MEAN SQUARED ERROR REDUCTION FOR GMM ESTIMATORS OF LINEAR TIME SERIES MODELS

By Guido M. Kuersteiner*

March 15, 2002

Abstract

In this paper we analyze Generalized Method of Moments (GMM) estimators for time series models as advocated by 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 Mean Squared Error (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.

An alternative to direct bias correction are k-class estimators introduced by Nagar. This approach is adapted to the time series case. The time series k-class estimator also corrects for the largest order bias and achieves an accelerated rate of convergence for the higher order asymptotic terms.

Key Words: time series, feasible GMM, number of instruments, bias correction, k-class

*I wish to thank Xiaohong Chen, Ronald Gallant, Jinyong Hahn, Jerry Hausman, Whitney Newey and Ken West for helpful comments. Financial support from NSF grant SES-0095132 is gratefully acknowledged. The views expressed in this paper are not necessarily shared by NSF. All remaining errors are my own.

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, Soderstrom 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 the context of time series models. 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, Soderstrom and Friedlander (1985) for the Autoregressive-Moving-Average (ARMA) model of order (1,1) 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 (1997,1999a) 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

MSE. The idea behind using the kernel weighted version of the GMM estimator is to dampen the importance of these bias terms and thus allow 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.

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 of the approximation around the limiting normal distribution 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 Limited Information Maximum :Likelihood (LIML) estimators and $O(n^{-\frac{2s}{2s+2d}})$ for Two Stage Least Squares (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 is asset pricing models.

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 is $O(n^{-\frac{2s}{2s+2}})$ 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. The 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 introduces the kernel weighted GMM estimator, 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. The proofs are collected in Appendix A and additional Lemmas are given in Appendix B.

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 $Ey_t^2 < \infty$. 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 $\mathcal{F}_t = \sigma(y_t, y_{t-1}, ...)$. Assume that 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 and conditionally homoskedastic martingale difference sequence. 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 a Moving-Average (MA) process of order m-1. 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 is Asset Pricing models. 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.

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 $\Xi = \left\{\xi \in \mathbb{R}^{r+1} \times \mathbb{R}^p \left| \xi_0 - \xi_1 z - ... - \xi_r z^r \neq 0, |z| \leq 1, \xi_j \in \mathbb{R}^p \right.\right\}$. 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. $\beta_i = \operatorname{vec} \delta_i(\beta) \in \mathbb{R}^p$. A normalization restriction $\beta_{0,1} = 1$ is imposed where $\beta_{0,1}$ is the first element of β_0.*

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 $|A|^2 = \operatorname{tr} AA^*$ 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 strictly stationary and ergodic, with $E\left(u_t|\mathcal{F}_{t-1}\right) = 0$, $E\left(u_t u_t'|\mathcal{F}_{t-1}\right) = \Sigma$ where Σ is a positive definite symmetric matrix of nonrandom constants. Let u_t^i be the i-th element of u_t and $\operatorname{cum}_{i_1,...,i_k}(t_1, ..., t_{k-1})$ the k-th order cross cumulant of $u_{t+t_1}^{i_1}, ..., u_t^{i_k}$ defined in*

(B.1) in the Appendix. Assume that

$$\sum_{t_1=-\infty}^{\infty} \cdots \sum_{t_{k-1}=-\infty}^{\infty} \left| \operatorname{cum}_{i_1,\ldots,i_k}(t_1,\ldots,t_{k-1}) \right| < \infty \text{ for } k \leq 8.$$

The fact that u_t is a martingale difference sequence arises naturally in rational expectations models. In our context it is needed together with the conditional homoskedasticity assumption to guarantee that the optimal GMM weight matrix is of a sufficiently simple form. This allows to construct estimates of the bias terms converging fast enough to increase the optimal rate of convergence to the asymptotic limit distribution of a bias corrected GMM estimator.

The conditional homoskedasticity condition $E(u_t u_t' | \mathcal{F}_{t-1}) = E u_t u_t'$ is restrictive as it rules out time changing variances. Relaxing this restriction results in more complicated GMM weight matrices of the type analyzed in Kuersteiner (1997, 1999b). In principle the higher order moment restriction implied by conditional homoskedasticity 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. The cumulant summability condition used here is similar but slightly stronger than the second part of Condition A in Andrews (1991). What is needed both in Andrews (1991) and here are restrictions on the eighth-moment dependence of the underlying process 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 $s > 1 + \delta$ and some $\delta > 0$.*

Note that $s = \infty$ if y_t follows a vector ARMA process. The following definitions will be used throughout the paper and are given next. Let y_t satisfy Assumption (C(s)). Partition $y_t = \left(y_t^1, y_t^{2\prime}\right)'$ where y_t^1 is the first element of y_t. Then define $x_t = \left(y_t^{2\prime}, y_{t-1}', \ldots, y_{t-r}'\right)'$. Let $\mu_y = E y_t$ and $\mu_x = E x_t$. Define $w_{t,i} = (x_{t+m} - \mu_x)\left(y_{t-i+1} - \mu_y\right)'$, $\Gamma_i^{xy} = E w_{t,i}$, $\Gamma_{-i}^{yx} = E w_{t,i}'$ and let $\breve{w}_{t,i} = w_{t,i} - \Gamma_i^{xy}$. Next define $w_{t,j-i}^y = \left(y_{t-i} - \mu_y\right)\left(y_{t-j} - \mu_y\right)'$ with $E w_{t,j-i}^y = \Gamma_{j-i}^{yy}$. Let $\varepsilon_t = \Phi(L) u_t$ and define $v_{t,i} = \varepsilon_{t+m}(y_{t-i+1} - \mu_y)$. Also define $E \varepsilon_t x_s = \Gamma_{t-s}^{\varepsilon x}$ and $E \varepsilon_{t+m} y_s = \Gamma_{t-s}^{\varepsilon y}$. For $a, b \in \{"x", "y", "\varepsilon"\}$ we define the following second order spectral densities

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

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 $\sum_{j=-\infty}^{\infty} |j|^s |\zeta_j|$ for all $s < \infty$.*

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 $[M]$ lagged observations as instruments where $[M]$ denotes the largest integer smaller than $M \in \mathbb{R}$. We therefore define instrument vectors $z_{t,M} = (y_t', y_{t-1}', ..., y_{t-[M]+1}')'$ and let 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.

For this purpose let $\mathbf{1}_{[M]}$ be an $[M] \times 1$ vector of ones, $\Omega_M = \sum_{l=-m+1}^{m-1} \gamma_l^\varepsilon \Omega_M(l)$ with $\Omega_M(l) = E z_{t,M} z_{t-l,M}'$, $P_M' = E(x_{t+m} - \mu_x)(z_{t,M} - \mathbf{1}_{[M]} \otimes \mu_y)'$ and $D_M = P_M' \Omega_M^{-1} P_M$. Also denote the i,j-th $p \times p$ block of $\Omega_M(l)$ by $\omega_{i,j}(l)$.The infeasible estimator of β is based on a nonrandom matrix $D_M^{-1} P_M' \Omega_M^{-1}$ and is given by

$$\beta_{n,M} = D_M^{-1} P_M' \Omega_M^{-1} \frac{1}{n} \sum_t (y_t^1 - \mu_y^1)(z_{t,M} - \mathbf{1}_{[M]} \otimes \mu_y)$$

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 $x = \{x_i\}_{i=1}^\infty$ with elements $x_i \in \mathbb{R}^p$ such that $x \in l^2$ if $\sum_i \|x_i\| < \infty$. We define the operator Ω component-wise by its image for all $x \in l^2$ by $b_i = \lim_{m\to\infty} \sum_j^m \omega_{i,j} x_j$ 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. $x\Omega^{-1} \in l^2$ for all $x \in l^2$. Denote by $\vartheta_{k,j}$ the k,j-th element of Ω^{-1}. From Whittle (1951) it is well known that $\vartheta_{k,j} = \int_{-\pi}^{\pi} f_\Omega^{-1}(\lambda) e^{i\lambda(j-k)} d\lambda$ such that $\vartheta_{k,j} \equiv \vartheta_{j-k}$. 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. Details of these arguments can be found in Kuersteiner (1999b).

Let $D = \lim_M P_M' \Omega_M^{-1} P_M = P'\Omega^{-1}P$ and $d_0 = \lim_M P_M' \Omega_M^{-1} \frac{1}{\sqrt{n}} \sum (z_{t,M} - \mathbf{1}_{[M]} \otimes \mu)\varepsilon_t$ almost surely. It can be shown that $D^{-1}d_0 \xrightarrow{p} N(0, D^{-1})$ as $n \to \infty$ under the assumptions made about y_t. It is also true that $\sqrt{n}\left(\hat{\beta}_M - \beta\right) - D^{-1}d_n \xrightarrow{p} 0$ as $n, M \to \infty$. The last statement is no longer true, at least without specifying the rate at which M goes to infinity, once we replace $\beta_{n,M}$ by a feasible estimator.

A feasible version of $\beta_{n,M}$ 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 GMM efficiency lower bound as long as there are no additional restrictions placed on the lag polynomial $C(L)$.

Once the infeasible estimator has been replaced by a feasible version 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 the resulting parameter estimates $\hat{\beta}_{n,M}$.

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. A fully data dependent procedure is developed in Section (4).

3. Kernel Weighted GMM

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

$$k(j/M)E\varepsilon_{t+m}y_{t-j-1} = 0.$$

The truncated kernel is $k(j/M) = \{|j/M| \leq 1\}$ where we use $\{.\}$ to denote the indicator function. The general kernel weighted approach therefore covers the standard GMM procedure as a special case when the truncated kernel is used. 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.

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* $k : \mathbb{R} \mapsto [-1, 1]$, $k(0) = 1$, $k(x) = k(-x) \forall x \in \mathbb{R}$, $k(x) = 0$ *for* $|x| > 1$, $k(.)$ *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) except that we also require $k(x) = 0$ for $|x| > 1$. This assumption ensures that only a finite number of moment conditions, controlled by the bandwidth parameter, are used in estimation. The assumption could be relaxed at the cost of having to introduce additional bandwidth parameters for estimation of the optimal weight matrix. This seems unattractive from a practical point of view and is not pursued here.

Assumption (E) rules out certain parametric kernel functions such as the Quadratic Spectral kernel but is satisfied by a number of well known kernels such as the Truncated, Bartlett, Parzen and Tukey-Hanning kernels. Assumption (F) rules out the Truncated kernel. For the Parzen and Tukey-Hanning kernels $q = 2$ and for the Bartlett kernel $q = 1$.

We define the matrix

$$k_M = \text{diag}(k(1/M), ..., k(1))'$$

having kernel weight $k(j/M)$ in the j-th diagonal element and zeros otherwise. An instrument selection matrix $S_M(t) = \text{diag}(\{t \geq 1\}, ...\{t \geq [M]\})$ is introduced to exclude instruments for which there is no data in the sample. The vector of available instruments is denoted by $\bar{z}_{t,M} = S_M(t)(z_{t,M} - \mathbf{1}_{[M]} \otimes \bar{y})$. The empirical analogue to the moment condition is then

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

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 $k_M = \mathbf{1}_{[M]}$ such that

$$K_M = I_{[M]p}.$$

Given the definition of the instrument vector $\bar{z}_{t,M}$ one has to estimate an M dimensional covariance matrix Ω^{-1}. We define $\hat{\Omega}_M(j) = \frac{1}{n} \sum_{t=\max(1+j,1)}^{\min(n,n+j)} \bar{z}_{t,M} \bar{z}'_{t-j,M}$. We denote the l,k-th block of $\hat{\Omega}_M(j)$ by $\hat{\omega}_{l,k}(j) = \frac{1}{n} \sum_{t=\max(1+j+k,1+l)}^{\min(n,n+j-k)} (y_{t-l} - \bar{y}) (y_{t-j-k} - \bar{y})'$. The optimal weight matrix is then given by

$$\hat{\Omega}_M = \sum_{j=-m+1}^{m-1} \hat{\gamma}^{\varepsilon}(j) \hat{\Omega}_M(j) \tag{3.1}$$

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 $\tilde{\beta}_{n,M}$. The l,k-th block of $\hat{\Omega}_M$ is defined correspondingly as $\hat{\omega}_{l,k} = \sum_{j=-m+1}^{m-1} \hat{\gamma}^{\varepsilon}(j) \hat{\omega}_{l,k}(j)$. Note that $\hat{\Omega}_n$ is symmetric but not necessarily positive definite. This is unimportant as long as the estimator $\hat{\beta}_{n,M}$ 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 $M \geq d/p$. Under Assumption (A) the structural parameters β are identified and $\hat{\beta}_{n,M}$ has a closed form expression. Let Z_k be the matrix of stacked instruments $Z_M = [\bar{z}_{\max(1,r-m+1),M}, ..., \bar{z}_{n-m,M}]'$ and $X = [x_{\max(m+1,r+1)} - \bar{x}, ..., x_n - \bar{x}]'$ the matrix of regressors. Also, Y is the stacked vector of the first demeaned element in y_t. Then define the $d \times p\,[M]$ matrix

$$\hat{P}'_M = n^{-1} X' Z_M \tag{3.2}$$

with elements $\hat{\Gamma}_j^{xy} = \frac{1}{n} \sum_{t=\max(j+1,r+1-m)}^{n-m} (x_{t+m} - \bar{x})(y_{t-j} - \bar{y})'$. The estimator $\beta_{n,M}$ can now be written as

$$\hat{\beta}_{n,M} = \left(\hat{P}'_M K_M \hat{\Omega}_M^{-1} K_M \hat{P}_M \right)^{-1} \hat{P}'_M K_M \hat{\Omega}_M^{-1} K_M \frac{Z'_M Y}{n} \tag{3.3}$$

where $\hat{P}'_M K_M \hat{\Omega}_M^{-1} K_M \hat{P} = \sum_{i,j=1}^{n} \hat{\Gamma}_i^{xy} k(i/M) \hat{\vartheta}_{i,j} k(j/M) \hat{\Gamma}_{-j}^{yx}$ and $\hat{\vartheta}_{i,j} = \left[\hat{\Omega}_M^{-1} \right]_{i,j}$ is the i,j-th block of $\hat{\Omega}_M^{-1}$. 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 from Lemmas (B.9-B.28) that $\left\| \hat{D}_M - D \right\| = O_p(M/n^{1/2}) = o_p(1)$ and $\left\| \hat{d}_M - d_0 \right\| = O_p(M/n^{1/2})$ where $\hat{D}_M = \hat{P}'_M K_M \hat{\Omega}_M^{-1} K_M \hat{P}_M$ and $\hat{d}_M = \hat{P}'_M K_M \hat{\Omega}_M^{-1} K_M \frac{Z'_n \varepsilon}{\sqrt{n}}$.

Representation (3.3) makes the effects of using kernel weighted moments transparent. In essence using the kernel weight matrix K_M introduces an inefficiency by using $K_M \hat{\Omega}_M^{-1} K_M$ instead of the optimal $\hat{\Omega}_M^{-1}$ as weight matrix. As Corollary (3.2) below shows, this inefficiency however is not related to the first order asymptotic properties of the estimator in the sense that $\hat{\beta}_{n,M}$ is first order asymptotically equivalent to $D^{-1}d_0$ as long as $M, n \to \infty$ and $M/n^{1/2} \to 0$.

