}^{n} |j_1|^q \Gamma_{j_1}^{xy} \frac{(1 - k(\frac{j_1}{M}))}{|j_1/M|^q} \vartheta_{j_1,j_2} \frac{(1 - k(\frac{j_2}{M}))}{|j_2/M|^q} |j_2|^q \\ &\quad \times \sum_{t,s=1}^{n} \gamma_{t-s}^{\varepsilon} \Gamma_{t-j_2-s+j_3}^{yy} |j_3|^q \frac{(1 - k(\frac{j_3}{M}))}{|j_3/M|^q} \vartheta_{j_3,j_4} \frac{(1 - k(\frac{j_2}{M}))}{|j_4/M|^q} |j_4|^q \Gamma_{j_4}^{yx}. \end{aligned}$$

Using the fact that $\left|\frac{(1-k(x))}{|x|^q}\right| < C$ for some $C < \infty$ and $\left\|n^{-1}\sum_{t,s=1}^n \gamma_{t-s}^{\varepsilon}\Gamma_{t-j_2-s+j_3}^{yy}\right\|$ is uniformly bounded in j_2 and j_3 leads to

$$\|Ed_{41}d_{41}'\| \leq C_1 M^{-4q}\left(\sum_{j_1,j_2=0}^{n} |j_1|^q\,|j_2|^q\,\left\|\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}\right\|\right)^2$$

where $\sum_{j_1,j_2,j_3,j_4=0}^n |j_1|^q\,|j_2|^q\,\left\|\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}\right\| = O(1)$ by Lemma 5.2. in Kuersteiner (1998). Thus $Ed_{41}d_{41}' = O(M^{-4q})$. ■

Lemma A.22. *Let* $d_{42} = -\frac{1}{\sqrt{n}}\sum_{t=1}^n\sum_{j_1,j_2=0}^n \Gamma_{j_1}^{xy}\left[(1-k(\frac{j_1}{M}))\vartheta_{j_1,j_2}k(\frac{j_2}{M}) + k(\frac{j_1}{M})\vartheta_{j_1,j_2}(1-k(\frac{j_2}{M}))\right]v_{t,j_2}$. *Then* $Ed_{42}d_{42}' = O(M^{-2q})$.

Proof. For d_{42} consider

$$\begin{aligned}&\sum_{j_1,j_2,j_3,j_4=0}^{n}\left\|\Gamma_{j_1}^{xy}(1-k(\frac{j_1}{M}))\vartheta_{j_1,j_2}k(\frac{j_2}{M})\right\|\left\|\frac{1}{n}\sum_{t,s=1}^{n}E(v_{t,j_2}v_{s,j_3}')\right\|\left\|(1-k(\frac{j_3}{M}))\vartheta_{j_3,j_4}k(\frac{j_4}{M})\Gamma_{j_4}^{yx}\right\| \\ \leq\ & C_1 M^{-2q}\left(\sum_{j_1,j_2=0}^{n}|j_1|^q\left\|\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}\right\|\right)\left(\sum_{j_1,j_2=0}^{n}|j_2|^q\left\|\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}\right\|\right)\end{aligned}$$

with the remaining terms being also of the same order such that $Ed_{42}d_{42}' = O(M^{-2q})$. ■

Lemma A.23. *Let* $d_5 = \frac{1}{\sqrt{n}}\sum_{t=1}^n\sum_{j_1,j_2=0}^n\left(\hat{\Gamma}_{j_1}^{xy} - \tilde{\Gamma}_{j_1}^{xy}\right)k(j_1/M)\vartheta_{j_1,j_2}k(j_2/M)v_{t,j_2}$. *Then* $d_5 = O_p(M/n)$.

Proof. For d_5 we consider

$$\begin{aligned}&\left\|\frac{1}{n}\sum_{t,s=1}^{n}\sum_{j_1,j_2,j_3,j_4=0}^{n}k(\frac{j_1}{M})\vartheta_{j_1,j_2}k(\frac{j_2}{M})E(v_{t,j_2}v_{s,j_3}')k(\frac{j_3}{M})\upsilon_{j_3,j_4}k(\frac{j_4}{M})\right\| \\ \leq\ & \sum_{j_1,j_2,j_3,j_4=0}^{n}k(\frac{j_1}{M})\left\|\vartheta_{j_1,j_2}\right\|\left\|\omega_{j_2,j_3}\right\|\left\|\upsilon_{j_3,j_4}\right\|k(\frac{j_4}{M}) = O(M^2)\end{aligned}$$

and using $\hat{\Gamma}_i - \tilde{\Gamma}_i = (\bar{x}-\mu_x)(\bar{y}-\mu_y)' = O_p(n^{-1})$ shows that $d_5 = O_p(M/n)$. ■

Lemma A.24. *Let* $d_6 = \frac{1}{\sqrt{n}}\sum_{t=1}^n\sum_{j_1,j_2=0}^n\left(\tilde{\Gamma}_{j_1}^{xy} - \Gamma_{j_1}^{xy}\right)k(j_1/M)\vartheta_{j_1,j_2}k(j_2/M)v_{t,j_2}$. *Then* $d_6 = O_p(M/\sqrt{n})$.

Proof. We consider

$$E\left\|d_6\right\| \leq \sum_{i=0}^{n}\sum_{j=0}^{n}\left(E\left\|\left(\tilde{\Gamma}_i - \Gamma_i\right)k(\frac{i}{M})\vartheta_{i,j}k(\frac{j}{M})\right\|^2 E\left\|\sum_{t=1}^{n}\frac{v_{t,i}}{\sqrt{n}}\right\|^2\right)^{1/2}$$

where $v_{t,i} = \varepsilon_{t+m}(y_{t-i} - \mu_y)$. Then

$$
\begin{aligned}
E\left\|\sum_{t=1}^{n}\frac{v_{t,i}}{\sqrt{n}}\right\|^2 &= trE(\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}v_{t,i}v_{s,i}^{'}) \\
&= \frac{1}{n}\sum_{t=1-m}^{n-m}\sum_{s=t}^{t+2m\vee n}\gamma_{t-s}^{\varepsilon}tr\left[\Gamma_{t-s}^{yy}\right] \leq 2m\sup_j\left|tr\left[\Gamma_j^{yy}\right]\right|\sup_j\left|\gamma_j^{\varepsilon}\right|
\end{aligned}
$$

and defining $w_{t,i} = (x_{t+m} - \mu_x)(y_{t-i} - \mu_y)'$

$$
\begin{aligned}
E\left\|\left(\tilde{\Gamma}_i - \Gamma_i\right)\vartheta_{i,j}\right\|^2 &= tr\left[E\left(\frac{1}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n}w_{t,i}w_{s,i}' - \Gamma_i\Gamma_i^{'}\right)\vartheta_{ij}\vartheta_{i,j}^{'}\right] \\
&= tr\left[\left(\frac{1}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n}\Gamma_{s-t+i}\Gamma_{t-s+i}^{'} + \gamma_{t-s}\Gamma_{t-s}^{xx'} + cum\right)\vartheta_{ij}\vartheta_{i,j}^{'}\right] \\
&\leq \left\|\frac{1}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n}\Gamma_{s-t+i}\Gamma_{t-s+i}^{'} + \gamma_{t-s}\Gamma_{t-s}^{xx'} + cum\right\|\left\|\vartheta_{ij}\vartheta_{i,j}^{'}\right\|
\end{aligned}
$$

where $\left\|\frac{1}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n}\Gamma_{s-t+i}\Gamma_{t-s+i}^{'} + \gamma_{t-s}\Gamma_{t-s}^{xx'} + cum\right\| = O(n^{-1})$ uniformly in i by the Toeplitz Lemma and $\left\|\vartheta_{ij}\vartheta_{i,j}^{'}\right\| \leq \|\vartheta_{ij}\|^2$. Summability of $\|\vartheta_{ij}\|$ over i shows that $E\|d_6\| = O(M/\sqrt{n})$. ■

Lemma A.25. *Let* $d_7 = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sum_{j_1,j_2=0}^{n}\left(\hat{\Gamma}_{j_1}^{xy} - \Gamma_{j_1}^{xy}\right)k(j_1/M)\vartheta_{j_1,j_2}(\hat{\omega}_{ij}-\omega_{ij})\vartheta_{j_1,j_2}k(j_2/M)v_{t,j_2}$. *Then* $d_7 = O_p(M/n)$.