The bandwidth parameter M is chosen such that the MSE of a weighted sum of the elements of $\beta_{n,M}$ is minimized. We approximate the MSE by first expanding $\beta_{n,M}$ around $\hat{\beta}_M$ 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 $M/\sqrt{n} \to 0$. We decompose the expansion into $\hat{D}_M - D = H_1 + ... + H_4$ and $\hat{d}_n = d_0 + d_1 + ... + d_9$ where H_i and d_i are defined in Equations (A.2) through (A.24) in Appendix A such that

$$\sqrt{n}(\beta_{n,M} - \beta) = D^{-1} \sum_{i=0}^{9} d_i - D^{-1} \sum_{i=1}^{4} \sum_{j=0}^{9} H_i D^{-1} d_j + o_p(M/\sqrt{n}).$$

We now denote by $\sqrt{n}\,(b_{n,M} - \beta)$ all the terms $D^{-1}d_i$ and $D^{-1}H_i D^{-1} d_j$ which are $O_p(M/\sqrt{n})$ or terms that are $O_p(M^{-2q})$. The remaining terms $R_{n,M} = \sqrt{n}(\beta_{n,M} - b_{n,M})$ are of order $o_p(M/\sqrt{n})$. The size of the mean squared error of the estimator is given in the next lemma. Define the approximate mean squared error of $\hat{\beta}_{n,M}$ as

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

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 $s > q$ 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 $O(M^{-2q})$ variance 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}. Intuitively, the kernel function distorts the optimal weight matrix resulting in an increased variance of higher order terms in the expansion. As will be shown, this increased variance is traded off against a reduction in the bias.

The second part of the Lemma shows that using the truncated 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 $M/\sqrt{n} \to 0$.

Corollary 3.2. *Suppose Assumptions (A), (B) and (C(s)) hold with $s > q$ and $k(.)$ satisfies Assumptions (E) and (F). If $n, M \to \infty$ and $M/\sqrt{n} \to 0$ as $n \to \infty$ then $\sqrt{n}\left(\hat{\beta}_{n,M} - \beta\right) - D^{-1}d_0 = o_p(1)$.*

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 $D^{-1}d_0$. The optimal rate of expansion for M is much slower than $o(\sqrt{n})$. The corollary also shows that the distortion introduced by using kernel weights thus effectively using an inefficient weight matrix are of second order and do not affect the first order asymptotic properties of $\hat{\beta}_{n,M}$ under the stated conditions.

The next proposition gives an expression for the asymptotic MSE using the largest in probability terms depending on M and n.

Proposition 3.3. *Suppose Assumptions (A), (B) and (C(s)) hold with $s > q$ and $k(.)$ satisfies Assumptions (E) and (F). If $n, M \to \infty$ and $M^{2q+2}/n \to \kappa$ with $0 < \kappa < \infty$ 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 $\mathcal{B}^{(q)} = 1/2\ell' D^{-1/2}(\mathcal{B}_2^{(q)} - \mathcal{B}_1^{(q)} D^{-1} \mathcal{B}_1^{(q)'}) D^{-1/2}\ell$ where $\mathcal{A}_1, \mathcal{B}_1^{(q)}$ and $\mathcal{B}_2^{(q)}$ are defined as

$$\mathcal{A}_1 = (2\pi)^2 \int_{-\pi}^{\pi} \left(\mathrm{vec}\, f_\Omega(\lambda)^{-1}\right)' \left(\mathrm{vec}\left(f_{yy}(\lambda)'\right) \otimes f_{\varepsilon x}(\lambda)'\right) d\lambda, \tag{3.4}$$

$$\mathcal{B}_1^{(q)} = \sum_{k=1,j=1}^{\infty} \left(\Gamma_k^{xy} \vartheta_{k,j} |j|^q \Gamma_{-j}^{yx} + \Gamma_k^{xy} |k|^q \vartheta_{k,j} \Gamma_{-j}^{yx} \right), \tag{3.5}$$

$$\begin{aligned} \mathcal{B}_2^{(q)} &= \sum_{k=1,j=1}^{\infty} |k|^q |j|^q \Gamma_k^{xy} \vartheta_{k,j} \Gamma_{-j}^{yx} \\ &\quad + \sum_{j_1,\ldots,j_4=1}^{\infty} \Gamma_{j_1}^{xy} \vartheta_{j_1,j_2} |j_2| \omega_{j_2,j_3} |j_3|^q \vartheta_{j_3,j_4} \Gamma_{-j_4}^{yx} + \mathcal{B}_1^{(2q)}. \end{aligned} \tag{3.6}$$

The Mean Squared error displays a trade off between higher efficiency due to more included instruments 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 uncorrelated with $v_{t,j}$.

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 lagged instruments leads to significant changes in the parameter estimates. These changes are in fact due to the presence of the Bias term $Mn^{-1/2} D^{-1} \mathcal{A} \int k^2(x)dx$.

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 $s > q$ and $k(x) = \{|x| \leq 1\}$. If $n, M \to \infty$ and $M^{2q+2}/n \to \kappa$ with $0 < \kappa < \infty$ 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 $n^{-1}M^2$. 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 $\hat{\beta}_{n,M}$. From well known arguments we deduce that the optimal lag length choice, $\bar{M}^*$, is given by

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

Using $\bar{M}^*$ 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.

For technical reasons one needs to guarantee that $\bar{M}^*$ does not coincide with $\left[\bar{M}^*\right]$ or in other words does not fall on points of discontinuity of $[.]$. We therefore specify M^* for any $1 \gg \varepsilon_M > 0$ as $M^* = \left[\bar{M}^*\right] + \max(\bar{M}^* - \left[\bar{M}^*\right], \varepsilon_M)$ or equivalently as

$$M^* = \begin{cases} \bar{M}^* & \text{if } \bar{M}^* - \left[\bar{M}^*\right] \geq \varepsilon_M \\ \left[\bar{M}^*\right] + \varepsilon_M & \text{if } \bar{M}^* - \left[\bar{M}^*\right] < \varepsilon_M \end{cases}$$

This definition guarantees that $M^* - [M^*] \geq \varepsilon_M$. Since ε_M can be chosen arbitrarily small the definition of M^* does not affect the optimal MSE.

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 $\tilde{\beta}_{n,M}$ with $M \geq d/p$ 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 $[M]\,p$ dimensional vector of lagged observations where $[M]$ indicates the highest lag. As long as M is fixed and finite

$\partial\bar{g}(\theta)/\partial\theta \xrightarrow{p} P_M$. Classical results show that $\tilde{\beta}_{n,M}$ is consistent and asymptotically normal with $\sqrt{n}(\tilde{\beta}_{n,M} - \beta) \xrightarrow{d} N(0, (P_M' P_M)^{-1} P_M' \Omega_M^{-1} P_M (P_M' P_M)^{-1})$. Typically, one chooses a small number of instruments for the first stage estimate. 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 $\hat{\Omega}_M^{-1}$ and the constants $\mathcal{A}_1, \mathcal{B}_1^{(q)}$ and $\mathcal{B}_2^{(q)}$. Estimation of $\hat{\Omega}_M^{-1}$ 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 $\mathcal{A}_1$ can be done nuisance parameter free in the sense that consistent estimates of $\mathcal{A}_1$ do not depend on additional unknown parameters. Unfortunately the same is not true for $\mathcal{B}_1^{(q)}$ and $\mathcal{B}_2^{(q)}$ 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 first consider the simpler estimation problem for the constant $\mathcal{A}_1$. For this purpose note that $f_\Omega(\lambda) = 2\pi f_\varepsilon(\lambda) f_y(\lambda)$ such that

$$f_{yy}(\lambda) f_\Omega^{-1}(\lambda) = (2\pi)^{-1} f_\varepsilon^{-1}(\lambda).$$

While $f_\varepsilon^{-1}(\lambda)$ could be estimated nonparametrically from the autocovariances of the estimated innovations $\hat{\varepsilon}_t$ 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.

To express the constant $\mathcal{A}$ we use the same definitions as before. From

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

it follows that $(\operatorname{vec} f_\Omega^{-1}(\lambda))' [\operatorname{vec} f_y(\lambda)' \otimes f_{\varepsilon x}(\lambda)'] = (2\pi)^{-1} p f_\varepsilon^{-1}(\lambda) f_{\varepsilon x}(\lambda)'$. The spectral density $f_{\varepsilon y}(\lambda)$ 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 $(2\pi)^{-1} \int f_{\varepsilon x}(\lambda)' f_\varepsilon^{-1}(\lambda) d\lambda$ which is

$$(2\pi)^{-2} \sum_{k=-\infty}^{\infty} \zeta_k \Gamma_k^{\varepsilon x'} = (2\pi)^{-2} \sum_{k=-\infty}^{\infty} \left[\zeta_k \Gamma_{k-m}^{\varepsilon y} E', \zeta_k \Gamma_{k+1-m}^{\varepsilon y}, ..., \zeta_k \Gamma_{k+r-m}^{\varepsilon y}\right]$$

where E is a $(p-1) \times p$ matrix defined as $E = (0, I_{p-1})$. Denoting the consistent MA(q) parameters by $\tilde{\theta}_j$ the coefficients ζ_j can be obtained from $\hat{\zeta}_k = (2\pi)^2 \sigma^{-2} \sum_{j=0}^{\infty} e_1' B^j e_1 e_1' B^{j+k} e_1$ where e_1 is the first unit vector in $\mathbb{R}^m$ and

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

Consistent estimates of the MA(q) representation of ε_t can be obtained by using consistent estimates of the parameter β to obtain estimated $\hat{\varepsilon}_t$. An MA(q) model is then estimated for $\hat{\varepsilon}_t$. This can be done by using a nonlinear least squares or pseudo maximum likelihood procedure as described in chapter 8 of Brockwell and Davis (1991). This procedure is outlined in the proof of Lemma (4.1). The matrices $\Gamma_k^{\varepsilon y}$ can be replaced by simple sample averages based on estimated residuals

$$\hat{\Gamma}_k^{\varepsilon y} = n^{-1} \sum_{t=\min(k+1,1)}^{\min(n-m,n-k)} \hat{\varepsilon}_{t+m} y_{t-k}.$$

Using these estimates one then estimates $\widehat{\mathcal{A}}$ by

$$\widehat{\mathcal{A}}_1 = \frac{p}{4\pi^2} \left[\sum_{k=-n+1}^{n} \hat{\zeta}_k \hat{\Gamma}_{k-m}^{\varepsilon y} E', ..., \sum_{k=-n+1}^{n} \hat{\zeta}_k \hat{\Gamma}_{k+r-m}^{\varepsilon y} \right] \tag{4.2}$$

and

$$\widehat{\mathcal{A}} = \widehat{\mathcal{A}}_1 \hat{D}^{-1/2} \ell \ell' \hat{D}^{-1/2} \widehat{\mathcal{A}}_1' \tag{4.3}$$

The intuition why quantities of the form $\sum_k \hat{\zeta}_k \hat{\Gamma}_{k-m}^{\varepsilon y}$ 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 parameters $\mathcal{B}_1^{(q)}, \mathcal{B}_2^{(q)}$ and D are harder to estimate. 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 Vector Autoregessive (VAR) model of order τ as approximating process for $C(L)$ such that

$$y_t = A_1 y_{t-1} + ... + A_\kappa y_{t-\tau} + v_t. \tag{4.4}$$

The choice of τ is guided mainly by practical considerations. If the number of variables p in the system is large then τ should be chosen small, i.e. close to one. Alternatively, in the simulations reported in Section (6) consistent model selection criteria are used to select an optimal τ.

In order to calculate the impulse response coefficients associated with (4.4) 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 $\tau p \times \tau p$ and $\tau p \times p$. 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 $\left\| E_1'(\sum_j^{T_\epsilon} A^j - (I-A)^{-1})E_1 \right\| < \epsilon$. The autocovariance function Γ_j^{yy} is then approximated by $\tilde{\Gamma}_j^{yy} = \sum_{l=0}^{T_\epsilon} \tilde{C}_l \tilde{\Sigma} \tilde{C}_{l+j}$ where $\tilde{\Sigma}$ is the covariance matrix of the residuals in the approximating model. Likewise we obtain approximations to the optimal weight matrix $\tilde{\Omega} = \sum_{j=-m+1}^{m} \gamma^\varepsilon(j) \tilde{\Omega}(j)$ where $\tilde{\Omega}(j)$ has typical k,l-th block $\tilde{\Gamma}^{yy}(l-k-j)$. In accordance with earlier definitions we denote the k,l-th block of $\tilde{\Omega}$ by $\tilde{\omega}_{k,l}$ and the k,l-th block of $\tilde{\Omega}^{-1}$ by $\tilde{\vartheta}_{k,j}$. We define $\tilde{\mathcal{B}}_1^{(q)}, \tilde{\mathcal{B}}_2^{(q)}$ by and $\tilde{D}$ substituting $\tilde{\Gamma}_k^{xy}$ for Γ_k^{xy} and $\tilde{\vartheta}_{k,j}$ for $\vartheta_{k,j}$ in definitions (3.5) and (3.6) and by letting $\tilde{D} = \sum_{k=1,j=1}^{\infty} \tilde{\Gamma}_k^{xy} \tilde{\vartheta}_{k,j} \tilde{\Gamma}_{-j}^{yx}$.

Substituting estimates $\widehat{\tilde{C}}_j$ for $\tilde{C}_j$ in $\tilde{\mathcal{B}}_1^{(q)}, \tilde{\mathcal{B}}_2^{(q)}$ and $\tilde{D}$ leads to an estimates $\widehat{\tilde{\mathcal{B}}}_1^{(q)}, \widehat{\tilde{\mathcal{B}}}_2^{(q)}$ and $\widehat{\tilde{D}}$. While $\widehat{\mathcal{A}}$ is positive by construction the same is not true for $\widehat{\tilde{\mathcal{B}}}^{(q)}$ when based directly on its definition in Proposition (3.3). In practice we therefore use an alternative version of $\widehat{\tilde{\mathcal{B}}}^{(q)}$. We use the approximate autocovariance matrices $\tilde{\Gamma}_j^{yy}$ to form the matrices $\tilde{P}_{T_1}' = \left[\tilde{\Gamma}_1^{xy}, ..., \tilde{\Gamma}_{T_1}^{xy}\right]$ and $\tilde{\Omega}_{T_1} = \sum_{j=-m+1}^{m} \gamma^\varepsilon(j) \tilde{\Omega}_{T_1}(j)$ where $\tilde{\Omega}_{T_1}(j)$ has typical k,l-th block $\tilde{\Gamma}^{yy}(l-k-j)$. We then form the matrix $J = [\operatorname{diag}(1,2,...,T_1) \otimes I_2]$ and compute the matrix $\tilde{b}_{T_1}^{(q)} = \tilde{P}_{T_1}' J^q \tilde{\Omega}_{T_1}^{-1} + \tilde{P}_{T_1}' \tilde{\Omega}_{T_1}^{-1} J^q - \mathcal{B}_1^{(q)} D^{-1} \tilde{P}_{T_1} \tilde{\Omega}_{T_1}^{-1}$. The parameter $\tilde{\mathcal{B}}_{T_1}^{(q)}$ is obtained from

$$\tilde{\mathcal{B}}_{T_1}^{(q)} = \tilde{b}_{T_1}^{(q)} \tilde{\Omega}_{T_1} \tilde{b}_{T_1}^{(q)\prime}. \tag{4.5}$$

It can be shown that as $T_1 \to \infty$ the approximate $\tilde{\mathcal{B}}_{T_1}^{(q)}$ tends to $\tilde{\mathcal{B}}^{(q)}$. An estimate $\widehat{\tilde{\mathcal{B}}}_{T_1}^{(q)}$ of $\tilde{\mathcal{B}}_{T_1}^{(q)}$ based on estimated coefficients $\widehat{\tilde{C}}_j$ then is $\sqrt{n}$-consistent for $\tilde{\mathcal{B}}_{T_1}^{(q)}$ as long as the coefficient estimates are $\sqrt{n}$-consistent for $\tilde{C}_j$ and can be made arbitrarily close to $\widehat{\tilde{\mathcal{B}}}^{(q)}$ by choosing T_ϵ and T_1 large enough.