Proof. For d_7 consider

$$
\begin{aligned}
\|d_7\| \leq& \sum_{i=0}^{n}\sum_{j=0}^{n}\left\|\left(\tilde{\Gamma}_i^{xy} - \Gamma_i^{xy}\right)k(\frac{i}{M})\vartheta_{i,j}(\hat{\omega}_{ij} - \bar{\omega}_{ij})\vartheta_{i,j}k(\frac{j}{M})\sum_{t=1}^{n}\frac{v_{t,j}}{\sqrt{n}}\right\| \\
&+\sum_{i=0}^{k}\sum_{j=0}^{k}\left\|\left(\tilde{\Gamma}_i^{xy} - \Gamma_i^{xy}\right)k(\frac{i}{M})\vartheta_{i,j}(\bar{\omega}_{ij} - \omega_{ij})\vartheta_{i,j}k(\frac{j}{M})\sum_{t=1}^{n}\frac{v_{t,j}}{\sqrt{n}}\right\|
\end{aligned}
$$

where $\hat{\omega}_{ij} = \sum_{l=-m+1}^{m-1}\hat{\gamma}_l^{\varepsilon}\hat{\Gamma}_{i-j-l}^{yy}$ and $\bar{\omega}_{i,j} = \sum_{l=-m+1}^{m-1}\gamma_l^{\varepsilon}\hat{\Gamma}_{i-j-l}^{yy}$ with $\hat{\gamma}_j^{\varepsilon} = \frac{1}{n}\sum_{t=r+k-m+1}^{n-m}\hat{\varepsilon}_t\hat{\varepsilon}_{t-j}$ and $\hat{\varepsilon}_t = \Delta(L,\tilde{\theta})(y_{t+m} - \bar{y})$, $\hat{\Gamma}_{i-j-l}^{yy} = \frac{1}{n}\sum_{t=\max(1+l+k,1)}^{\min(n,n+l)}(y_{t-i} - \bar{y})(y_{t-l-j} - \bar{y})'$. First

$$
\begin{aligned}
&\left\|\left(\tilde{\Gamma}_i^{xy} - \Gamma_i^{xy}\right)k(\frac{i}{M})\vartheta_{i,j}(\hat{\omega}_{ij} - \bar{\omega}_{ij})\vartheta_{i,j}k(\frac{j}{M})\sum_{t=1}^{n}\frac{v_{t,j}}{\sqrt{n}}\right\| \\
\leq& \sum_{l=-m+1}^{m-1}\left|\hat{\gamma}_l^{\varepsilon} - \gamma_l^{\varepsilon}\right|\left\|\left(\tilde{\Gamma}_i^{xy} - \Gamma_i^{xy}\right)k(\frac{i}{M})\vartheta_{i,j}\hat{\Gamma}_{i-j-l}^{yy}\vartheta_{i,j}k(\frac{j}{M})\sum_{t=1}^{n}\frac{v_{t,j}}{\sqrt{n}}\right\|
\end{aligned}
$$

where $|\hat{\gamma}_l^\varepsilon - \gamma_l^\varepsilon| = O_p(n^{-1/2})$ and

$$E\sum_{i=0}^{k}\sum_{j=0}^{k}\left\|\left(\tilde{\Gamma}_i^{xy} - \Gamma_i^{xy}\right)k(\frac{i}{M})\vartheta_{i,j}\hat{\Gamma}_{i-j-l}^{yy}\vartheta_{i,j}k(\frac{j}{M})\sum_{t=1}^{n}\frac{v_{t,j}}{\sqrt{n}}\right\|$$

$$\leq \sum_{i=0}^{n}\sum_{j=0}^{n}\left(E\left\|\left(\tilde{\Gamma}_i^{xy} - \Gamma_i^{xy}\right)k(\frac{i}{M})\vartheta_{i,j}\right\|^2 E\left\|\hat{\Gamma}_{i-j-l}^{yy}\vartheta_{i,j}k(\frac{j}{M})\sum_{t=1}^{n}\frac{v_{t,i}}{\sqrt{n}}\right\|^2\right)^{1/2} \quad \text{(A.2)}$$

As before $E\left\|\left(\tilde{\Gamma}_i^{xy} - \Gamma_i^{xy}\right)k(\frac{i}{M})\vartheta_{i,j}\right\|^2 \leq E\left\|\left(\tilde{\Gamma}_i^{xy} - \Gamma_i^{xy}\right)\right\|^2\|\vartheta_{i,j}\|^2$ and $E\left\|\left(\tilde{\Gamma}_i^{xy} - \Gamma_i^{xy}\right)\right\|^2 = O(n^{-1})$ uniformly in i. Also $E\left\|\hat{\Gamma}_{i-j-l}^{yy}\vartheta_{i,j}\sum_{t=1}^{n}\frac{v_{t,i}}{\sqrt{n}}\right\|^2 = O(1)$ such that (A.2) can be bounded by $CM/n\sum_{i=0}^{n}\sum_{j=0}^{n}\|\vartheta_{i,j}\|\left|k(\frac{j}{M})/M\right|$ for some constant C. It follows that the first term in d_7 is $O_p(M/n)$. For the second term we have

$$E\sum_{i=0}^{n}\sum_{j=0}^{n}\left\|\left(\tilde{\Gamma}_i^{xy} - \Gamma_i^{xy}\right)k(\frac{i}{M})\vartheta_{i,j}(\bar{\omega}_{ij} - \omega_{ij})\vartheta_{i,j}k(\frac{j}{M})\sum_{t=1}^{n}\frac{v_{t,i}}{\sqrt{n}}\right\|$$

$$\leq \sum_{l=-m+1}^{m-1}\gamma_l^\varepsilon\sum_{i=0}^{n}\sum_{j=0}^{n}\left(E\left\|\left(\tilde{\Gamma}_i^{xy} - \Gamma_i^{xy}\right)k(\frac{i}{M})\vartheta_{i,j}\right\|^2 E\left\|(\hat{\Gamma}_{i-j-l}^{yy} - \Gamma_{i-j-l}^{yy})\vartheta_{i,j}k(\frac{j}{n})\sum_{t=1}^{n}\frac{v_{t,i}}{\sqrt{n}}\right\|^2\right)^{1/2}$$

by Lemma (A.8) $E\left\|(\hat{\Gamma}_{i-j-l}^{yy} - \Gamma_{i-j-l}^{yy})\vartheta_{i,j}\sum_{t=1}^{n}\frac{v_{t,i}}{\sqrt{n}}\right\|^2 = O(n^{-1})$ uniformly in i and j and summable over i thus the second term of d_7 is $O_p(M/n)$. ■

Lemma A.26. *Let* $d_8 = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sum_{j_1,j_2=0}^{n}\Gamma_{j_1}^{xy}k(j_1/M)\vartheta_{j_1,j_2}(\hat{\omega}_{ij} - \omega_{ij})\vartheta_{j_1,j_2}k(j_2/M)v_{t,j_2}$. *Then* $d_8 = O_p(n^{-1})$.
Proof. *We bound* d_8 *by*

$$\|d_8\| \leq \frac{1}{\sqrt{n}}\sum_{j_1,j_2=0}^{n}\left\|\Gamma_{j_1}^{xy}k(j_1/M)\vartheta_{j_1,j_2}\right\|\left\|(\hat{\omega}_{ij} - \omega_{ij})\vartheta_{j_1,j_2}k(j_2/M)\sum_{t=1}^{n}v_{t,j_2}\right\|$$

such that $E\|d_8\| = O(n^{-1}\sum_{j_1,j_2=0}^{n}\left\|\Gamma_{j_1}^{xy}k(j_1/M)\vartheta_{j_1,j_2}\right\|) = O(n^{-1})$. ■

Lemma A.27. *Let* $d_9 = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varepsilon_{t+m}\sum_{j_1,j_2=0}^{n}\hat{\Gamma}_{j_1}^{xy}k(j_1/M)\hat{\upsilon}_{i,j}k(j_2/M)(\bar{y} - \mu_y)$. *Then* $d_9 = O_p(M/n)$.

Proof. For d_9 note that $(\bar{y} - \mu_y) = O_p(n^{-1/2})$, $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varepsilon_{t+m} = O_p(1)$ and

$$\left\|\sum_{j=1}^{n}\sum_{i=0}^{n}\hat{\Gamma}_i^{xy}k(\frac{i}{M})\hat{\upsilon}_{i,j}k(\frac{j}{M})\right\| \leq \sum_{j=1}^{n}\sum_{i=0}^{n}\left\|(\hat{\Gamma}_i^{xy} - \Gamma_i^{xy})k(\frac{i}{M})\vartheta_{i,j}k(\frac{j}{M})\right\|$$

$$+\sum_{j=1}^{n}\sum_{i=0}^{n}\left\|\hat{\Gamma}_i^{xy}\right\|\left\|k(\frac{i}{M})\vartheta_{i,j}(\bar{\omega}_{ij}-\omega_{ij})\vartheta_{i,j}k(\frac{j}{M})\right\|$$
$$+\sum_{j=1}^{n}\sum_{i=0}^{n}\left\|\Gamma_i^{xy}k(\frac{i}{M})\vartheta_{i,j}k(\frac{j}{M})\right\|$$

where $E\left\|(\hat{\Gamma}_i^{xy}-\Gamma_i^{xy})k(\frac{i}{M})\vartheta_{i,j}k(\frac{j}{M})\right\|^2 = O(n^{-1})$ and $\sum_{j=1}^{k}\sum_{i=0}^{k}\left\|\Gamma_i^{xy}k(\frac{i}{M})\vartheta_{i,j}k(\frac{j}{M})\right\| = O(1)$. Using

$$\begin{aligned}\|\vartheta_{i,j}(\hat{\omega}_{ij}-\omega_{ij})\vartheta_{i,j}\| &\leq \|\vartheta_{i,j}(\hat{\omega}_{ij}-\bar{\omega}_{ij})\vartheta_{i,j}\|+\|\vartheta_{i,j}(\bar{\omega}_{ij}-\omega_{ij})\vartheta_{i,j}\| \\ &\leq \sum_{l=-m+1}^{m-1}|\hat{\gamma}_l^{\varepsilon}-\gamma_l^{\varepsilon}|\left\|\vartheta_{i,j}\hat{\Gamma}_{i,j}^{yy}(l)\vartheta_{i,j}\right\| \\ &\quad +|\gamma_l^{\varepsilon}|\left\|\vartheta_{i,j}\left(\hat{\Gamma}_{i,j}^{yy}(l)-\Gamma_{i,j}^{yy}(l)\right)\vartheta_{i,j}\right\|\end{aligned}$$

such that $E\|\vartheta_{i,j}(\hat{\omega}_{ij}-\omega_{ij})\vartheta_{i,j}\|^2 = O(n^{-1})$ by previous results and thus $d_9 = O_p(M/n)$. ■

Lemma A.28. *Let d_0 be as defined in Lemma (A.18) and d_2 and d_3 as defined in Lemma (A.20). Then $Ed_0d_2 = o(n^{-2s})$.*

Proof.

$$\begin{aligned}\left\|Ed_0d_2^{'}\right\| &\leq \frac{1}{n}\sum_{t,s}\sum_{j_1,j_2}\sum_{j_3=1}^{n}\sum_{j_4=n+1}^{\infty}\left\|\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}Ev_{t,j_2}v_{s,j_3}'\vartheta_{j_3,j_4}\Gamma_{-j_4}^{yx}\right\| \\ &\leq \sum_{j_1,j_2}\sum_{j_3=1}^{n}\left\|\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}\right\|\|\omega_{j_2,j_3}\|\,n^{-2s}\sum_{j_4=n+1}^{\infty}j_4^s\|\vartheta_{j_3,j_4}\|\sum_{h_1=n+1}^{\infty}h_1^s\|\Gamma_{h_1}\| \\ &\leq n^{-2s}\sup_{j_2}\left(\sum_{j_3=1}^{n}\|\omega_{j_2,j_3}\|\sum_{j_4=n+1}^{\infty}j_4^s\|\vartheta_{j_3,j_4}\|\sum_{h_1=n}^{\infty}h_1^s\|\Gamma_{h_1}\|\right)\sum_{j_1,j_2}\|\Gamma_{j_1}\vartheta_{j_1,j_2}\| \\ &= o(n^{-2s})\end{aligned}$$

where $Ed_0d_3' = o(n^{-2s})$ by the same arguments. ■

Lemma A.29. *Let d_0 be as defined in Lemma (A.18) and d_{42} as defined in Lemma (A.22). Then $Ed_0d_{42} = -M^{-q}k_q\mathcal{B}_1^{(q)}/2 + o(M^{-q})$.Next*

$$\begin{aligned}Ed_0d_{42}^{'} &= -\frac{1}{n}\sum_{t,s=1}^{n}\sum_{j_1,j_2=0}^{\infty}\sum_{j_3,j_4=0}^{n}\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}E(v_{t,j_2}v_{s,j_3}') \\ &\quad\times\left[(1-k(\frac{j_3}{M}))\vartheta_{j_3,j_4}k(\frac{j_4}{M})+k(\frac{j_3}{M}))\vartheta_{j_3,j_4}(1-k(\frac{j_4}{M}))\right]\Gamma_{j_4}^{yx}\end{aligned}$$

$$= M^{-q}k_q \sum_{j_1,j_2,j_3,j_4=0}^{\infty} \Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}\omega_{j_2,j_3}\left[|j_3|^q \vartheta_{j_3,j_4} + \vartheta_{j_3,j_4}|j_4|^q\right]\Gamma_{-j_4}^{yx} + o(M^{-q})$$

$$= M^{-q}k_q 2^{-1}\left[\int f_{xy}^{(q)}(\lambda)f_{\Omega}^{-1}(\lambda)f_{yx}(\lambda)d\lambda + \int f_{xy}(\lambda)f_{\Omega}^{-1}(\lambda)f_{yx}^{(q)}(\lambda)d\lambda\right] + o(M^{-q})$$

The second equality uses Lemma (A.21) to replace $k(\frac{j_4}{M})$ by 1. For the third equality note that $\sum_{j_2=0}^{\infty}\vartheta_{j_1,j_2}\omega_{j_2,j_3} = 1$ if $j_1 = j_3$ and zero otherwise.