In the same way we define $\tilde{D}_{T_1} = \tilde{P}'_{T_1}\tilde{\Omega}^{-1}_{T_1}\tilde{P}_{T_1}$ with a corresponding estimate $\widehat{\tilde{D}}$ based on estimated coefficients $\widehat{\tilde{C}}_j$.

We assume that $\widehat{\tilde{\mathcal{B}}}^{(q)}_{T_1}$ is estimated such that it is $\sqrt{n}$-consistent for $\tilde{\mathcal{B}}^{(q)}_{T_1}$.

Assumption G. *For all $T_1 \geq 1$ fixed, $\sqrt{n}(\widehat{\tilde{\mathcal{B}}}^{(q)}_{T_1} - \tilde{\mathcal{B}}^{(q)}_{T_1}) = O_p(1)$.*

It is then established in the following lemma that the estimates for $\mathcal{B}^{(q)}/\mathcal{A}$ formed by $\widehat{\tilde{\mathcal{B}}}^{(q)}/\widehat{\mathcal{A}}$ where $\widehat{\tilde{\mathcal{B}}}^{(q)}$ is the estimated version of (4.5) are well enough behaved to be used in a plug in procedure.

Lemma 4.1. *Let $\widehat{\mathcal{A}}$ be defined in (4.3) and $\widehat{\tilde{\mathcal{B}}}^{(q)}_{T_1}$ be based on (4.5) and satisfy (G) for all fixed $T_1 \geq 1$. Then $\sqrt{n}(\widehat{\mathcal{A}} - \mathcal{A}) = O_p(1)$ and $\sqrt{n}(\widehat{\tilde{\mathcal{B}}}^{(q)}_{T_1}/\widehat{\mathcal{A}} - \tilde{\mathcal{B}}^{(q)}_{T_1}/\mathcal{A}) = O_p(1)$ where $\tilde{\mathcal{B}}^{(q)}_{T_1}$ is defined in (4.5).*

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 $|k(x) - k(y)| \leq C\,|x - y|\ \forall x, y \in \mathbb{R}$ 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 one of the main results 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. Following Andrews (1991) we define the truncated mean squared error as

$$\varphi_{n,h}(M, \ell, k(.), b_{n,M}) = E \min\left\{n^2/M^2 \ell' D^{1/2}(b_{n,M} - \beta)(b_{n,M} - \beta)' D^{1/2\prime}\ell, h\right\} - 1$$

and state the following theorem.

Theorem 4.2. *Suppose Assumptions (A), (B) and (C(s)) hold with $s > q$ and $k(.)$ satisfies Assumptions (E), (F) and (H). If $\widehat{\tilde{\mathcal{B}}}^{(q)}_{T_1}$ satisfies (G) and $\hat{M}^* = \left(nqk_q\widehat{\tilde{\mathcal{B}}}^{(q)}_{T_1}/\widehat{\mathcal{A}}\left(\int k(x)^2dx\right)^2\right)^{\frac{1}{2+2q}}$ then $n/\sqrt{M}(\hat{\beta}_{n,\hat{M}^*} - \hat{\beta}_{n,M^*}) = o_p(1)$ and*

$$\lim_{h\to\infty}\lim_{n\to\infty}\left(\varphi_{n,h}(\tilde{M}^*, \ell, k(.), b_{n,\hat{M}^*}) - \varphi_{n,h}(\tilde{M}^*, \ell, k(.), b_{n,\tilde{M}^*})\right) = 0$$

where $\tilde{M}^* = \left(nqk_q\tilde{\mathcal{B}}_{T_1}^{(q)}/\mathcal{A}\left(\int k(x)^2dx\right)^2\right)^{\frac{1}{2+2q}}$ *for* $T_1 \geq 1$ *fixed.*

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. An immediate consequence of the Theorem is also that $\hat{\beta}_{n,\hat{M}^*}$ is first order asymptotically equivalent to the infeasible estimator $D^{-1}d_0$.

5. Bias Reduction and Bias Correction

In this section we analyze the asymptotic bias of $\hat{\beta}_{n,M}$ 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* $s > q$ *and* $k(.)$ *satisfies Assumptions (E). If* $M \to \infty$ *and* $M/n^{1/2} \to 0$ *then*

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

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* $n, M \to \infty$, $M/n^{1/2} \to 0$ *and* $\int k^2(x)dx \leq 2$ *then*

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

where $\beta_{n,M}^T$ *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 $k_P(x) = (1 - 6x^2 + 6|x|^3)\{|x| \leq 1/2\} + 2(1 - |x|^3)\{1/2 \leq |x| \leq 1\}$ and the Tukey-Hanning $k_T(x) = (1 + \cos(\pi x))/2\{|x| \leq 1\}$.

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 and 3/4 for the Tukey-Hanning. 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

$$\hat{\beta}^*_{n,M} = \hat{\beta}_{n,M} - \left(\hat{P}'_M K_M \hat{\Omega}_M^{-1} K_M \hat{P}_M\right)^{-1} \frac{M}{n} \widehat{\mathcal{A}}'_1 \int k^2(x)dx. \tag{5.1}$$

The bias term $\mathcal{A}_1$ can be estimated by the methods described in the previous section. The quality of the estimator $\mathcal{A}_1$ 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 $\hat{\beta}_{n,M}$. If $\widehat{\mathcal{A}}_1 - \mathcal{A}_1 = O_p(n^{-\eta})$ for $\eta \in (0, 1/2]$ then the convergence rate of the estimator is improved. The mean squared error of the bias corrected estimator is defined as

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

and we obtain the following result.

Theorem 5.3. *Suppose Assumptions (A), (B) and (C(s)) hold with $s > q$ 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.

An alternative to direct recentering of the estimator by subtracting an asymptotically correct estimator of the bias is to compensate $\hat{\beta}_{n,M}$ in a way that eliminates higher order bias terms. The classical case of such a procedure is Nagar's (1959) k-class estimator.

In our context a k-class estimator can be defined as follows. Let $\Bbbk = M/n \int k^2(x)dx$ and define the n-dimensional matrix $A_n(\Phi)$ with typical element k, j given by $[A_n(\Phi)]_{k,j} = \zeta_{k-j}$. The k-class estimator is then

$$\beta^{\Bbbk}_{n,M} = \left[\hat{P}'_M K_M \hat{\Omega}_M^{-1} K_M \hat{P}' - n^{-1}\Bbbk X' A_n(\hat{\Phi}) X\right]^{-1} n^{-1} \left(\hat{P}'_M K_M \hat{\Omega}_M^{-1} K_M Z'_n - \Bbbk X' A_n(\hat{\Phi})\right) Y$$

This formulation takes the serial correlation in ε_t into account. It can be readily seen that $\beta^{\Bbbk}_{n,M}$ is equivalent to the kernel weighted GMM estimator for $\Bbbk = 0$. An analogy to the formulation in Nagar (1959) and $\beta^{\Bbbk}_{n,M}$ can be drawn by defining $Q_n = Z_n K_M \hat{\Omega}_n^{-1} K_M Z_n'$, $T_n = A_n(\hat{\Phi}) - Q_n$ and $W = n^{-1} X' T_n$. Then the k-class estimator can be written as

$$\beta^{\mathbb{N}}_{n,M} = \left[X' A_n(\hat{\Phi}) X - (1+\Bbbk) n^{-1} X' T_n X\right]^{-1} \left(X' A_n(\hat{\Phi}) - (1+\Bbbk) W\right) Y.$$

The two versions $\beta^{\Bbbk}_{n,M}$ and $\beta^{\mathbb{N}}_{n,M}$ differ by terms that are of order $O_p(M/n)$.

The next theorem establishes that the k-class estimator achieves the same rate of convergence for the higher order terms as the bias corrected estimator.

Theorem 5.4. *Suppose Assumptions (A), (B) and (C(s)) hold with $s > q$ and $k(.)$ satisfies Assumptions (E) and (F). Let*

$$\varphi^a_n(M, \ell, k(.)) = n D^{1/2} \ell' E(b^a_{n,M} - \beta)(b^a_{n,M} - \beta)' \ell D^{1/2} - 1$$

for $a \in \{\Bbbk, \mathbb{N}\}$ where $b^a_{n,M} = \beta^a_{n,M} - R^a_{n,M}$ and $R^a_{n,M} = o_p(M/n)$. 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 $M = O(n^{1/(1+2q)})$.

6. Monte Carlo Simulations

A small Monte Carlo experiment is conducted in order to assess the relevance of the asymptotic approximations derived in the previous sections. We are using the following kernel functions in addition to a standard GMM estimator with a finite number of M instruments. The kernels used are the Bartlett $k_B(x)$, the Parzen $k_P(x)$ and the Tukey-Hanning $k_T(x)$ which were all defined in Section 5.For the simulations we consider the following data generating process

$$\begin{aligned} y_{1t} &= \beta y_{2t} + u_t - \theta u_{t-1} \\ y_{2t} &= \phi y_{2t-1} + v_t. \end{aligned} \tag{6.1}$$

with $\beta_0 = 1$ and $[u_t, v_t]' \sim N(0, \Sigma)$ where Σ has elements $\sigma_1^2 = \sigma_2^2 = 1$ and σ_{12}. The parameter σ_{12} is one of the determinants of the small sample bias of both Ordinary Least Squares (OLS) and GMM estimators and is set to .5. The parameter ϕ controls the quality of lagged instruments and is chosen in $\{.1, .2, .3, .6\}$. The parameter θ finally is set to $\{-.9, -.6, -.3, 0, .3, .6, .9\}$.

The optimal M^* can be computed using the constants published in Andrews (1991). We have $k_1 = 1$ for the Bartlett, $k_2 = 6$ for the Parzen and $k_2 = \pi^2/4$ for the Tukey-Hanning kernel. Moreover, $\int k(x)^2dx$ is 2/3 for the Bartlett, .539285 for the Parzen and 3/4 for the Tukey-Hanning. This leads to

$$M^* = \begin{cases} 1.22474\left(n\mathcal{B}^{(q)}/\mathcal{A}\right)^{1/3} & \text{Bartlett} \\ 2.50582\left(n\mathcal{B}^{(q)}/\mathcal{A}\right)^{1/6} & \text{Parzen} \\ 1.66942\left(n\mathcal{B}^{(q)}/\mathcal{A}\right)^{1/6} & \text{Tukey-Hanning} \end{cases}.$$

We generate samples of size $n = 128$ from Model (6.1). Starting values are $y_0 = 0$ and $\varepsilon_0 = 0$. In each sample the first 1,000 observations are discarded to eliminate dependence on initial conditions.

Standard GMM estimators are obtained from applying Formula (3.3) with $K_M = I_M$. In order to obtain an estimate for Ω_M we first construct an inefficient but consistent estimate $\tilde{\beta}_{n,1}$ based on (3.3) setting $K_M = I_M$ and $\Omega_M = I_M$. We then construct residuals $\tilde{\varepsilon}_t = y_{1t} - \tilde{\beta}_{n,M}y_{2t}$ and estimate $\hat{\Omega}_M$ as described in (3.1). Kernel weighted GMM estimators (KGMM) are based on the same inefficient initial estimate such that the estimate for $\hat{\Omega}_M$ is identical to the weight matrix used for the standard GMM estimators. In the second stage we again apply (3.3) with $\hat{\Omega}_M$ and the appropriate matrix K_M corresponding to the respective kernel function.

The estimated optimal bandwidth $\hat{M}^*$ is computed according to the procedure laid out in Section (4). For each simulation replication we obtain a consistent first stage estimate $\tilde{\beta}_{n,1}$ to generate residuals $\tilde{\varepsilon}_t$. We estimate θ by fitting an MA(1) model to $\tilde{\varepsilon}_t$ using the GAUSS procedure `arima.src`. We then estimate the sample autocovariances $\Gamma_j^{\varepsilon y}$ for $j = 0, ..., n/2$ where n is the sample size and form an estimate of $\mathcal{A}_1$ based on Formula (4.2). Next we use the BIC (see Reinsel, 1995, p. 92) criterion to determine the optimal specification of the approximating VAR for $y_t = [y_{1t}, y_{2t}]'$ allowing for a maximum of 10 lags. Based on the optimal lag length specification we compute the impulse coefficients of the VAR and estimate $\tilde{\mathcal{B}}_{T_1}^{(q)}$ and $\widehat{\tilde{D}}_{T_1}$ for $T_\epsilon = T_1 = 100$. Experiments with larger values for T_ϵ and T_1 indicate that the results are not sensitive to the choice of these parameters.

In Tables 1-4 we compare the performance of feasible kernel weighted GMM with Bartlett (k_B), Parzen (k_P) and Tukey-Hanning (k_T) kernels to an infeasible optimal GMM estimator. The infeasible optimal GMM estimator is obtained by estimating β by standard GMM for M fixed at $M = 1, 2, 3, 4, 5, 10, 15, 20$. We then calculate the empirical mean squared error of parameter estimates based on 1,000 Monte Carlo replications and chose the specification that leads to the lowest MSE statistic. A table entry "IV 10" for example means that GMM with instruments up to lag 10 achieved the lowest MSE in the simulations. We compare this estimator to feasible kernel weighted GMM based on $\hat{M}^*$. Keeping in mind that the infeasible GMM estimators performance is unattainable in practice the results for the feasible procedures are very encouraging. For cases with weak identification, ie. when ϕ is close to zero, the KGMM based on the Bartlett kernel actually outperforms the infeasible procedure both in terms of median bias and mean squared error. For values of $\phi = .3$ and $.6$ the performance is still quite good although not quite as good as the infeasible procedure. Kernel based estimators tend to have problems with the existence of second moments, particularly with the Parzen and Tukey-Hanning kernels. This explains the sometimes inflated MSE values compared to the MAE statistics.

In Tables 5-8 we analyze the optimality of $\hat{M}^*$. For this purpose we compare the Mean Absolute Error (MAE) of GMM and KGMM based on a fixed number of instruments M to the feasible procedure. The tables reveal that choosing a suboptimal number of instruments can have large effects on the MAE, with increases of 50% and more relative to the best possible performance in some cases. The tables also show that, keeping M fixed, using a kernel weighting scheme for the moment conditions can reduce the MAE. This is particularly true when ϕ is small and generally when M is large. The latter effect is related to the bias reduction property of kernel weighting which becomes important for large values of M. Most importantly, the feasible procedure successfully minimizes the MAE for values of ϕ close to zero and performs reasonably well for values of $\phi > .3$. This is partly due to the fact that KGMM is less sensitive to severe overidentification than standard GMM. Kernel weighting is therefore a very useful tool for developing feasible procedures with reasonable finite sample properties.

In Tables 9 and 10 we focus on the median bias properties of GMM, KGMM with Bartlett kernel as well as bias corrected KGMM. The bias reduction property of the kernel weights established in Corollary (5.2) is extremely robust across the entire parameter space. The magnitude of the bias reduction relative to standard GMM can reach up to 50% of the original bias when $\phi = .3$. Experiments with implementations of the bias corrected estimator (5.1) indicate fairly good performance as far as bias reduction is concerned but have lead to severely inflated MAE and MSE statistics. For this reason we report results for an alternative version of (5.1) defined as

$$\hat{\beta}^{**}_{n,M} = \hat{\beta}_{n,M} - \widehat{\tilde{D}}_{T_1}^{-1} \frac{M}{n} \widehat{\mathcal{A}}_1' \int k^2(x)dx$$

where $\widehat{\tilde{D}}_{T_1}$ is computed based on an approximating VAR as described before. The performance of this bias corrected estimator is mixed. When the instruments are weak, ie. if ϕ is small then the effect of the bias correction on the bias is small, especially for $|\theta|$ large. In Table 9 the best result is achieved for $\theta = 0$ and $M = 1$ where the bias is essentially eliminated. Generally speaking the bias correction is quite sensitive to the choice of M and does not perform well for large values of M. When ϕ is set to .3 the bias correction works better giving an additional bias reduction of up to 50% over the kernel weighted procedure. Again, the performance of $\hat{\beta}^{**}_{n,M}$ deteriorates for $M > 5$. As with any bias corrected procedure one worries about the impact of the bias correction on the variability of the estimator. Tables 11 and 12 compare the MAE of $\hat{\beta}^{**}_{n,M}$ to GMM and KGMM. For values of $M \leq 5$ the MAE of $\hat{\beta}^{**}_{n,M}$ is generally in line with the two other procedures. In some cases ($\phi = .3$, $|\theta| = .3$, $M = 4$) it even outperforms the other procedures in terms of MAE. On the other hand for $M > 5$ the MAE starts to increase quite significantly such that a combination of bias correction and severe overidentification can not be recommended based on the simulation results.

7. 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.

Massachusetts Institute of Technology, Dept. of Economics, 50 Memorial Drive E52-251B, Cambridge, MA 02142, USA. Email: gkuerste@mit.edu. Web: http://web.mit.edu/gkuerste/www/.

A. Proofs

Proof of Lemma (3.1). Recall $\hat{D}_M = n^{-2}X'Z_MK_M\hat{\Omega}_M^{-1}K_MZ_M'X$. First we will split the error $\hat{D}_M - D$ into four different parts as

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

where $H_1 = P_M'K_M\Omega_M^{-1}K_MP_M - P'\Omega^{-1}P$, $H_2 = \hat{P}_M'K_M\Omega_M^{-1}K_M\hat{P}_M - P_M'K_M\Omega_M^{-1}K_MP_M$, $H_3 = -\hat{P}_M'K_M\Omega_M^{-1}(\hat{\Omega}_M - \Omega_M)\Omega_M^{-1}K_M\hat{P}_M$ and H_4 is defined in (A.14). The terms H_3 and H_4 contain a Taylor series expansion of $\hat{\Omega}_M^{-1}$ around Ω_M^{-1} given by

$$\hat{\Omega}_M^{-1} = \Omega_M^{-1} - \Omega_M^{-1}(\hat{\Omega}_M - \Omega_M)\Omega_M^{-1} + B + o_p(\left\|\hat{\Omega}_M - \Omega_M\right\|^2) \tag{A.1}$$

where B has typical element k, l given by $\mathrm{vec}(\hat{\Omega}_M - \Omega_M)'\frac{\partial^2\vartheta_{kl}}{\partial\,\mathrm{vec}\,\Omega\partial\,\mathrm{vec}\,\Omega'}\mathrm{vec}(\hat{\Omega}_M - \Omega_M)$. In Lemmas (B.9) to (B.11) it is shown that $H_1 = H_{11} + H_{12} + H_{13} + H_{14}$ is

$$H_{11} \equiv P_M'\Omega_M^{-1}P_M - P'\Omega^{-1}P = o(M^{-2s}) \tag{A.2}$$

$$H_{12} \equiv P_M'(I - K_M)\Omega_M^{-1}(I - K_M)P_M = O(M^{-2q}) \tag{A.3}$$

$$H_{13} \equiv -P_M'\Omega_M^{-1}(I - K_M)P_M = O(M^{-q}) \tag{A.4}$$

$$H_{14} \equiv -P_M'(I - K_M)\Omega_M^{-1}P_M = O(M^{-q}) \tag{A.5}$$

where $\equiv$ means 'equal by definition'. In Lemmas (B.12) to (B.15) the term $H_2 = H_{211} + H_{212} + H_{221} + H_{222}$ is analyzed to be

$$H_{211} \equiv -\left(\hat{P}_M - \check{P}_M\right)'K_M\Omega_M^{-1}K_M(\hat{P}_M - \check{P}_M) = O_p(M/n) \tag{A.6}$$

$$H_{212} \equiv \hat{P}_M'K_M\Omega_M^{-1}K_M(\hat{P}_M - \check{P}_M) + (\hat{P}_M - \check{P}_M)'K_M\Omega_M^{-1}K_M\hat{P}_M = O_p(n^{-1/2}) \tag{A.7}$$

$$H_{221} \equiv -\left(\check{P}_M - P_M\right)'K_M\Omega_M^{-1}K_M'\left(\check{P}_M - P_M\right) = O_p(M/n) \tag{A.8}$$

$$H_{222} \equiv \check{P}_M'K_M\Omega_M^{-1}K_M(\check{P}_M - P_M) + (\check{P}_M - P_M)'K_M\Omega_M^{-1}K_M\check{P}_M = O_p(M/n^{1/2}). \tag{A.9}$$

where $\hat{P}_M$ is defined in (3.2) and $\check{P}_M' = \left[\check{\Gamma}_1^{xy}, ..., \check{\Gamma}_{[M]}^{xy}\right]$ where $\check{\Gamma}_j^{xy} = n^{-1}\sum_{t=\max(j+1,r-m)+1}^{n} w_{t,j}$.

Lemmas (B.16) and (B.17) show that $H_3 = H_{31} + H_{32} + H_{33} + H_{34}$ is

$$H_{31} \equiv (\hat{P}_M - P_M)'K_M\Omega_M^{-1}(\hat{\Omega}_M - \Omega_M)\Omega_M^{-1}K_M(\hat{P}_M - P_M) = o_p(M/n) \tag{A.10}$$

$$H_{32} \equiv -\hat{P}_M'K_M\Omega_M^{-1}(\hat{\Omega}_M - \Omega_M)\Omega_M^{-1}K_M(\hat{P}_M - P_M) = o_p(M/n) \tag{A.11}$$

$$H_{33} \equiv -(\hat{P}_M - P_M)'K_M\Omega_M^{-1}(\hat{\Omega}_M - \Omega_M)\Omega_M^{-1}K_M\hat{P}_M = o_p(M/n) \tag{A.12}$$

$$H_{34} \equiv -P_M'K_M\Omega_M^{-1}(\hat{\Omega}_M - \Omega_M)\Omega_M^{-1}K_MP_M = O_p(n^{-1/2}) \tag{A.13}$$

and H_4 which is a remainder term defined as

$$H_4 \equiv \hat{P}_M' K_M(\hat{\Omega}_M^{-1} - \Omega_M^{-1} + \Omega_M^{-1}(\hat{\Omega}_M - \Omega_M)\Omega_M^{-1})K_M \hat{P}_M = o_p(M/n) \tag{A.14}$$

where the last equality follows from Lemma (B.18).

Next we turn to the analysis of $\hat{d}_M = \hat{P}_M' K_M \hat{\Omega}_M^{-1} n^{-1/2} \sum_{t=1}^{n-m} \varepsilon_{t+m} \bar{z}_{t,n} K_M$ which is decomposed as $\hat{d}_k = \sum_j^9 d_j$. Define $V_M = \left[n^{-1/2} \sum_t v_{t,1}', ..., n^{-1/2} \sum_t v_{t,[M]}'\right]'$ with $V \equiv V_\infty$ such that it follows from Lemmas (B.19) to (B.28) that

$$
\begin{aligned}
&\text{(A.15)} \qquad & d_0 &\equiv P'\Omega^{-1}V = O_p(1) \\
&\text{(A.16)} \qquad & d_1 &\equiv P_M'\Omega_M^{-1}V_M - P'\Omega^{-1}V = o_p(M^{-s}) \\
&\text{(A.17)} \qquad & d_2 &\equiv P_M'(I - K_M)\Omega_M^{-1}(I - K_M)V_M = O_p(M^{-2q}) \\
&\text{(A.18)} \qquad & d_3 &\equiv -P_M'(I - K_M)\Omega_M^{-1}V_M - P_M'\Omega_M^{-1}(I - K_M)V_M = O_p(M^{-q}) \\
&\text{(A.19)} \qquad & d_4 &\equiv \left(\hat{P}_M - \check{P}_M\right)' K_M \Omega_M^{-1} K_M V_M = O_p(M/n) \\
&\text{(A.20)} \qquad & d_5 &\equiv \left(\check{P}_M - P_M\right)' K_M \Omega_M^{-1} K_M V_M = O_p(M/n^{1/2}) \\
&\text{(A.21)} \qquad & d_6 &\equiv \left(\hat{P}_M - P_M\right)' K_M \Omega_M^{-1}(\Omega_M - \hat{\Omega}_M)\Omega_M^{-1} K_M V_M = O_p(M/n) \\
&\text{(A.22)} \qquad & d_7 &\equiv P_M' K_M \Omega_M^{-1}(\Omega_M - \hat{\Omega}_M)\Omega_M^{-1} K_M V_M = O_p(n^{-1/2}) \\
&\text{(A.23)} \qquad & d_8 &\equiv \hat{P}_M' K_M B K_M V_M + o_p\left(M/n\right) = O_p(M/n) \\
&\text{(A.24)} \qquad & d_9 &\equiv n^{-1/2} \sum_t \varepsilon_t \hat{P}_M' K_M \hat{\Omega}_M^{-1} K_M \left[\mathbf{1}_{[M]} \otimes (\bar{y} - \mu_y)\right] = O_p(M/n^{3/2}).
\end{aligned}
$$

We consider the terms in the expansion $D^{-1}\sum_{i=0}^9 d_i - D^{-1}\sum_{i=1}^4 \sum_{j=0}^9 H_i D^{-1} d_j$ of the estimator which depend on M and n and are largest in probability. From the results in Equations (A.2) to (A.24) it follows that the largest such terms are $H_{12}, H_{13}, H_{14}, H_{222}, d_0, d_2, d_3$ and d_5. Of those terms we examine cross products of the form $Ed_i d_j'$, $Ed_i d_0' D^{-1} H_i$ and $EH_i D^{-1} d_0 d_0' D^{-1} H_j$. The largest terms vanishing at rate M^{-q} as $M \to \infty$ are $Ed_0 d_3' = -M^{-q} k_q \mathcal{B}_1^{(q)} + o(M^{-q})$ as shown in Lemma (B.31) and $-Ed_0 d_0' D^{-1}(H_{13} + H_{14}) = M^{-q} k_q \mathcal{B}_1^{(q)} + o(M^{-q})$ by Lemmas (B.19) and (B.34). The two terms cancel because they are of opposite sign.

Terms of order M^{-2q} include $Ed_0 d_2' = M^{-2q} k_q^2 \mathcal{B}_0^{(q)} + o(M^{-2q})$ by Lemma (B.30) and $-Ed_0 d_0' D^{-1} H_{12}' = -M^{-2q} k_q^2 \mathcal{B}_0^{(q)} + o(M^{-2q})$ by Lemma (B.29). Since $Ed_0 d_2'$ and $-Ed_0 d_0' D^{-1} H_{12}'$ are of opposite sign these terms cancel. We are left with $E(d_3 - (H_{13} + H_{14})D^{-1} d_0)(d_3 - (H_{13} + H_{14})D^{-1} d_0)' = O(M^{-2q})$ by Lemmas (B.11), (B.19), (B.31) and (B.35).

Terms that grow with M and are highest in order are $H_{222} D^{-1} d_0$ and d_5. It follows by Lemma (B.33) that the cross product term $EH_{222} D^{-1} d_0 d_5'$ is of lower order. We are left with $EH_{222}D^{-1}d_0d_0'D^{-1}H_{222}' = O(n^{-1})$ by Lemma (B.32) and $Ed_5d_5' = O(M^2/n)$ by Lemma (B.36). ■

Proof of Proposition (3.3) From the proof of Lemma (**3.1**) we only need to consider the terms $A_n = Ed_5d_5'$ and $B_n = E(d_3 - (H_{13} + H_{14})D^{-1}d_0)(d_3 - (H_{13} + H_{14})D^{-1}d_0)'$. Since for all $n \geq 1$ we have $A_n \geq 0$ and $B_n \geq 0$ 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 (B.36) it follows that

$$E\ell' D^{-1/2}d_5d_5'D^{-1/2}\ell = M^2/n\left(\int k^2(x)dx\right)^2 \mathcal{A}_1 D^{-1/2}\ell\ell' D^{-1/2}\mathcal{A}_1' + o(M^2/n).$$

From Lemma (B.31) it follows that $M^q Ed_0d_3 = -k_q\mathcal{B}_1^{(q)} + o(1)$ and from Lemma (B.19) it follows that $Ed_0d_0' = D + o(1)$ such that

$$M^{2q}E(H_{13} + H_{14})D^{-1}d_0d_0'D^{-1}(H_{13} + H_{14})' = k_q^2\mathcal{B}_1^{(q)}D^{-1}\mathcal{B}_1^{(q)\prime} + o(1).$$

This implies that

$$E(H_{13} + H_{14})D^{-1}d_0d_3' - E(H_{13} + H_{14})D^{-1}d_0d_0'D^{-1}(H_{13} + H_{14})' = o(M^{-2q})$$

or in other words $B_n = Ed_3d_3' - E(H_{13}+H_{14})D^{-1}d_0d_0'D^{-1}(H_{13}+H_{14}) + o(M^{-2q})$. Here $Ed_3d_3' = M^{-2q}\mathcal{B}_2^{(q)} + o(M^{-2q})$ as shown in Lemma (B.35) where $\mathcal{B}_2^{(q)}$ is defined in (3.6). ■

Proof of Lemma (4.1) The only difficulty here is to show that $\sum \hat{\zeta}_{j+m}\hat{\Gamma}_{j-k}^{\varepsilon y}$ 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 $g(\lambda, \theta) = \left|\theta(e^{i\lambda})\right|^2$ with $\theta(z) = 1 - \theta_1 z - ...\theta_{m-1}z^{m-1}$. Define the parameter space $\Theta_1 \subset \mathbb{R}^{m-1}$ such that $\theta = (\theta_1, ..., \theta_{m-1}) \in \Theta_1$ if $\theta(z) \neq 0$ for $|z| \leq 1$. By Assumption (D) $\exists\ \Theta_2 \subset \operatorname{int}\Theta_1, \Theta_2$ compact such that $\theta_0 \in \Theta_2$.