Lemma A.30. *Let d_0 be as defined in Lemma (A.18), d_6 as defined in Lemma (A.24) and d_8 as defined in Lemma (A.26). Let $\ell \in \mathbb{R}^d$ such that $\ell'\ell = 1$ .Then $E\ell' d_0 d_6' \ell = O(n^{-2})$ and $E\ell' d_0 d_8' \ell = O(n^{-2})$.*

Proof.

$$E\ell' d_0 d_6' \ell = \frac{1}{n}\sum_{t,s}\sum_{j_1,j_2=0}^{\infty}\sum_{j_3,j_4=0}^{n} k(\frac{j_3}{M})k(\frac{j_4}{M})\ell'\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}Ev_{t,j_2}v_{s,j_3}'\vartheta_{j_3,j_4}(\hat{\Gamma}_{-j_4}^{yx} - \Gamma_{-j_4}^{yx})'\ell$$

where

$$\begin{aligned}
\ell'\Gamma_i\vartheta_{i,j}v_{t,j}v_{s,l}'\vartheta_{l,h}(\hat{\Gamma}_h - \Gamma_h)'\ell &= vec\vartheta_{i,j}(v_{t,j}\otimes\Gamma_i'\ell)(\ell'(\hat{\Gamma}_h - \Gamma_h)\otimes v_{s,l}')vec\vartheta_{l,h} \\
&= tr(\vartheta_{i,j}v_{t,j}(\hat{\Gamma}_h - \Gamma_h)'\ell)(\vartheta_{l,h}'v_{s,l}'\Gamma_i'\ell) \\
&= tr(\vartheta_{i,j}v_{t,j}v_{s,l}')(\vartheta_{l,h}(\hat{\Gamma}_h - \Gamma_h)'\ell\ell'\Gamma_i)
\end{aligned}$$

such that all products of expectations with less than three terms are zero. Denote the elements of $\left[\ell'\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}\right]_a = g_{j_1 j_2}^a$ and $[\vartheta_{j_3,j_4}]_{a,b} = \vartheta_{j_3,j_4}^{a,b}$. Then

$$E\ell' d_0 d_6' \ell = n^{-2}\sum_{t,s,r}\sum_{j_1,j_2}^{\infty}\sum_{j_3,j_4}^{n}\sum_{a,b,c} k(\frac{j_3}{M})k(\frac{j_4}{M})g_{j_1 j_2}^a\vartheta_{j_3,j_4}^{b,c}cum(\varepsilon_{t+m},\varepsilon_{s+m},y_{t-j}^a,y_{s-l}^b,x_r'\ell,y_{r-h}^c)$$

which is of order $O(n^{-2})$ because the cumulant is summable over t,s,r and j_4. The term $Ed_0d_8' = O(n^{-2})$ by the same arguments. ■

Lemma A.31. *Let d_2 and d_3 be as defined in Lemma (A.20). Then $Ed_2d_3' = o(n^{-4s})$.*

Proof. Consider

$$\begin{aligned}
\|Ed_2d_3'\| &\leq \frac{1}{n}\sum_{t,s}\sum_{j_1,j_3}^{n}\sum_{j_2,j_4=n+1}^{\infty}\left\|\Gamma_{j_1}^{xy}\vartheta_{j_2,j_3}Ev_{t,j_2}v_{s,j_3}'\vartheta_{j_3,j_4}\Gamma_{-j_4}^{yx}\right\| \\
&\leq \sum_{j_1,j_3}^{n}\sum_{j_2,j_4=n+1}^{\infty}\left\|\Gamma_{j_1}^{xy}\vartheta_{j_2,j_3}\omega_{j_2,j_3}\vartheta_{j_3,j_4}\Gamma_{j_4}^{yx}\right\| = o(n^{-4s}).
\end{aligned}$$

■

Lemma A.32. *Let* d_6 *as defined in Lemma (A.24). Then* $E\ell' d_6 d_6' \ell = M^2/n\mathcal{A}_1\ell\ell'\mathcal{A}_1' + o(M^2/n)$.

Proof. By Lemma (A.4) we have

$$\begin{aligned} E\ell d_6 d_6' \ell &= \frac{1}{n^3}\sum_{t_1,t_2,s_1,s_2}\sum_{j_1,j_2,j_3,j_4}^{n}\prod_{l=1}^{4}k(\tfrac{j_l}{M})\left\{\left(vec\vartheta_{j_1,j_2}'\right)' E(v_{t_1,j_2}\otimes\bar{w}_{s_1,j_1}')(\ell'\otimes\ell)E(\bar{w}_{s_2,j_4}\otimes v_{t_2,j_3}')vec\vartheta_{j_3,j_4}\right. \\ &+tr\left[\vartheta_{j_1,j_2}E(v_{t_1,j_2}\ell'\bar{w}_{s_2,j_4})\vartheta_{j_3,j_4}E(v_{t_2,j_3}\ell'\bar{w}_{s_1,j_1})\right] \\ &\left.+tr\left[\vartheta_{j_1,j_2}E(v_{t_1,j_2}v_{t_2,j_3})\vartheta_{j_3,j_4}E(\bar{w}_{s_1,j_1}'\ell\ell'\bar{w}_{s_2,j_4})\right]\right\} + \mathcal{K}_8 \end{aligned}$$

where the eight order cumulant term $\mathcal{K}_8$ is summable by Lemma (A.8). Using the results in Lemma (A.6) the first term can be written as

$$\frac{1}{n}\sum_{t_1,s_1}\sum_{j_1,j_2}^{n}\prod_{l=1}^{4}k(\tfrac{j_l}{M})\left(vec\vartheta_{j_1,j_2}'\right)' E(v_{t_1,j_2}\otimes\bar{w}_{s_1,j_1}')\ell\ell'\sum_{t_2,s_2}\sum_{j_3,j_4}^{k}E(\bar{w}_{s_2,j_4}\otimes v_{t_2,j_3}')vec\vartheta_{j_3,j_4}$$

where

$$E(v_{t_1,j_2}\otimes\bar{w}_{s_1,j_1}') = \left(vec\Gamma^{yy}_{t_1-s_1+j_1-j_2}\otimes\Gamma^{\varepsilon x\prime}_{t_1-s_1+m}\right) + K_{pp}(\Gamma^{\varepsilon y}_{t_1-s_1+m+j_1}\otimes\Gamma^{yx}_{t_1-s_1-j_2}) + \mathcal{K}_4^1$$

such that

$$\frac{1}{n^3}\sum_{t_1,s_1}E(v_{t_1,j_2}\otimes\bar{w}_{s_1,j_1}') = \sum_{h=-n+1}^{n-1}(1-\frac{|h|}{n})\left[\left(vec\Gamma^{yy}_{h+j_1-j_2}\otimes\Gamma^{\varepsilon x\prime}_{h+m}\right) + K_{pp}(\Gamma^{\varepsilon y}_{h+m+j_1}\otimes\Gamma^{yx}_{h-j_2})\right] + O(n^{-1}).$$