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

$$\tilde{\theta} = \arg\min_{\theta} \Lambda_n^{\hat{\varepsilon}}(\theta) \tag{A.25}$$

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

$\hat{\alpha}_0 - \alpha_0 = \bar{y} - \mu_y - \tilde{\beta}'(\bar{x} - \mu_x)$. It follows that

$$\begin{aligned}
\hat{I}_n^{\varepsilon}(\lambda) &= I_n^{\varepsilon}(\lambda) + (\tilde{\beta} - \beta)' I_n^{x}(\lambda)(\tilde{\beta} - \beta) + I_n^{\alpha}(\lambda) \\
&\quad + 2\left(\tilde{\beta} - \beta\right)' I_n^{\varepsilon x}(\lambda) + 2 I_n^{\varepsilon\alpha}(\lambda) + 2(\hat{\alpha}_0 - \alpha_0)(\tilde{\beta} - \beta)' I_n^{\alpha x}(\lambda).
\end{aligned}$$

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

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

From standard arguments (see Brockwell and Davis 1991, ch 10) it follows that $\Lambda_n^{ab}(\theta) \overset{a.s.}{\to} \Lambda^{ab}(\theta)$ with $\Lambda^{ab}(\theta) = 2\pi \int f_{ab}(\lambda)/g(\lambda, \theta) d\lambda$ and $\partial^k \Lambda_n^{ab}(\theta)/\partial\theta \overset{a.s.}{\to} \partial^k \Lambda^{ab}(\theta)/\partial\theta$ for $k < \infty$ such that $\Lambda_n^{\hat{\varepsilon}}(\theta) \to 2\pi \int f_{\varepsilon\varepsilon}(\lambda)/g(\lambda, \theta) d\lambda$ uniformly in $\theta \in \Theta_2$. Consistency of $\tilde{\theta}$ follows from standard arguments.

To establish $\sqrt{n}$-consistency note that $\sqrt{n}\partial\Lambda_n^{\varepsilon}(\theta_0)/\partial\theta = O_p(1)$, $n^{-1/2}\sum_t \varepsilon_t = O_p(1)$ and $n^{-1/2}\sum_t (x_t - \mu_x) = O_p(1)$. Therefore

$$\sqrt{n}\partial\Lambda_n^{\hat{\varepsilon}}(\theta)/\partial\theta = \sqrt{n}\partial\Lambda_n^{\varepsilon}(\theta)/\partial\theta + \sqrt{n}2(\tilde{\beta} - \beta)'\partial\Lambda^{\varepsilon x}(\theta)/\partial\theta + o_p(1). \tag{A.26}$$

We also define $\partial\Lambda^{\varepsilon}(\theta)/\partial\theta = 2\pi \int f_{\varepsilon}(\lambda) \partial g^{-1}(\lambda, \theta)/\partial\theta d\lambda$ such that

$$\left\| \frac{\partial\Lambda^{\varepsilon}(\tilde{\theta})}{\partial\theta} \right\| \leq \left\| \frac{\partial\Lambda^{\varepsilon}(\tilde{\theta})}{\partial\theta} - \frac{\partial\Lambda_n^{\hat{\varepsilon}}(\tilde{\theta})}{\partial\theta} \right\| + \left\| \frac{\partial\Lambda_n^{\hat{\varepsilon}}(\tilde{\theta})}{\partial\theta} - \frac{\partial\Lambda_n^{\hat{\varepsilon}}(\theta_0)}{\partial\theta} \right\| + \left\| \frac{\partial\Lambda_n^{\hat{\varepsilon}}(\theta_0)}{\partial\theta} \right\| \tag{A.27}$$

where $\left\|\partial\Lambda_n^{\hat{\varepsilon}}(\theta_0)/\partial\theta\right\| = O_p(n^{-1/2})$ by (A.26). Definition (A.25) for $\tilde{\theta}$ implies that

$$\left\| \frac{\partial\Lambda_n^{\hat{\varepsilon}}(\tilde{\theta})}{\partial\theta} - \frac{\partial\Lambda_n^{\hat{\varepsilon}}(\theta_0)}{\partial\theta} \right\| \leq 2 \left\| \frac{\partial\Lambda_n^{\hat{\varepsilon}}(\theta_0)}{\partial\theta} \right\| = O_p(n^{-1/2}).$$

Finally

$$\begin{aligned}
\frac{\partial\Lambda_n^{\hat{\varepsilon}}(\tilde{\theta})}{\partial\theta} - \frac{\partial\Lambda^{\varepsilon}(\tilde{\theta})}{\partial\theta} &= \partial\Lambda_n^{\varepsilon}(\tilde{\theta})/\partial\theta - \int 2\pi f_{\varepsilon}(\lambda) \partial g^{-1}(\lambda, \tilde{\theta})/\partial\theta d\lambda \\
&\quad + (\tilde{\beta} - \beta)'(\partial(\Lambda^{\varepsilon x}(\theta_0) + \Lambda^{\varepsilon x}(\theta_0)))/\partial\theta + o_p(1)
\end{aligned}$$

where the second term is $O_p(n^{-1/2})$ since $(\tilde{\beta}-\beta)=O_p(n^{-1/2})$. The first term can be written as

$$\begin{aligned}
\partial\Lambda_n^{\varepsilon}(\tilde{\theta})/\partial\theta - \int 2\pi f_\varepsilon(\lambda)\partial g^{-1}(\lambda,\tilde{\theta})/\partial\theta d\lambda \\
&= n^{-1}\sum_j \left[I_n^{\varepsilon}(\lambda_j)-2\pi f_\varepsilon(\lambda_j)\right]\partial g^{-1}(\lambda_j,\tilde{\theta})/\partial\theta \\
&\quad + n^{-1}\sum_j 2\pi f_\varepsilon(\lambda_j)\partial g^{-1}(\lambda_j,\tilde{\theta})/\partial\theta - \int 2\pi f_\varepsilon(\lambda)\partial g^{-1}(\lambda,\tilde{\theta})/\partial\theta d\lambda
\end{aligned}$$

where the second term is $O(n^{-1})$. Now define $\xi_j(\theta)=(2\pi)^{-1}\int \partial g^{-1}(\lambda,\theta)/\partial\theta e^{i\lambda j}d\lambda$ such that $\partial g^{-1}(\lambda,\theta)/\partial\theta=\sum_j \xi_j(\theta)e^{-i\lambda j}$ and

$$\begin{aligned}
& n^{-1}\sum_j \left[I_n^{\varepsilon}(\lambda_j)-2\pi f_\varepsilon(\lambda_j)\right]\partial g^{-1}(\lambda_j,\tilde{\theta})/\partial\theta \\
= \;& n^{-2}\sum_j\sum_{t,s=1}^{n}\sum_{l=-\infty}^{\infty}\left(\varepsilon_t\varepsilon_s - E\varepsilon_t\varepsilon_s\right)\xi_l(\tilde{\theta})e^{i\lambda_j(t-s-l)} \\
= \;& n^{-1}\sum_{l=-n}^{n}\sum_{t=\max(l,1)}^{n-|\min(l,0)|}\left(\varepsilon_t\varepsilon_{t-l}-E\varepsilon_t\varepsilon_{t-l}\right)\xi_l(\tilde{\theta}) \\
\leq \;& \left(\sum_{l=-n}^{n}|l|^{-2}\left(n^{-1}\sum\nolimits_t\left(\varepsilon_t\varepsilon_{t-l}-E\varepsilon_t\varepsilon_{t-l}\right)\right)^2\right)^{1/2}\left(\sum_{l=-n}^{n}|l|_l^2\,\xi_l(\tilde{\theta})^2\right)^{1/2} + n^{-1}\sum_t\left(\varepsilon_t^2-E\varepsilon_t^2\right)
\end{aligned}$$

where the second equality follows from $n^{-1}\sum_j e^{i\lambda_j(t-s)}=0$ for $t\neq s$ and the inequality follows from the Cauchy-Schwarz inequality. Then note that $n^{-1}\sum_t\left(\varepsilon_t^2-E\varepsilon_t^2\right)=O_p(n^{-1/2})$,

$$E\sum_{l=-n}^{n}|l|^{-2}\left(n^{-1}\sum\nolimits_t\left(\varepsilon_t\varepsilon_{t-l}-E\varepsilon_t\varepsilon_{t-l}\right)\right)^2=\sum_{l=-n}^{n}|l|^{-2}\,n^{-2}\sum\nolimits_t\left[E\varepsilon_t^2\varepsilon_{t-l}^2-\left(E\varepsilon_t\varepsilon_{t-l}\right)^2\right]=O(n^{-1}),$$

and $\sum_{l=-n}^{n}|l|^2\,c_l(\theta)^2$ is uniformly converging for θ with $|\theta-\theta_0|<\delta$ for some $\delta>0$ such that $\theta(z)$ has no zeros on or inside the unit circle. Consistency of $\tilde{\theta}$ then implies $\sum_{l=-n}^{n}|l|^2\,c_l(\tilde{\theta})^2=O_p(1)$. These results establish that $\left\|\frac{\partial\Lambda_n^{\hat{\varepsilon}}(\tilde{\theta})}{\partial\theta}-\frac{\partial\Lambda^{\varepsilon}(\tilde{\theta})}{\partial\theta}\right\|=O_p(n^{-1/2})$. From (A.27) it then follows that $\left\|\frac{\partial\Lambda^{\varepsilon}(\tilde{\theta})}{\partial\theta}\right\|=O_p(n^{-1/2})$ such that by a continuity argument $\sqrt{n}(\tilde{\theta}-\theta)=O_p(1)$.

We next show that $\sum_{j=-n+1}^{n-1}\tilde{\zeta}_{j+m}\hat{\Gamma}_{j-k}^{\varepsilon y}-\sum_{j=-n+1}^{n-1}\zeta_{j+m}\Gamma_{j-k}^{\varepsilon y}=O_p(n^{-1/2})$. Write

$$\sum_{j=-n+1}^{n-1}\tilde{\zeta}_{j+m}\hat{\Gamma}_{j-k}^{xy}-\sum_j\zeta_{j+m}\Gamma_{j-k}^{xy}=\sum_{j=-n+1}^{n-1}\tilde{\zeta}_{j+m}\left(\hat{\Gamma}_{j-k}^{xy}-\Gamma_{j-k}^{xy}\right)-\sum_{j=-n+1}^{n-1}\left(\tilde{\zeta}_{j+m}-\zeta_{j+m}\right)\Gamma_{j-k}^{xy}.$$

First consider

$$n^{1/2}\sum_{j=-n+1}^{n-1}\left\|\tilde{\zeta}_j-\zeta_j\right\|\left\|\Gamma^{xy}_{j-k}\right\|\leq n^{1/2}\sup_j\left\|\tilde{\zeta}_j-\zeta_j\right\|\sum_{j=-n+1}^{n-1}\left\|\Gamma^{xy}_{j-k}\right\|$$

where $P(\sup_j\left\|\tilde{\zeta}_j-\zeta_j\right\|>Cn^{-1/2})$ goes to zero for some C large enough by the previous result.

For any δ such that $|\theta-\theta_0|<\delta$ implies $\theta(z)$ has no zeros on or inside the unit circle consider

$$\begin{aligned}P\left(n^{1/2}\sum\nolimits_{j=-n+1}^{n-1}\left|\tilde{\zeta}_j\right|\left\|\hat{\Gamma}^{xy}_{j-k}-\Gamma^{xy}_{j-k}\right\|>\eta\right) &\leq P\left(n^{1/2}\sup_{|\theta-\theta_0|<\delta}\sum\nolimits_{j=-n+1}^{n-1}\left|\zeta_k\left(\theta\right)\right|\left\|\hat{\Gamma}^{xy}_{j-k}-\Gamma^{xy}_{j-k}\right\|>\eta\right)\\ &\quad +P\left(\left|\tilde{\theta}-\theta_0\right|\geq\delta\right)\end{aligned}$$

We use the triangular inequality $\left\|\hat{\Gamma}^{xy}_{j-k}-\Gamma^{xy}_{j-k}\right\|\leq\left\|\hat{\Gamma}^{xy}_{j-k}-\check{\Gamma}^{xy}_{j-k}\right\|+\left\|\check{\Gamma}^{xy}_{j-k}-\check{\Gamma}^{xy}_{j-k}\right\|$ such that

$$n^{1/2}\sup_{|\theta-\theta_0|<\delta}\sum\nolimits_{j=-n+1}^{n-1}\left|\zeta_j\left(\theta\right)\right|\left\|\hat{\Gamma}^{xy}_{j-k}-\check{\Gamma}^{xy}_{j-k}\right\|=O_p(1)$$

by Equation (B.2) and the fact that $\sup_{|\theta-\theta_0|<\delta}\sum_{j=-n+1}^{n-1}\left|\zeta_j\left(\theta\right)\right|=O(1)$. In the same way it follows from Equation (B.3) that

$$n^{1/2}\sup_{|\theta-\theta_0|<\delta}\sum\nolimits_{j=-n+1}^{n-1}\left|\zeta_j\left(\theta\right)\right|E\left\|\check{\Gamma}^{xy}_{j-k}-\Gamma^{xy}_{j-k}\right\|=O(1).$$

This establishes that $\widehat{\mathcal{A}}_1-\mathcal{A}_1=O_p(n^{-1/2})$. ■

Proof of Theorem (4.2) We first show that $n/\sqrt{M^*}(\hat{\beta}_{n,\hat{M}^*}-\hat{\beta}_{n,M^*})=o_p(1)$. The decomposition $\sqrt{n}\left(\hat{\beta}_{n,\hat{M}^*}-\hat{\beta}_{n,M^*}\right)=\hat{D}^{-1}_{M^*}\left(\hat{D}_{M^*}-\hat{D}_{\hat{M}^*}\right)\hat{D}^{-1}_{\hat{M}^*}\hat{d}_{\hat{M}^*}-\hat{D}^{-1}_{M^*}(\hat{d}_{\hat{M}^*}-\hat{d}_{M^*})$ is used. Note that $\hat{D}_{M^*}=O_p(1)$ and $d_{M^*}=O_p(1)$. The following calculations also establish $\hat{D}_{\hat{M}^*}=O_p(1)$ and $d_{\hat{M}^*}=O_p(1)$. It is therefore enough to show that $\sqrt{n/M^*}(\hat{D}_{M^*}-\hat{D}_{\hat{M}^*})=o_p(1)$ and $\sqrt{n/M^*}(\hat{d}_{\hat{M}^*}-\hat{d}_{M^*})=o_p(1)$. Define $\tilde{k}_M=\text{diag}(k(1/M),...,k((n-m)/M))'$ and $\tilde{K}_M=(\tilde{k}_M\otimes I_p)$ and let $\hat{P}'_{n-m}=n^{-1}X'Z_{n-m}$ and

$$\Omega^*_M=\begin{bmatrix}\Omega_M & 0\\ 0 & I_{n-[M]-m}\end{bmatrix},\Omega^{*-1}_M=\begin{bmatrix}\Omega^{-1}_M & 0\\ 0 & I_{n-[M]-m}\end{bmatrix}$$

with $\hat{\Omega}^*_M$ and $\hat{\Omega}^{*-1}_M$ defined in the same way replacing Ω_M and Ω^{-1}_M by $\hat{\Omega}_M$ and $\hat{\Omega}^{-1}_M$. Using these definitions we can rewrite $\hat{d}_M=n^{-1/2}\hat{P}'_{n-m}\tilde{K}_M\hat{\Omega}^{*-1}_M\tilde{K}_MZ'_{n-m}\varepsilon$. First consider

$$\begin{aligned}\sqrt{n/M^*}(\hat{d}_{\hat{M}^*}-\hat{d}_{M^*}) &= \sqrt{n/M^*}(\hat{P}'_{n-m}\tilde{K}_{\hat{M}^*}\hat{\Omega}^{*-1}_{\hat{M}^*}\tilde{K}_{\hat{M}^*}\frac{Z'_{n-m}\varepsilon}{\sqrt{n}}-\hat{P}'_{n-m}\tilde{K}_{M^*}\hat{\Omega}^{*-1}_{M^*}\tilde{K}_{M^*}\frac{Z'_{n-m}\varepsilon}{\sqrt{n}})\\ &= \sqrt{n/M^*}(\hat{P}'_{n-m}\left[\Delta_{\hat{M}^*}\hat{\Omega}^{*-1}_{\hat{M}^*}\Delta_{\hat{M}^*}+\tilde{K}_{M^*}\hat{\Omega}^{*-1}_{\hat{M}^*}\Delta_{\hat{M}^*}+\Delta_{\hat{M}^*}\hat{\Omega}^{*-1}_{\hat{M}^*}\tilde{K}_{M^*}\right.\\ &\quad \left.+\tilde{K}_{M^*}\left(\hat{\Omega}^{*-1}_{\hat{M}^*}-\hat{\Omega}^{*-1}_{M^*}\right)\tilde{K}_{M^*}\right]\frac{Z'_{n-m}\varepsilon}{\sqrt{n}})\end{aligned}$$

with $\Delta_{\hat{M}^*} = \tilde{K}_{\hat{M}^*} - \tilde{K}_{M^*}$. From Assumption (H) it follows for $c = \mathcal{A}/\mathcal{B}^{(q)}$ that $\left|k(j/\hat{M}^*) - k(j/M^*)\right| \leq C_1 n^{-1/(2q+2)} |j| \left|\hat{c}^{1/(2q+2)} - c^{1/(2q+2)}\right|$ for some constant C_1. Denote the k,j-th element of $\Omega^{*-1}_{M^*}$ by $\vartheta^*_{j_1,j_2}$. Then

$$\begin{aligned}
\sqrt{n/M^*}\hat{P}'_{n-m}\Delta_{\hat{M}^*}\hat{\Omega}^{*-1}_{\hat{M}^*}\tilde{K}_{M^*}\frac{Z'_{n-m}\varepsilon}{\sqrt{n}} &= n^{\frac{2q+1}{4q+4}}\frac{C_2}{\sqrt{n}}\sum_{t=1}^{n}\sum_{j_1,j_2=1}^{n-m}\hat{\Gamma}^{xy}_{j_1}\left[k(j_1/\hat{M}^*) - k(j_1/M^*)\right]\hat{\vartheta}^*_{j_1,j_2}k(j_2/M)v_{t,j_2} \\
&= n^{\frac{2q+1}{4q+4}}C_2\left[d_4^{\Delta} + d_5^{\Delta} + d_6^{\Delta} + d_7^{\Delta} + d_8^{\Delta} + d_9^{\Delta} + d^{\Delta}\right]
\end{aligned}$$

where $d^{\Delta} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sum_{j_1,j_2=1}^{n-m}\Gamma^{xy}_{j_1}\left[k(j_1/\hat{M}^*) - k(j_1/M^*)\right]\vartheta^*_{j_1,j_2}k(j_2/M^*)v_{t,j_2}$,

$$d_4^{\Delta} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sum_{j_1,j_2=1}^{n-m}\left(\hat{\Gamma}^{xy}_{j_1} - \check{\Gamma}^{xy}_{j_1}\right)\left[k(j_1/\hat{M}^*) - k(j_1/M^*)\right]\vartheta^*_{j_1,j_2}k(j_2/M^*)v_{t,j_2}$$

and similarly for $d_5^{\Delta},...,d_9^{\Delta}$ corresponding to Definitions (A.20-A.24) for $d_5,...,d_9$ where we replace K_M by $\tilde{\Delta}_{\hat{M}^*}$ and $\hat{\Omega}_M$ by $\hat{\Omega}^*_M$ in the same way as in d_4^{Δ}. We consider the largest term d_5^{Δ}

$$\left\|n^{\frac{2q+1}{4q+4}}d_5^{\Delta}\right\| \leq C_1 n^{-\frac{3}{4q+4}}\left|\hat{c}^{1/(2q+2)} - c^{1/(2q+2)}\right|\sum_{t=1}^{n}\sum_{j_1,j_2=1}^{n-m}|j_1|\left\|\left(\check{\Gamma}^{xy}_{j_1} - \Gamma^{xy}_{j_1}\right)\vartheta^*_{j_1,j_2}k(j_2/M^*)v_{t,j_2}\right\|.$$

By the same arguments as in the proof of Lemma (B.24) it follows that

$$\text{(A.28)}\qquad \sum_{j_1,j_2=1}^{n-m}|j_1|\left\|\left(\check{\Gamma}^{xy}_{j_1} - \Gamma^{xy}_{j_1}\right)\vartheta^*_{j_1,j_2}k(j_2/M^*)\sum\nolimits_{t=1}^{n}v_{t,j_2}\right\| = O_p(M^*) = O_p(n^{1/(2q+2)})$$

where we have used that $\sum_{j_1}|j_1|\left\|\vartheta^*_{j_1,j_2}\right\| < \infty$ uniformly in j_2 since $\vartheta^*_{j_1,j_2}$ has the same summability properties as ϑ_{j_1,j_2}. Also note that $k(j_2/M^*) = 0$ for $|j_2| > M^*$. The bound (A.28) implies that $n^{\frac{2q+1}{4q+4}}d_5^{\Delta} = O_p(n^{-(2q+3)/(4q+4)})$ as long as $\left|\hat{c}^{1/(2q+2)} - c^{1/(2q+2)}\right| = O_p(n^{-1/2})$ which follows from Lemma (4.1). Using similar arguments based on the proofs of Lemmas (B.23,B.25-B.28) it can be shown that the remaining terms $d_4^{\Delta}, d_6^{\Delta}..., d_9^{\Delta}$ are of smaller order. For d^{Δ} note that $\sum_{j_1,j_2=1}^{n}|j_1|\left\|\Gamma^{xy}_{j_1}\vartheta^*_{j_1,j_2}\right\| = O(1)$ such that

$$\sum_{j_1,j_2=1}^{n}|k(j_2/M^*)|\,|j_1|\left\|\Gamma^{xy}_{j_1}\vartheta^*_{j_1,j_2}\right\|\left(E\left\|\sum\nolimits_{t=1}^{n}v_{t,j_2}/\sqrt{n}\right\|^2\right)^{1/2} = O(1)$$

and thus $n^{\frac{2q+1}{4q+4}}d^{\Delta} = O_p(n^{-(2q+3)/(4q+4)})$ as long as $\left|\hat{c}^{1/(2q+2)} - c^{1/(2q+2)}\right| = O_p(n^{-1/2})$.

For $\sqrt{n/M^*}\hat{P}'_{n-m}\Delta_{\hat{M}^*}\hat{\Omega}^{*-1}_{\hat{M}^*}\Delta_{\hat{M}^*}\frac{Z'_{n-m}\varepsilon}{\sqrt{n}} = n^{\frac{2q+1}{4q+4}}C_2\left[d_4^{\Delta\Delta} + ... + d_9^{\Delta\Delta} + d^{\Delta\Delta}\right]$ we define

$$d^{\Delta\Delta} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sum_{j_1,j_2=0}^{n-m}\Gamma^{xy}_{j_1}\left[k(j_1/\hat{M}^*) - k(j_1/M^*)\right]\vartheta^*_{j_1,j_2}\left[k(j_2/\hat{M}^*) - k(j_2/M^*)\right]v_{t,j_2}$$

and similarly for the other terms. From $\sum_{j_1,j_2=1}^{n-m}\left(|j_1|\,|j_2|\left\|\Gamma^{xy}_{j_1}\vartheta^*_{j_1,j_2}\right\|^2 E\left\|\sum_{t=1}^{n}v_{t,j_2}/\sqrt{n}\right\|^2\right)^{1/2} = O(1)$ it follows that $n^{\frac{2q+1}{4q+4}}d^{\Delta\Delta} = O_p(n^{-(2q+7)/(4q+4)})$ using the fact that $\left|\hat{c}^{1/(2q+2)} - c^{1/(2q+2)}\right| = O_p(n^{-1/2})$ by Lemma (4.1). For $n^{\frac{2q+1}{4q+4}}d_5^{\Delta\Delta}$ note that

$$\left\|d_5^{\Delta\Delta}\right\| \leq \sum_{j_1,j_2=1}^{\max(\left[\hat{M}^*\right],[M^*])}|j_1|\,|j_2|\left\|\left(\check{\Gamma}^{xy}_{j_1} - \Gamma^{xy}_{j_1}\right)\vartheta^*_{j_1,j_2}\sum\nolimits_{t=1}^{n}v_{t,j_2}\right\|$$

such that for any finite $\epsilon > 0$ and some C

$$\begin{aligned} & P\left(M^{*-2}\left\|d_5^{\Delta\Delta}\right\| > C\right) \\ \leq\ & P\left(M^{*-1}\sum_{j_1,j_2=0}^{\lceil M^*+\epsilon\rceil}|j_1|\,|j_2|\left\|\left(\check{\Gamma}^{xy}_{j_1} - \Gamma^{xy}_{j_1}\right)\vartheta^*_{j_1,j_2}\sum\nolimits_{t=1}^{n}v_{t,j_2}\right\| > C\right) + P(\hat{M}^* > M^* + \epsilon). \end{aligned}$$

where $\lceil M^* + \epsilon\rceil$ denotes the smallest integer larger than $M^* + \epsilon$. Using the Markov inequality it follows that

$$M^{*-2}\sum_{j_1,j_2=0}^{\lceil M^*+\epsilon\rceil}|j_1|\,|j_2|\left(E\left\|\check{\Gamma}^{xy}_{j_1} - \Gamma^{xy}_{j_1}\right\|^2\right)^{1/2}\left\|\vartheta^*_{j_1,j_2}\right\|\left(E\left\|\sum\nolimits_{t=1}^{n}v_{t,j_2}/\sqrt{n}\right\|^2\right)^{1/2} = O(n^{-1/2})$$

by similar arguments as in the proof of Lemma (B.24). Therefore $n^{\frac{2q+1}{4q+4}}d_5^{\Delta\Delta} = O_p(n^{-(4q+5)/(4q+4)})$. The remaining terms are of smaller order by the same arguments as before.

Finally, for $\sqrt{n/M^*}\hat{P}'_{n-m}\tilde{K}_{M^*}\left(\hat{\Omega}^{*-1}_{\hat{M}^*} - \hat{\Omega}^{*-1}_{M^*}\right)\tilde{K}_{M^*}\frac{Z'_{n-m}\varepsilon}{\sqrt{n}}$ we expand $\hat{\Omega}^{*-1}_{\hat{M}^*}$ around $\hat{\Omega}^{*-1}_{M^*}$ and $\hat{\Omega}^{*-1}_{M^*}$ around $\Omega^{*-1}_{M^*}$ as in (A.1) leading to

$$\hat{\Omega}^{*-1}_{\hat{M}^*} = \hat{\Omega}^{*-1}_{M^*} - \hat{\Omega}^{*-1}_{M^*}(\hat{\Omega}^*_{M^*} - \hat{\Omega}^*_{\hat{M}^*})\hat{\Omega}^{*-1}_{M^*} + o_p(\left\|\hat{\Omega}^*_{\hat{M}^*} - \hat{\Omega}^*_{M^*}\right\|) \tag{A.29}$$

and

$$\hat{\Omega}^{*-1}_{M^*} = \Omega^{*-1}_{M^*} - \Omega^{*-1}_{M^*}(\Omega^*_{M^*} - \hat{\Omega}^*_{M^*})\Omega^{*-1}_{M^*} + o_p(\left\|\Omega^*_{M^*} - \hat{\Omega}^*_{M^*}\right\|) \tag{A.30}$$

Using the fact that $P(\hat{M}^* > M^* + \epsilon)$ tends to zero we can show that $\left\|\hat{\Omega}^*_{\hat{M}^*} - \hat{\Omega}^*_{M^*}\right\| = O_p(M^*/n^{1/2}) = O_p(n^{\frac{-q}{2q+2}})$ while $\left\|\Omega^*_{M^*} - \hat{\Omega}^*_{M^*}\right\| = O_p(n^{-q/(2q+2)})$ by Lemma (B.8). Combining (A.29) and (A.30) then leads to

$$\hat{\Omega}^{*-1}_{\hat{M}^*} - \hat{\Omega}^{*-1}_{M^*} = -\mathcal{O}_{M^*}(\hat{\Omega}^*_{M^*} - \hat{\Omega}^*_{\hat{M}^*})\mathcal{O}_{M^*} + o_p(\left\|\Omega^*_{M^*} - \hat{\Omega}^*_{M^*}\right\|)$$

with $\mathcal{O}_{M^*} = \Omega^{*-1}_{M^*} - \Omega^{*-1}_{M^*}(\Omega^*_{M^*} - \hat{\Omega}^*_{M^*})\Omega^{*-1}_{M^*}$. We then consider

$$\begin{aligned}
\text{(A.31)}\quad d_{10} &\equiv \sqrt{n/M^*}\hat{P}'_{n-m}\tilde{K}_{M^*}\Omega^{*-1}_{M^*}(\hat{\Omega}^*_{M^*} - \hat{\Omega}^*_{\hat{M}^*})\Omega^{*-1}_{M^*}\tilde{K}_{M^*}\frac{Z'_{n-m}\varepsilon}{\sqrt{n}} \\
\text{(A.32)}\quad d_{11} &\equiv \sqrt{n/M^*}\hat{P}'_{n-m}\tilde{K}_{M^*}\Omega^{*-1}_{M^*}(\Omega^*_{M^*} - \hat{\Omega}^*_{M^*})\Omega^{*-1}_{M^*}(\hat{\Omega}^*_{M^*} - \hat{\Omega}^*_{\hat{M}^*})\Omega^{*-1}_{M^*}\tilde{K}_{M^*}\frac{Z'_{n-m}\varepsilon}{\sqrt{n}} \\
&\quad + \sqrt{n/M^*}\hat{P}'_{n-m}\tilde{K}_{M^*}\Omega^{*-1}_{M^*}(\hat{\Omega}^*_{M^*} - \hat{\Omega}^*_{\hat{M}^*})\Omega^{*-1}_{M^*}(\Omega^*_{M^*} - \hat{\Omega}^*_{M^*})\Omega^{*-1}_{M^*}\tilde{K}_{M^*}\frac{Z'_{n-m}\varepsilon}{\sqrt{n}} \\
\text{(A.33)}\quad d_{12} &\equiv \sqrt{n/M^*}\hat{P}'_{n-m}\tilde{K}_{M^*}\Omega^{*-1}_{M^*}(\Omega^*_{M^*} - \hat{\Omega}^*_{M^*})\Omega^{*-1}_{M^*}(\hat{\Omega}^*_{M^*} - \hat{\Omega}^*_{\hat{M}^*}) \\
&\quad \times \Omega^{*-1}_{M^*}(\Omega^*_{M^*} - \hat{\Omega}^*_{M^*})\Omega^{*-1}_{M^*}\tilde{K}_{M^*}\frac{Z'_{n-m}\varepsilon}{\sqrt{n}}.
\end{aligned}$$

By Lemma (B.37) it follows that $d_{10} = o_p(1)$, Lemma (B.38) establishes $d_{11} = o_p(1)$ and Lemma (B.39) shows $d_{12} = o_p(1)$. It now follows that

$$\sqrt{n/M^*}\hat{P}'_{n-m}\tilde{K}_{M^*}\left(\hat{\Omega}^{*-1}_{\hat{M}^*} - \hat{\Omega}^{*-1}_{M^*}\right)\tilde{K}_{M^*}\frac{Z'_{n-m}\varepsilon}{\sqrt{n}} = d_{10} + d_{11} + d_{12} + o_p(1)$$

by the Taylor expansion theorem.