It now follows from Parzen (1957) that

$$\begin{aligned} &\frac{1}{nM}\sum_{t_1,s_1}\sum_{j_1,j_2}^{n}\prod_{l=1}^{2}k(\tfrac{j_l}{M})\left(vec\vartheta_{j_1,j_2}'\right)' E(v_{t_1,j_2}\otimes\bar{w}_{s_1,j_1}') \\ \to &\int k(x)^2dx\int\left(vecf_{\Omega}^{-1\prime}\right)'\left(vecf^{yy}\otimes f^{\varepsilon x\prime} + K_{pp}(f_{\varepsilon y}\otimes f_{yx})\right)d\lambda \end{aligned}$$

which establishes the first part of (3.1). Next turn to

$$\begin{aligned} &\vartheta_{j_1,j_2}E(v_{t_1,j_2}\ell'\bar{w}_{s_2,j_4})\vartheta_{j_3,j_4}E(v_{t_2,j_3}\ell'\bar{w}_{s_1,j_1}) \\ = &\vartheta_{j_1,j_2}(\ell'\Gamma^{\varepsilon x}_{t_1-s_2+m}\Gamma^{yy}_{t_1-s_2-j_2+j_4} + \Gamma^{\varepsilon y}_{t_1-s_2+m+j_4}\ell'\Gamma^{xy}_{s_2-t_1+j_2} + \mathcal{K}_4^2) \\ &\times\vartheta_{j_3,j_4}(\ell'\Gamma^{\varepsilon x}_{t_2-s_1+m}\Gamma^{yy}_{t_2-s_1-j_3+j_1} + \Gamma^{\varepsilon y}_{t_2-s_1+m+j_1}\ell'\Gamma^{xy}_{s_1-t_2+j_3} + \mathcal{K}_4^2) \end{aligned}$$

where for a typical term in this product we have

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

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

$$\begin{aligned}
&\left\| \sum_{u_1,u_2,u_3,j_4} \prod_{l=1}^{4} k(\tfrac{j_l}{M}) \vartheta_{u_1} \ell' \sum_{h_1} \left[(1 - \tfrac{|h_1|}{n}) \Gamma^{\varepsilon x}_{h_1+m} \Gamma^{yy}_{h_2+u_2} \right] \vartheta_{u_3} \sum_{k_2} (1 - \tfrac{|k_2 - j_1|}{n}) \Gamma^{\varepsilon y}_{k_2+m} \ell' \Gamma^{xy}_{k_2+u_1-u_2+u_3} \right\| \\
\leq & \sum_{u_1,u_2,u_3,j_4} \left| k(\tfrac{j_l}{M}) \right| \|\vartheta_{u_1}\| \left\| \ell' \sum_{h_1} (1 - \tfrac{|h_1|}{n}) \Gamma^{\varepsilon x}_{h_1+m} \Gamma^{yy}_{h_2+u_2} \right\| \|\vartheta_{u_3}\| \left\| \sum_{k_2} (1 - \tfrac{|k_2 - j_1|}{n}) \Gamma^{\varepsilon y}_{k_2+m} \ell' \Gamma^{xy}_{k_2+u_1-u_2+u_3} \right\| \\
= & \; O(M).
\end{aligned}$$

Similar arguments show that the second and third terms of $E\ell d_6 d_6' \ell$ are both $O(M/n)$. ■

Lemma A.33. *Let d_6 as defined in Lemma (A.24) and d_8 as defined in Lemma (A.26). Then $E\ell d_6 d_8' \ell = O(M/n)$ and $E\ell d_8 d_8' \ell = O(M/n)$.*

Proof. Consider

$$\begin{aligned}
E\ell d_6 d_8' \ell = & \frac{1}{n^3} \sum_{\substack{t_1, t_2, s_1, s_2 \\ j_1, .., j_4}}^{n} \prod_{l=1}^{4} k(\tfrac{j_l}{M}) \Big\{ \Big((vec\vartheta'_{j_1,j_2})' E(v_{t_1,j_2} \otimes \bar{w}'_{s_1,j_1}) (\ell' \Gamma_{j_4} \vartheta_{j_4,j_3} \otimes \ell) \\
& \times E(\bar{w}^y_{s_2,j_3,j_4} \otimes v'_{t_2,j_3}) vec\vartheta_{j_3,j_4} \Big) \\
& + tr \left[\vartheta_{j_1,j_2} E(v_{t_1,j_2} \ell' \Gamma_{j_4} \vartheta_{j_4,j_3} \bar{w}^y_{s_2,j_3,j_4}) \vartheta_{j_3,j_4} E(v_{t_2,j_3} \ell' \bar{w}_{s_1,j_1}) \right] \\
& + tr \left[\vartheta_{j_1,j_2} E(v_{t_1,j_2} v_{t_2,j_3}) \vartheta_{j_3,j_4} E(\bar{w}'_{s_1,j_1} \ell \ell' \Gamma_{j_4} \vartheta_{j_4,j_3} \bar{w}^y_{s_2,j_3,j_4}) \right] + \mathcal{K}_8 \Big\}
\end{aligned}$$

where the first term is contains

$$E(\bar{w}^y_{s_2,j_3,j_4} \otimes v'_{t_2,j_3}) = \Gamma^{\varepsilon y'}_{t_2-s_2+j_4} \otimes \Gamma^{yy}_{s_2-t_2} + (vec(\Gamma^{yy}_{t_2-s_2+j_4-j_3})' \otimes \Gamma^{\varepsilon y}_{s_2-t_2+j_3}) + \mathcal{K}^1_4$$

leading to

$$\begin{aligned}
& \left\| \sum_{u,j_4}^{n} k(\tfrac{j_4-u}{M}) k(\tfrac{j_4}{M}) (\ell' \Gamma_{j_4} \vartheta_u \otimes \ell) \sum_{k} (1 - \tfrac{|k|}{M}) (\Gamma^{\varepsilon y'}_{k+j_4} \otimes \Gamma^{yy}_{k}) vec\vartheta_u \right\| \\
\leq & \sum_{u,j_4}^{n} \|\ell' \Gamma_{j_4} \vartheta_u \otimes \ell\| \|vec\vartheta_u\| \sum_{k} \left\| (\Gamma^{\varepsilon y'}_{k+j_4} \otimes \Gamma^{yy}_{k}) \right\| = O(1)
\end{aligned}$$

with the same holding for $vec(\Gamma^{yy}_{t_2-s_2+j_4-j_3})' \otimes \Gamma^{\varepsilon y}_{s_2-t_2+j_3}$. This establishes that the first term is of lower order. The second term is of lower order by the same arguments used before. The third term is again of lower order showing that $E\ell d_6 d_8' \ell = O(M/n)$.