Next consider

$$\begin{aligned}
\sqrt{n/M^*}(\hat{D}_{\hat{M}^*} - \hat{D}_{M^*}) &= \sqrt{n/M^*}(\hat{P}'_{n-m}\left[\Delta_{\hat{M}^*}\hat{\Omega}^{*-1}_{\hat{M}^*}\Delta_{\hat{M}^*} + \tilde{K}_{M^*}\hat{\Omega}^{*-1}_{\hat{M}^*}\Delta_{\hat{M}^*} + \Delta_{\hat{M}^*}\hat{\Omega}^{*-1}_{\hat{M}^*}\tilde{K}_{M^*}\right. \\
&\quad \left. + \tilde{K}_{M^*}\left(\hat{\Omega}^{*-1}_{\hat{M}^*} - \hat{\Omega}^{*-1}_{M^*}\right)\tilde{K}_{M^*}\right]\hat{P}).
\end{aligned}$$

First, we analyze $\hat{P}'_{n-m}\Delta_{\hat{M}^*}\hat{\Omega}^{*-1}_{\hat{M}^*}K'_{M^*}\hat{P}_{n-m} = H^{\Delta}_2 + H^{\Delta}_3 + H^{\Delta}$ where $H^{\Delta}_2 = H^{\Delta}_{211} + H^{\Delta}_{212} + H^{\Delta}_{221} + H^{\Delta}_{222}$, $H^{\Delta}_3 = H^{\Delta}_{31} + H^{\Delta}_{32} + H^{\Delta}_{33} + H^{\Delta}_{34}$ and the definitions follow from the definitions in (A.6)-(A.13) with the appropriate substitutions for $\hat{\Omega}^{*-1}_{\hat{M}^*}$ and $\Delta_{\hat{M}^*}$. Furthermore let $H^{\Delta} = \sum_{i=0}^{n}\sum_{j=0}^{n}\Gamma^{xy}_i(k(i/\hat{M}^*) - k(i/M^*))\vartheta^*_{i,j}k(j/M^*)\Gamma^{yx}_{-j}$. Then

$$\begin{aligned}
\sqrt{n/M^*}\left\|H^{\Delta}\right\| &\leq n^{\frac{2q+1}{4q+4}}C_1 n^{-1/(2q+2)}\left|\hat{c}^{1/(2q+2)} - c^{1/(2q+2)}\right|\sum_{j_1,j_2=1}^{n-m}|j_1|\left\|\Gamma^{xy}_{j_1}\vartheta^*_{j_1,j_2}k(j_2/M^*)\Gamma^{yx}_{-j_2}\right\| \\
&= n^{-3/(4q+4)}C_1\sqrt{n}\left|\hat{c}^{1/(2q+2)} - c^{1/(2q+2)}\right|O(1) = O_p(n^{-3/(4q+4)}).
\end{aligned}$$

Now consider H_{222}^{Δ}

$$\sqrt{n/M^*}\left\|H_{222}^{\Delta}\right\| \leq n^{-3/(4q+4)}C_1\sqrt{n}\left|\hat{c}^{1/(2q+2)} - c^{1/(2q+2)}\right|\sum_{j_1,j_2=1}^{n-m}|j_1|\left\|\left(\check{\Gamma}_{j_1}^{xy} - \Gamma_{j_1}^{xy}\right)\vartheta_{j_1,j_2}^* k(j_2/M^*)\check{\Gamma}_{-j_2}^{yx}\right\|$$

and by the proof of Lemma (B.15) it follows that $E\sum_{j_1,j_2=1}^{n-m}|j_1|\left\|\left(\check{\Gamma}_{j_1}^{xy} - \Gamma_{j_1}^{xy}\right)\vartheta_{j_1,j_2}^* k(j_2/M^*)\check{\Gamma}_{-j_2}^{yx}\right\| = O(n^{-1/2}M^*)$ such that $\sqrt{n/M^*}\left\|H_{222}^{\Delta}\right\| = o_p(n^{-3/(4q+4)})$. Using the results of Lemma (B.17) we can show in the same way that $\sqrt{n/M^*}\left\|H_{34}^{\Delta}\right\| = o_p(1)$. All the remaining terms are of lower order by Lemmas (B.12-B.16).

Next, we turn to $\hat{P}_{n-m}'\Delta_{\hat{M}^*}\hat{\Omega}_{\hat{M}^*}^{*-1}\Delta_{\hat{M}^*}\hat{P}_{n-m} = H_2^{\Delta\Delta} + H_3^{\Delta\Delta} + H^{\Delta\Delta}$ where $H_2^{\Delta\Delta}, H_3^{\Delta\Delta}, H^{\Delta\Delta}$ are defined in the obvious way. It follows immediately that

$$\begin{aligned}\sqrt{n/M^*}\left\|H^{\Delta\Delta}\right\| &\leq n^{-(2q+5)/(4q+4)}C_1 n\left|\hat{c}^{1/(2q+2)} - c^{1/(2q+2)}\right|^2\sum_{j_1,j_2=1}^{n-m}|j_1|\,|j_2|\left\|\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}^*\Gamma_{-j_2}^{yx}\right\| \\ &= O_p(n^{-(2q+5)/(4q+4)}).\end{aligned}$$

For $H_{222}^{\Delta\Delta}$ we note that $E\sum_{j_1,j_2=1}^{n-m}|j_1|\,|j_2|\left\|\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}^*\Gamma_{-j_2}^{yx}\right\| = O(1)$ such that again $\sqrt{n/M^*}\left\|H_{222}^{\Delta\Delta}\right\| = o_p(1)$. The same type of arguments also establish $\sqrt{n/M^*}\left\|H_{34}^{\Delta\Delta}\right\| = o_p(1)$. All the other terms are of lower order.

Finally, we turn to $\sqrt{n/M^*}\hat{P}_{n-m}'\tilde{K}_{M^*}\left(\hat{\Omega}_{\hat{M}^*}^{*-1} - \hat{\Omega}_{M^*}^{*-1}\right)\tilde{K}_{M^*}\hat{P}_{n-m}$ for which we consider

$$\begin{aligned}
&(\text{A.34})\ H_5 \equiv \sqrt{n/M^*}\hat{P}_{n-m}'\tilde{K}_{M^*}\Omega_{M^*}^{*-1}(\hat{\Omega}_{M^*}^* - \hat{\Omega}_{\hat{M}^*}^*)\Omega_{M^*}^{*-1}\tilde{K}_{M^*}\hat{P}_{n-m} \\
&(\text{A.35})\ H_6 \equiv \sqrt{n/M^*}\hat{P}_{n-m}'\tilde{K}_{M^*}\Omega_{M^*}^{*-1}(\Omega_{M^*}^* - \hat{\Omega}_{M^*}^*)\Omega_{M^*}^{*-1}(\hat{\Omega}_{M^*}^* - \hat{\Omega}_{\hat{M}^*}^*)\Omega_{M^*}^{*-1}\tilde{K}_{M^*}\hat{P}_{n-m} \\
&\qquad + \sqrt{n/M^*}\hat{P}_{n-m}'\tilde{K}_{M^*}\Omega_{M^*}^{*-1}(\hat{\Omega}_{M^*}^* - \hat{\Omega}_{\hat{M}^*}^*)\Omega_{M^*}^{*-1}(\Omega_{M^*}^* - \hat{\Omega}_{M^*}^*)\Omega_{M^*}^{*-1}\tilde{K}_{M^*}\hat{P}_{n-m} \\
&(\text{A.36})\ H_7 \equiv \sqrt{n/M^*}\hat{P}_{n-m}'\tilde{K}_{M^*}\Omega_{M^*}^{*-1}(\Omega_{M^*}^* - \hat{\Omega}_{M^*}^*)\Omega_{M^*}^{*-1}(\hat{\Omega}_{M^*}^* - \hat{\Omega}_{\hat{M}^*}^*) \\
&\qquad \times\Omega_{M^*}^{*-1}(\Omega_{M^*}^* - \hat{\Omega}_{M^*}^*)\Omega_{M^*}^{*-1}\tilde{K}_{M^*}\hat{P}_{n-m}.
\end{aligned}$$

Lemmas (B.40-B.42) establish that $H_5, H_6, H_7 = o_p(1)$.

Next we show that

$$\lim_{h\to\infty}\lim_{n\to\infty}\left[\varphi_{n,h}(\tilde{M}^*, \ell, k(.), b_{n,\hat{M}^*}) - \varphi_{n,h}(\tilde{M}^*, \ell, k(.), b_{n,\tilde{M}^*})\right] = 0.$$

This result was shown by Andrews (1991), Theorem 3(c) and follows immediately from $n\sqrt{1/M^*}(\hat{\beta}_{n,\hat{M}^*} - \hat{\beta}_{n,M^*}) = o_p(1)$. ■

Proof of Proposition (5.1) We consider Ed_i and EH_iDd_j. First, $Ed_i = 0$ for $i \leq 3$. The terms $d_4, d_6, ..., d_9$ are of lower order by Lemmas (B.23,B.25-B.28). The terms $EH_iD^{-1}d_j$ are all of lower order. The largest order term is therefore Ed_5. By the proof of Lemma (B.36) it follows that $Ed_5 = M/\sqrt{n}\mathcal{A}_1 \int k^2(x)dx + o(M/\sqrt{n})$. ■

Proof of Theorem (5.3) We consider the expansion of $\sqrt{n}\left(\beta^*_{n,M} - \beta\right)$as before. 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_5 by

$$\bar{d}_5 = d_5 - \frac{M}{\sqrt{n}}\mathcal{A}_1' \int k^2(x)dx$$

and the additional term $d_{13} = M/\sqrt{n}(\widehat{\mathcal{A}}_1 - \mathcal{A}_1)' \int k^2(x)dx$ needs to be considered. First note that $Ed_5 - \frac{M}{\sqrt{n}} D^{-1}\mathcal{A}_1' \int k^2(x)dx = o(1)$. Then $E\bar{d}_5\bar{d}_5' = E(d_5 - Ed_5)(d_5 - Ed_5)' + o(1)$. From the proof of Lemma (B.36) it follows that $E\bar{d}_5\bar{d}_5' = O(M/n)$. Also $E\ell' H_{222} D^{-1} d_0 \bar{d}_5' D^{-1/2}\ell = o(M/n)$ by Lemma (B.33) and $d_{13} = O_p(M/n)$ together with Lemma (B.32) shows that all remaining terms are at most of order M/n. ■

Proof of Theorem (5.4): First note that

$$\beta^{\mathbb{N}}_{n,M} = \left[(1+\Bbbk)\hat{P}_M' K_M \hat{\Omega}_M^{-1} K_M \hat{P}_M - n^{-1}\Bbbk X' A_n(\hat{\Phi}) X\right]^{-1} n^{-1}\left((1+\Bbbk)\hat{P}_M' K_M \hat{\Omega}_M^{-1} K_M Z_n' - \Bbbk X' A_n(\hat{\Phi})\right) Y$$

with $n^{-1/2}\Bbbk\hat{P}_M' K_M \hat{\Omega}_M^{-1} K_M Z_n'\varepsilon = O_p(M/n)$ and $\Bbbk\hat{P}_M' K_M \hat{\Omega}_M^{-1} K_M \hat{P}_M = O_p(M/n)$. It therefore follows that $\beta^{\mathbb{N}}_{n,M} - \beta^{\Bbbk}_{n,M} = O_p(M/n)$. Next note that

$$n^{-1/2}\Bbbk E X' A_n(\Phi)\varepsilon = M/n^{3/2}\sum_{t=1}^{n}\sum_{s=1}^{n}\zeta_{t-s}\Gamma^{\varepsilon x}_{t-s} + o(M/n^{1/2}) = M/n^{1/2}\int k(x)^2 dx \mathcal{A}_1' + o(M/n^{1/2})$$

by the Toeplitz lemma. The term $o(M/n^{1/2})$ stands for replacing $x_t - \bar{x}$ by $x_t - \mu_x$ in X and allows the sums not to be exactly from $t = 1, .., n$. The details of these calculations are omitted. The variance covariance matrix of $n^{-1/2}\Bbbk X' A_n(\Phi)\varepsilon$ is given by

$$\mathrm{Var}(n^{-1/2}\Bbbk X' A_n(\Phi)\varepsilon) = n^{-1}\Bbbk^2 \sum_{t_1,t_2=1}^{n}\sum_{s_1,s_2=1}^{n}\zeta_{t_1-s_1}\zeta_{t_2-s_2}\left[\Gamma^{x\varepsilon}_{t_1-s_2}\Gamma^{\varepsilon x}_{t_2-s_1} + \gamma^{\varepsilon}_{t_1-t_2}\Gamma^{xx}_{s_1-s_2}\right] + o(M^2/n)$$

and setting $u = t_1 - s_2$ and $v = t_2 - s_1$ leads to

$$\frac{M^2}{n^3} \sum_{t_1,t_2=1}^{n} \sum_{s_1,s_2=1}^{n} \zeta_{t_1-s_1}\zeta_{t_2-s_2} \left[\Gamma^{x\varepsilon}_{t_1-s_2}\Gamma^{\varepsilon x}_{t_2-s_1} + \gamma^{\varepsilon}_{t_1-t_2}\Gamma^{xx}_{s_1-s_2}\right]$$

$$= \frac{M^2}{n^3} \sum_{u,v=-n+1}^{n-1} \sum_{t_2=\max(v+1,1)}^{\min(n,n+v)} \sum_{s_2=\max(-u+1,1)}^{\min(n,n-u)} \zeta_{u+t_2-s_2}\zeta_{t_2-s_2} \left[\Gamma^{x\varepsilon}_{u}\Gamma^{\varepsilon x}_{v} + \gamma^{\varepsilon}_{u+s_2-t_2}\Gamma^{xx}_{v+s_2-t_2}\right]$$

$$= O(M^2/n^2)$$

which shows that $n^{-1/2}\mathbb{k}X'A_n(\Phi)\varepsilon - M/n^{1/2}\int k(x)^2 dx\mathcal{A}_1' = O_p(M/n)$ such that

$$d_{14} = n^{-1/2}\mathbb{k}X'A_n(\hat{\Phi})\varepsilon - M/n^{1/2}\int k(x)^2 dx\mathcal{A}_1' = O_p(M/n).$$