Finally $E\ell d_8 d_8' \ell = O(M/n)$ by the same arguments as before. ■

Lemma A.34. *Let* $H_{222} = \sum_{i=0}^{n} \sum_{j=0}^{n} \left(\tilde{\Gamma}_i^{xy} - \Gamma_i^{xy}\right) k(i/M)\vartheta_{i,j} k(j/M)\tilde{\Gamma}_{-j}^{yx}$*. Then*

$$E\ell' H_{222} D^{-1} d_0 d_0 D^{-1} H_{222}\ell = O(n^{-1}).$$

Proof. By Lemma (A.14) we can replace H_{222} by

$$\sum_{i=0}^{n} \sum_{j=0}^{n} \left(\tilde{\Gamma}_i^{xy} - \Gamma_i^{xy}\right) k(i/M)\vartheta_{i,j} k(j/M)\Gamma_{-j}^{yx} + o(M/n^{-1/2}). \tag{A.3}$$

Next define $a_j = \sum_{i=0}^{\infty} \Gamma_i^{xy} \vartheta_{i,j}$. Then, using only the dominant term in (A.3),

$$EH_{222} D^{-1} d_0 d_0 D^{-1} H_{222}' = n^{-3} \sum_{\substack{t_1, t_2, s_1, s_2 \\ j_1, ..., j_6}} \prod_{l=1}^{4} k(\tfrac{j_l}{M}) \bar{\omega}_{s_1,j_1} a_{j_1}' D^{-1} a_{j_2} v_{t_1,j_2} v_{t_2,j_3}' a_{j_3}' D^{-1} a_{j_4} \bar{\omega}_{s_2,j_4}'$$

Using the same arguments as in the proof of Lemma (A.32) it follows that the leading term in $E\ell' H_{222} D^{-1} d_0 d_0 D^{-1} H_{222}'\ell$ depends on

$$\mathcal{A}_2 = \sum_{j_1 j_2 = 0}^{\infty} \sum_{h=-\infty}^{\infty} vec(a_{j_2}' D^{-1} a_{j_1})' \left[\left(vec\Gamma_{h+j_1-j_2}^{yy} \otimes \Gamma_{h+m}^{\varepsilon x\prime}\right) + K_{pp}(\Gamma_{h+j_1}^{\varepsilon y} \otimes \Gamma_{h-j_2}^{yx})\right]$$

where $\mathcal{A}_2$ is well defined due to the summability properties of a_j. It follows that $E\ell' H_{222} D^{-1} d_0 d_0 D^{-1} H_{222}'\ell = O(n^{-1})$. ■

Lemma A.35. *Let* H_{222} *be as defined in Lemma (A.34) and* $d6$ *as defined in Lemma (A.24). Then* $E\ell' H_{222} D^{-1} d_0 d_6'\ell = O(M/n)$*.*

Proof. We again replace H_{222} by $\sum_{i=0}^{n} \sum_{j=0}^{n} \left(\tilde{\Gamma}_i^{xy} - \Gamma_i^{xy}\right) k(i/M)\vartheta_{i,j} k(j/M)\Gamma_{-j}^{yx}$ and consider

$$E\ell' H_{222} D^{-1} d_0 d_6'\ell = n^{-3} \sum_{\substack{t_1, t_2, s_1, s_2 \\ j_1, ..., j_6}} E \prod_{l=1}^{4} k(\tfrac{j_l}{M}) \bar{\omega}_{s_1,j_1} a_{j_1}' D^{-1} a_{j_2} v_{t_1,j_2} v_{t_2,j_3}' \vartheta_{j_3,j_4} \bar{\omega}_{s_2,j_4}'$$

where a_j is defined in Lemma (A.34). The dominant term in this expectation is given by $M/n \int k(x)^2 dx \mathcal{A}_2 \ell\ell' \mathcal{A}_1$ where $\mathcal{A}_2$ is defined in Lemma (A.34) and $\mathcal{A}_1$ is defined in (3.1). ■

Lemma A.36. *Let* H_{222} *be as defined in Lemma (A.34) and* $d8$ *as defined in Lemma (A.26). Then* $E\ell' H_{222} D^{-1} d_0 d_8'\ell = O(n^{-1})$*.*

Proof. By the same arguments as in the proof of Lemmas (A.34) and (A.30). ■

Proof of Lemma (3.1). First we will split the error $\hat{D}_M - D$ into three different parts. Recall $\hat{D}_k = n^{-2} X' Z_M (k_M \otimes I) \hat{\Omega}_{n,M}^{-1} (k_M \otimes I)' Z_M' X$. Let Ω_k^{-1} be the upper left side $kp \times kp$ block of the infinite dimensional inverse Ω^{-1}. The difference $\hat{D}_k - D$ is decomposed into the following terms

$$\hat{D}_k - D = H_1 + H_2 + H_3 + H_4$$

where $H_1 = H_{11} + H_{12} + H_{13} + H_{14}$ which are defined in Lemmas (A.9)-(A.11), $H_2 = H_{211} + H_{212} + H_{221} + H_{222}$ which are defined in Lemmas (A.12)-(A.15), $H_3 = H_{31} + H_{32} + H_{33} + H_{34}$ which are defined in Lemmas (A.16)-(A.17) and H_4 which is a remainder term of lower order.

Next we turn to the analysis of $\hat{d}_M = \hat{P}_M'(k_m \otimes I_p)\hat{\Omega}_n^{-1} n^{-1/2} \sum_{t=1}^{n-m} \varepsilon_{t+m} \bar{z}_{t,M}(k_M \otimes I_p)$ which is decomposed as $\hat{d}_k = \sum_j^9 d_j$ with $d_4 = d_{41} + d_{42}$ where all the terms are defined in Lemmas (A.18)-(A.27).

We consider cross products of the form $Ed_i d_j'$, $Ed_i d_0' D^{-1} H_i$ and $EH_i D^{-1} d_0 d_0' D^{-1} H_j$ which depend on M and n and are largest in probability. Lemmas (A.28)-(A.33) show that the largest terms vanishing as $M \to \infty$ are $E\left(d_0 d_{42} + d_{42} d_0'\right) = -M^{-q} k_q \mathcal{B}_2^{(q)}$ and $Ed_0 d_0' D^{-1}(H_{13} + H_{14}) = M^{-q} k_q \mathcal{B}_2^{(q)}$ which cancel because they are of opposite sign.

Terms of order M^{-2q} include $Ed_0 d_{42}'$, $Ed_0 d_{41}'$, $Ed_0 d_0' D^{-1} H_{12}'$ and $H_{12} D^{-1} Ed_0 d_0$. Since $Ed_0 d_{42}'$ and $Ed_0 d_0' D^{-1} H_{12}'$ are of opposite sign these terms cancel. We are left with $E(d_{42} - (H_{13} + H_{14}) D^{-1} d_0)(d_{42} - (H_{13} + H_{14}) D^{-1} d_0)' = O(M^{-2q})$.

Terms that grow with M and are highest in order are $H_{222} D^{-1} d_0$ and d_6. It follows that the cross product term $EH_{222} D^{-1} d_0 d_6'$ is of lower order by Lemma (A.35). We are left with $EH_{222} D^{-1} d_0 d_0' D^{-1} H_{222}' = O(n^{-1})$ by Lemma (A.34) and $Ed_6 d_6' = O(M^2/n)$. ■

Proof of Proposition (3.3) From the proof of Lemma (**3.1**) we only need to consider the terms $A_n = Ed_6 d_6'$ and $B_n = E(d_{42} - (H_{13} + H_{14}) D^{-1} d_0)(d_{42} - (H_{13} + H_{14}) D^{-1} d_0)'$. Since for all $n \geq 1$ we have $A \geq 0$ and $B \geq 0$ it follows that $\liminf_n A_n \geq 0$ and $\liminf_n B_n \geq 0$ such that $\mathcal{A}$ and $\mathcal{B}^{(q)}$ are nonnegative.

From Lemma (A.32) it follows that $E\ell' D^{-1/2} d_6 d_6' D^{-1/2} \ell = M^2/n \left(\int k^2(x) dx\right)^2 \mathcal{A}_1 D^{-1/2} \ell\ell' D^{-1/2} \mathcal{A}_1'$. It can be shown that $B_n = Ed_{42} d_{42}' - E(H_{13} + H_{14}) D^{-1} d_0 d_0' D^{-1}(H_{13} + H_{14})$ where $Ed_{42} d_{42}' = \mathcal{B}_1^{(q)}$ defined in (3.2) and $E(H_{13} + H_{14}) D^{-1} d_0 d_0' D^{-1}(H_{13} + H_{14})$ defined in (3.3) using the same arguments as in Lemma (A.32). ■

Proof of Proposition (4.1) The only difficulty here is to show that $\sum \hat{\zeta}_j \hat{\Gamma}_j^{ab}$ is $\sqrt{n}$-consistent. Let $\tilde{\beta}$ be a $\sqrt{n}$-consistent first stage estimate. The estimated residuals $\hat{\varepsilon}_t = (y_t - \bar{y}) - \tilde{\beta}'(x_t - \bar{x})$ are used to estimate $\hat{\zeta}_j$. Let $g(\lambda, \theta) = \left|1 - \theta_1 e^{i\lambda} - \ldots \theta_{m-1} e^{i\lambda(m-1)}\right|^2$. The periodogram of $\hat{\varepsilon}_t$ is $\hat{I}_n^{\varepsilon}(\lambda) = n^{-1} \sum_{t,s} \hat{\varepsilon}_t \hat{\varepsilon}_s e^{i\lambda(t-s)}$. The maximum likelihood estimator for θ is asymptotically equivalent to $\arg\min_\theta \Lambda_n^{\hat{\varepsilon}}(\theta)$ with $\Lambda_n^{\hat{\varepsilon}}(\theta) = n^{-1} \sum_j \hat{I}_n^{\varepsilon}(\lambda_j)/g(\lambda_j, \theta)$ for $\lambda_j = 2\pi j/n$, $j =$

$-n+1,...,0,...,n-1$. Define $I_n^{\varepsilon}(\lambda)=n^{-1}\sum_{t,s}\varepsilon_t\varepsilon_s e^{i\lambda(t-s)}$, $I_n^{\varepsilon x}(\lambda)=n^{-1}\sum_{t,s}\varepsilon_t(x_s-\mu_x)e^{i\lambda(t-s)}$, $I_n^{x\alpha}(\lambda)=n^{-1}(\hat{\alpha}_0-\alpha_0)\sum_{t,s}(x_t-\mu_x)e^{i\lambda(t-s)}$, $I_n^{\varepsilon\alpha}(\lambda)=n^{-1}(\hat{\alpha}_0-\alpha_0)\sum_{t,s}\varepsilon_t e^{i\lambda(t-s)}$ and $I_n^{\alpha}(\lambda)=n^{-1}(\hat{\alpha}_0-\alpha_0)^2\sum_{t,s}e^{i\lambda(t-s)}$ for $\hat{\alpha}_0-\alpha_0=\bar{y}-\mu_y-\beta'(\bar{x}-\mu_x)$. It follows that

$$\begin{aligned}\hat{I}_n^{\varepsilon}(\lambda) &= I_n^{\varepsilon}(\lambda)+(\tilde{\beta}-\beta)'(I_n^{\varepsilon x}(\lambda)+I_n^{x\varepsilon}(\lambda))+(\tilde{\beta}-\beta)'I_n^{x}(\lambda)(\tilde{\beta}-\beta)\\ &\quad +(\hat{\alpha}_0-\alpha_0)(I_n^{\varepsilon\alpha}(\lambda)+I_n^{\alpha\varepsilon}(\lambda))+(\tilde{\beta}-\beta)'(I_n^{\alpha x}(\lambda)+I_n^{\alpha x}(\lambda))+I_n^{\alpha}(\lambda).\end{aligned}$$

Note that $I_n^{\alpha}(\lambda_j)=I_n^{x\alpha}(\lambda_j)=I_n^{\varepsilon\alpha}(\lambda_j)=0$ for $j\neq 0$ and $I_n^{\alpha}(\lambda_j)=n(\hat{\alpha}_0-\alpha_0)^2$, $I_n^{\varepsilon\alpha}(\lambda)=(\hat{\alpha}_0-\alpha_0)\sum_t\varepsilon_t$ for $j=0$. We now have

$$\begin{aligned}\Lambda_n^{\hat{\varepsilon}}(\theta) &= \Lambda_n^{\varepsilon}(\theta)+(\tilde{\beta}-\beta)'(\Lambda_n^{\varepsilon x}(\theta)+\Lambda_n^{x\varepsilon}(\theta))+(\tilde{\beta}-\beta)'\Lambda_n^{x}(\theta)(\tilde{\beta}-\beta)\\ &\quad +\left[(\hat{\alpha}_0-\alpha_0)n^{-1}\sum_t(\varepsilon_t+(x_t-\mu_x))+(\hat{\alpha}_0-\alpha_0)^2\right]/g(0,\theta).\end{aligned}$$

From standard arguments (see Brockwell and Davis 1989, ch 10) it follows that $\Lambda_n^{ab}(\theta)\overset{a.s.}{\rightarrow}\Lambda^{ab}(\theta)$ with $\Lambda^{ab}(\theta)=\int f_{ab}(\lambda)g(\lambda,\theta)d\lambda$ and $\partial^k\Lambda_n^{ab}(\theta)/\partial\theta\overset{a.s.}{\rightarrow}\partial^k\Lambda^{ab}(\theta)/\partial\theta$ for $k<\infty$. Moreover $\sqrt{n}\partial\Lambda_n^{\varepsilon}(\theta)/\partial\theta=O_p(1)$ and $n^{-1/2}\sum_t\varepsilon_t=O_p(1)$ and $n^{-1/2}\sum_t x_t-\mu_x=O_p(1)$. Therefore

$$\sqrt{n}\partial\Lambda_n^{\hat{\varepsilon}}(\theta)/\partial\theta=\sqrt{n}\partial\Lambda_n^{\varepsilon}(\theta)/\partial\theta+\sqrt{n}(\tilde{\beta}-\beta)'(\partial(\Lambda^{\varepsilon x}(\theta)+\Lambda^{\varepsilon x}(\theta))/\partial\theta+o_p(1)$$

which shows that $\sqrt{n}(\tilde{\theta}-\theta)=O_p(1)$. Expanding $\zeta_k=(2\pi)^{-2}\sigma^4\sum_{j=0}^{\infty}e_1'B^je_1e_1'B^{j+k}e_1$ around θ_0 leads to

$$\begin{aligned}\tilde{\zeta}_k &= \zeta_k+(2\pi)^{-2}\sigma^4\sum_{j=0}^{\infty}e_1'B^{j+k}e_1\sum_{l=0}^{j}(e_1'B^{l-1}\otimes e_1'B^{j-l})\frac{\partial vecB}{\partial\theta}(\tilde{\theta}-\theta_0)\\ &\quad +(2\pi)^{-2}\sigma^4\sum_{j=0}^{\infty}e_1'B^je_1\sum_{l=0}^{j+k}(e_1'B^{l-1}\otimes e_1'B^{k+j-l})\frac{\partial vecB}{\partial\theta}(\tilde{\theta}-\theta_0)+o_p(n^{-1/2}).\end{aligned}$$

The matrices B are evaluated at θ_0 such that all the sums are absolutely convergent. It then follows that $\tilde{\zeta}_k-\zeta_k=O_p(n^{-1/2})$ uniformly in k.

We next show that $\sum_{k=-n+1}^{n-1}\tilde{\zeta}_k\hat{\Gamma}_{k+r}^{\varepsilon y}-\sum_k\zeta_k\Gamma_{k+r}^{\varepsilon y}=O_p(n^{-1/2})$. First consider

$$\begin{aligned}P(\sum_{k=-n+1}^{n-1}\left\|\tilde{\zeta}_k-\zeta_k\right\|\left\|\hat{\Gamma}_{k+r}^{\varepsilon y}\right\| &> \eta)\leq P(Cn^{-1/2}\sum_{k=-n+1}^{n-1}\left\|\hat{\Gamma}_{k-r}^{\varepsilon y}\right\|>\eta)\\ +P(\sup_k\left\|\tilde{\zeta}_k-\zeta_k\right\| &> Cn^{-1/2}).\end{aligned}$$

The second probability goes to zero by the previous result. Also

$$\frac{1}{n^2}\sum_{t,s}E\varepsilon_t(y_{t-k-r}-\mu_y)\varepsilon_s(y_{s-k-r}-\mu_y)'=\frac{1}{n}\sum_j(1-\tfrac{|j|}{n})(\gamma_j^{\varepsilon}\Gamma_{j-k-r}^{yy}+\Gamma_{k+r}^{\varepsilon y}\Gamma_{k+r}^{\varepsilon y}+\Gamma_{j+k+r}^{\varepsilon y}\Gamma_{j+k+r}^{\varepsilon y}+\mathcal{K}_4^3)$$

with similar expressions holding for other terms involving ε_t, x_t and y_t such that $E\sum_{k=-n+1}^{n-1}\left\|\hat{\Gamma}_{k+r}^{\varepsilon y}\right\| = O(1)$. Next consider $\sum_{k=-n+1}^{n-1}\zeta_k(\hat{\Gamma}_{k+r}^{\varepsilon y}-\Gamma_{k+r}^{\varepsilon y})$. By the previous result it follows that $\left\|\hat{\Gamma}_{k+r}^{\varepsilon y}-\Gamma_{k+r}^{\varepsilon y}\right\| = O_p(n^{-1/2})$ uniformly in k. The result then follows since ζ_k is summable. ■

Proof of Theorem (4.2) To be typed. ■

Proof of Proposition (5.1) We consider Ed_i and EH_iDd_j. First, $Ed_i = 0$ for $i \leq 4$. The terms d_5, d_7, d_8 and d_9 are of lower order by Lemmas (A.23,A.25-A.27). The terms EH_iDd_j are all of lower order. The largest order term is therefore Ed_6. By the proof of Lemma (A.32) it follows that $Ed_6 = M/\sqrt{n}\mathcal{A}_1\int k^2(x)dx + o(M/\sqrt{n})$. ■

Proof of Theorem (5.3) We consider the expansion of $\sqrt{n}\left(\beta_{n,M}^{*}-\beta\right)$as before. First note that

$$\sqrt{n}\left(\beta_{n,M}^{*}-\beta\right)=\sqrt{n}\left(\beta_{n,M}-\beta\right)-\frac{M}{\sqrt{n}}D^{-1}\mathcal{A}_1'\int k^2(x)dx+O_p(M/n)$$

since $\sqrt{n}(\widehat{\mathcal{A}}_1-\mathcal{A}_1) = O_p(n^{-1/2})$. The analysis of the MSE of $\sqrt{n}\left(\beta_{n,M}^{*}-\beta\right)$ is then the same as the analysis for $\sqrt{n}\left(\beta_{n,M}-\beta\right)$ where we replace d_6 by

$$\bar{d}_6 = d_6-\frac{M}{\sqrt{n}}\mathcal{A}_1'\int k^2(x)dx$$

and the additional term $d_{10} = M/\sqrt{n}(\widehat{\mathcal{A}}_1-\mathcal{A}_1)\int k^2(x)dx$ needs to be considered. First note that $Ed_6-\frac{M}{\sqrt{n}}\mathcal{A}_1'\int k^2(x)dx = o(1)$. Then $E\bar{d}_6\bar{d}_6' = E(d_6-Ed_6)(d_6-Ed_6)'+o(1)$. From the proof of Lemma (A.32) it follows that $E\bar{d}_6\bar{d}_6' = O(M/n)$. Also $E\ell' H_{222}D^{-1}d_0\bar{d}_6'D^{-1/2}\ell = o(M/n)$ and $Ed_{10}d_{10} = O(M/n)$ together with Lemma (A.34) shows that all remaining terms are at most of order M/n. ■

References

[1] D. W. Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858, 1991.

[2] D. W. Andrews. Consistent moment selection procedures for generalized method of moments estimation. *Econometrica*, pages 543–564, 1999.

[3] D. W. K. Andrews. Asymptotics for semiparametric econometric models via stochastic equicontinuity. *Econometrica*, 62:43–72, 1994.

[4] D. R. Brillinger. *Time Series, Data Analysis and Theory, Expanded Edition*. Holden-Day, Inc., 1981.

[5] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer Verlag-New York, Inc., 1987.

[6] G. Chamberlain. Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, 34:305–334, 1987.

[7] S. G. Donald and W. K. Newey. Choosing the number of instruments. *Draft, MIT*, 1997.

[8] L. P. Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1053, 1982.

[9] L. P. Hansen. A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators. *Journal of Econometrics*, 30:203–238, 1985.

[10] L. P. Hansen, J. C. Heaton, and M. Ogaki. Efficiency bounds implied by multiperiod conditional moment restrictions. *Journal of the American Statistical Association*, 83:863–871, 1988.

[11] L. P. Hansen and K. J. Singleton. Computing semiparametric efficiency bounds for linear time series models. In W. A. Barnett, J. Powell, and G. Tauchen, editors, *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, pages 387–412, 1991.

[12] L. P. Hansen and K. J. Singleton. Efficient estimation of linear asset-pricing models with moving average errors. *Journal of Business and Economic Statistics*, pages 53–68, 1996.

[13] F. Hayashi and C. Sims. Nearly efficient estimation of time series models with predetermined, but not exogenous, instruments. *Econometrica*, 51(3):783–798, 1983.

[14] G. M. Kuersteiner. Linear covariance matrix lower bounds for time series estimators. *Manuscript*, 1997.

[15] G. M. Kuersteiner. Optimal instrumental variables estimation for arma models. *MIT Working Paper*, 1998.

[16] O. Linton. Seocnd order approximation in the partially linear regression model. *Econometrica*, pages 1079–1112, 1995.

[17] O. B. Linton. Second order approximation for semiparametric intrumental variable estimators and test statistics. *mimeo Yale University*, 1997.

[18] J. R. Magnus and H. Neudecker. The commutation matrix: Some properties and applications. *Annals of Statistics*, 7(2):381–394, 1979.

[19] W. K. Newey. Adaptive estimation of regression models via moment restrictions. *Journal of Econometrics*, 38:301–339, 1988.

[20] W. K. Newey. Efficient instrumental variables estimation of nonlinear models. *Econometrica*, 58:809–837, 1990.

[21] W. K. Newey. The asymptotic variance of semiparametric estimators. *Econometrica*, pages 1349–1382, 1994.

[22] E. Parzen. On consistent estimates of the spectrum of a stationary time series. *Annals of Statistics*, 28:329–348, 1957.

[23] P. Stoica, T. SŤderstrŤm, and B. Friedlander. Optimal instrumental variable estimates of the ar parameters of an arma process. *IEEE Transactions on Automatic Control*, 30:1066–1074, 1985.

[24] Z. Xiao and P. C. Phillips. Higher order approximations for frequency domain time series regression. *Journal of Econometrics*, pages 297–336, 1998.