This follows from $n^{-1/2}\mathbb{k}X'\left(A_n(\hat{\Phi}) - A_n(\Phi)\right)\varepsilon = o_p(M/n)$. In the same way it can be shown that $n^{-1}\mathbb{k}X'A_n(\Phi)X = O_p(M/n)$. The result now follows from adding the term

$$-d_{14} - M/n^{1/2}\int k(x)^2 dx\mathcal{A}_1'$$

to the expansion of $\hat{d}_n$. ■

B. Lemmas

B.1. General Results

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 B.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*

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

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

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

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

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

Lemma B.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

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

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

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

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

For the summability of the cumulant note that

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

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

Definition B.4. *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).*

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

Lemma B.5. *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*

$$\begin{aligned} EW'AXY'BZ &= \left(\mathrm{vec}\, A'\right)' E(X \otimes W) E(Z' \otimes Y') \,\mathrm{vec}\, B' + \mathrm{tr}(EAXZ')(EB'YW') \\ &\quad + \mathrm{tr}(EAXY')(EBZW') + \mathcal{K}_4 \end{aligned}$$

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

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

Lemma B.6. *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*

$$\begin{aligned} E\,\mathrm{tr}\, WAXY'BZ &= \left(\mathrm{vec}\, B'\right)' E(Y \otimes Z) E(X' \otimes W') \,\mathrm{vec}\, A + \mathrm{tr}(EAXY')(EBZW') \\ &\quad + \mathrm{tr}(B' \otimes I) E(Y' \otimes W)(I \otimes A) E \,\mathrm{vec}(X) \,\mathrm{vec}(Z')' + \mathcal{K}_4 \end{aligned}$$

where $\mathcal{K}_4 = \sum\limits_k \sum\limits_{j_1,...,j_4} ... \sum a_{j_1,j_2} b_{j_3,j_4} \,\mathrm{cum}^*(w_{j_1,k}, x_{j_2}, y_{j_3}, z_{j_4,k})$*.*

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

Lemma B.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 $\ell \in \mathbb{R}^{pr+p-1}$ is a vector of constants such that $\ell'\ell = 1$ then*

i) $E(v_{t,i} \otimes \breve{w}'_{s,j}\ell) = ((\operatorname{vec}(\Gamma^{yy}_{s-t+i-j}) \otimes \left(\Gamma^{\varepsilon x}_{t-s}\right)') + K_{pp}(\Gamma^{\varepsilon y}_{t-s+j} \otimes \Gamma^{yx}_{t-i-s}) + \mathcal{K}^1_4)(I \otimes \ell)$ where $\mathcal{K}^1_4$ is a $p^2 \times (pr + p - 1)$ matrix with typical element (a, b) equal to

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

and K_{pp} is defined in (B.4).

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)

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

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)

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

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

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

v)$E(v_{t,i}vec(w'_{s,j})') = \left(\Gamma^{\varepsilon x}_{t-s}\right)' \otimes \Gamma^{yy}_{t-i+j-s} + \Gamma^{yx}_{t-i-s} \otimes (\Gamma^{\varepsilon y}_{t-s+j}) + \mathcal{K}^5_4(t, s, i, j)$ where $\mathcal{K}^5_4$ is a $p \times p$ matrix with typical element (a, b)

$$\left[\mathcal{K}^5_4(t, s, i, j)\right]_{a,b} = \operatorname{cum}^*(\varepsilon_{t+m}, y^a_{t-i}, y^{b \bmod p+1}_{s-j}, x^{[b/r]+1}_{s+m}).$$

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 B.8. *Let $\hat{\Omega}_M$ be defined in (3.1) and $\sqrt{n}(\tilde{\beta}_n - \beta_0) = O_p(1)$. Let $M \to \infty$ such that $M/n^{1/2} \to 0$. Then $\left\|\hat{\Omega}_M - \Omega_M\right\| = O_p(M/n^{1/2})$*

Proof. Note that

$$\begin{aligned}\left\|\hat{\Omega}_M - \Omega_M\right\| \;\; &\leq \sum_{l=-m+1}^{m-1} |\gamma_l^{\varepsilon}| \left\|\hat{\Omega}_M(l) - \Omega_M(l)\right\| \\ &\quad + \sum_{l=-m+1}^{m-1} |\hat{\gamma}_l^{\varepsilon} - \gamma_l^{\varepsilon}| \left(\left\|\hat{\Omega}_M(l) - \Omega_M(l)\right\| + \|\Omega_M(l)\|\right)\end{aligned}$$

where $|\hat{\gamma}_l^{\varepsilon} - \gamma_l^{\varepsilon}| = O_p(n^{-1/2})$ uniformly in l for a consistent first stage estimate $\tilde{\beta}_n$. For $\hat{\Omega}_M(l)$ we use

$$\left\|\hat{\Omega}_M(l) - \Omega_M(l)\right\|^2 = \sum_{i,j=1}^{M} \|\hat{\omega}_{i,j}(l) - \omega_{i,j}(l)\|^2$$

Define $r_1 = \max(i+1, j+l+1)$, $r_2 = \min(n, n+j)$ and

$$\breve{\omega}_{i,j}(l) = n^{-1} \sum_{t=r_1}^{r_2} w_{t,j+l-i}^{y} = n^{-1} \sum_{t=r_1}^{r_2} \breve{w}_{t,j+l-i}^{y} + \frac{n - r_1 + r_2}{n} \Gamma_{j+l-i}^{yy}$$

with $\breve{w}_{t,j+l-i}^{y} = w_{t,j+l-i}^{y} - \Gamma_{j+l-i}^{yy}$ such that $\hat{\omega}_{i,j}(l) - \breve{\omega}_{i,j}(l) = O_p(n^{-1/2})$ uniformly in i, j, l. For $M/n^{1/2} \to 0$ it thus follows that $\left\|\hat{\Omega}_M(l) - \Omega_M(l)\right\|^2 = \sum_{i,j=1}^{M} \|\breve{\omega}_{i,j}(l) - \omega_{i,j}(l)\|^2 + O_p(M^2/n)$. Now

$$\begin{aligned}\|\breve{\omega}_{i,j}(l) - \omega_{i,j}(l)\|^2 \leq \left\|n^{-1} \sum\nolimits_{t=r_1}^{r_2} \breve{w}_{t,j+l-i}^{y}\right\|^2 + \left(\frac{2M}{n}\right)^2 \left\|\Gamma_{j+l-i}^{yy}\right\|^2 \\ + 4\frac{M}{n} \left\|\Gamma_{j+l-i}^{yy}\right\| \left\|n^{-1} \sum\nolimits_{t=r_1}^{r_2} \breve{w}_{t,j+l-i}^{y}\right\|\end{aligned}$$

where

$$\begin{aligned}E\left\|n^{-1} \sum\nolimits_{t=r_1}^{r_2} \breve{w}_{t,j+l-i}^{y}\right\|^2 &\leq n^{-2} \sum\nolimits_{t=r_1}^{r_2} E\,\mathrm{tr}(w_{t,j+l-i} - \Gamma_{j+l-i}^{yy})(w_{s,j+l-i} - \Gamma_{j+l-i}^{yy})' \\ &= n^{-2} \sum\nolimits_{t=r_1}^{r_2} \left[\gamma_{t-i+j+s}^{yy}\gamma_{t-s}^{yy} + \gamma_{t-i+s}^{yy}\gamma_{t-s+j}^{yy} + \mathcal{K}_4^4\right] = O(n^{-1})\end{aligned}$$

uniformly in i and j. Moreover $E\left\|n^{-1} \sum_{t=r_1}^{r_2} \breve{w}_{t,j+l-i}^{y}\right\|^2$ is summable in one of the indices, say j. Thus $\sum_{i,j=1}^{M} \left\|n^{-1} \sum_{t=r_1}^{r_2} \breve{w}_{t,j+l-i}^{y}\right\|^2 = O_p(M/n^{1/2})$ while $\left(\frac{2M}{n}\right)^2 \sum_{i,j=1}^{M} \left\|\Gamma_{j+l-i}^{yy}\right\|^2 = O(M^3/n^2) = o(M/n)$ and $\frac{M}{n} \sum_{i,j=1}^{M} \left\|\Gamma_{j+l-i}^{yy}\right\| \left\|n^{-1} \sum_{t=r_1}^{r_2} \breve{w}_{t,j+l-i}^{y}\right\| = O_p(M^2/n^{3/2}) = o_p(M/n)$. The result then follows from the Markov inequality,

$$E \sum_{l=-m+1}^{m-1} \left\|\hat{\Omega}_M(l) - \Omega_M(l)\right\| \leq \sum_{l=-m+1}^{m-1} \left(E\left\|\hat{\Omega}_M(l) - \Omega_M(l)\right\|^2\right)^{1/2}$$

and the fact that $\|\Omega_M(l)\| = O(M)$. ■

B.2. Results for Lemma 3.1

Lemma B.9. *Let H_{11} be defined as in (A.2). Then $H_{11} = o(M^{-2s})$ where s is defined in Assumption (C).*

Proof. Assumption (C) implies that $\sum_j |j|^s \|\Gamma^{yy}\| < \infty$. We let $H_{11} = H_{111} + H_{112}$ with $H_{111} = P'\left(\Omega_M^{*-1} - \Omega^{-1}\right)P$ and $H_{112} = P_M'\Omega_M^{-1}P_M - P\Omega_M^{*-1}P$ where Ω_M^* is an infinite dimensional matrix defined by

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

and I_∞ stands for the infinite dimensional identity matrix (see Kuersteiner 1999b, Lemma 4.2 for details). Let $\vartheta_{i,j}^M$ be the i,j-th $p \times p$ block of Ω_M^{*-1}

$$\begin{aligned}
\|H_{111}\| &= \left\|P'\Omega^{-1}\left(\Omega - \Omega_M^*\right)\Omega_M^{*-1}P\right\| \\
&\leq \textstyle\sum_{j_3>M}\sum_{j_1 j_4}\left\|\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}\right\|\|\omega_{j_3,j_3} - I\|\left\|\vartheta_{j_3,j_4}^M\Gamma_{-j_4}^{yx}\right\| \\
&\quad +2\textstyle\sum_{j_3>M}\sum_{j_1,j_2,j_4}\left\|\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}\right\|\|\omega_{j_2,j_3}\|\left\|\vartheta_{j_3,j_4}^M\Gamma_{-j_4}^{yx}\right\| \\
&\leq M^{-2s}\textstyle\sum_{j_3>M}\sum_{j_1 j_4}|j_3|^{2s}\left\|\Gamma_{j_1}^{xy}\vartheta_{j_1,j_3}\right\|\|\omega_{j_3,j_3} - I\|\left\|\vartheta_{j_3,j_4}^M\Gamma_{-j_4}^{yx}\right\| \\
&\quad +M^{-2s}2\textstyle\sum_{j_3>M}\sum_{j_1,j_2,j_4}|j_3|^{2s}\left\|\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}\right\|\|\omega_{j_2,j_3}\|\left\|\vartheta_{j_3,j_4}^M\Gamma_{-j_4}^{yx}\right\|
\end{aligned}$$

which tends to zero as $M \to \infty$ such that $\|H_{111}\| = o(M^{-2s})$. The last line follows from the fact that $\sum_{j_1,j_3}|j_3|^s\left\|\Gamma_{j_1}^{xy}\vartheta_{j_1,j_3}\right\| < \infty$ which can be shown by arguments similar to the proof of Lemma 5.2 in Kuersteiner (1999b). Note that

$$H_{112} = -\sum_{j_1=[M]+1}^{\infty}\Gamma_{j_1}^{xy}\Gamma_{-j_1}^{yx}$$

and use the inequality

$$\|H_{112}\| \leq \left(\sum_{j=[M]+1}^{\infty}\left\|\Gamma_j^{xy}\right\|\right)^2 \leq M^{-2s}\left(\sum_{j=[M]+1}^{\infty}|j|^s\left\|\Gamma_j^{xy}\right\|\right)^2 = o(M^{-2s})$$

leading to $\|H_{11}\| = o(M^{-2s})$. ∎

Lemma B.10. *Let H_{12} be defined in (A.3). Then $H_{12} = O(M^{-2q})$.*

Proof. We write $H_{12} = M^{-2q} \sum_{j_1,j_2=1}^{[M]} \Gamma_{j_1}^{xy} |j_1|^q \frac{1-k(j_1/M)}{|j_1/M|^q} \vartheta_{j_1,j_2}^M \frac{1-k(j_2/M)}{|j_2/M|^q} |j_2|^q \Gamma_{-j_2}^{yx}$. By the Dominated Convergence Theorem

$$\sum_{j_1,j_2=1}^{[M]} \Gamma_{j_1}^{xy} |j_1|^q \frac{1-k(j_1/M)}{|j_1/M|^q} \vartheta_{j_1,j_2}^M \frac{1-k(j_2/M)}{|j_2/M|^q} |j_2|^q \Gamma_{-j_2}^{yx}$$

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

where we have used Assumption (F) such that $H_{12} = M^{-2q} k_q^2 \mathcal{B}_0^{(q)} + o(M^{-2q}) = O(M^{-2q})$. ■

Lemma B.11. *Let $H_{13}$and H_{14} be defined in (A.4) and (A.5). Then $H_{13} + H_{14} = O(M^{-q})$.*

Proof. Follows from Lemma (B.34). ■

Lemma B.12. *Let H_{211} be defined in (A.6). Then $H_{211} = O_p(M/n)$.*

Proof. We write $H_{211} = \sum_{j_1,j_2=1}^{[M]} \left(\hat{\Gamma}_{j_1}^{xy} - \check{\Gamma}_{j_1}^{xy}\right) k(j_1/M) \vartheta_{j_1,j_2}^M k(j_2/M) \left(\hat{\Gamma}_{j_2}^{xy} - \check{\Gamma}_{j_2}^{xy}\right)'$ where $\check{\Gamma}_j^{xy} = n^{-1} \sum_{t=j+1}^{n-m} w_{t,j}$. First note that

$$\|H_{211}\| \leq \sum_{j_1,j_2=1}^{[M]} \left\|\hat{\Gamma}_{j_1}^{xy} - \check{\Gamma}_{j_1}^{xy}\right\| \left\|\vartheta_{j_1,j_2}^M\right\| \left\|\hat{\Gamma}_{j_2}^{xy} - \check{\Gamma}_{j_2}^{xy}\right\|.$$

One obtains

$$\left\|\hat{\Gamma}_j^{xy} - \check{\Gamma}_j^{xy}\right\| \leq \|\bar{x} - \mu_x\| \left\|n^{-1} \sum\nolimits_{t=j+1}^{n-m} (y_t - \bar{y})\right\| + \|\bar{y} - \mu_y\| \left\|n^{-1} \sum\nolimits_{t=j+1}^{n-m} (x_t - \bar{x})\right\| + \left\|\left(\bar{y} - \mu_y\right)\left(\bar{x} - \mu_x\right)\right\|$$

with $E\left\|\left(\bar{y} - \mu_y\right)\left(\bar{x} - \mu_x\right)\right\|^2 = O(n^{-2})$. Further, $\left\|n^{-1} \sum_{t=j+1}^{n-m} (y_t - \bar{y})\right\| \leq \|\bar{y}\| + n^{-1} \sum_{t=1}^{n} \|y_t\| = O_p(1)$ leading to

$$\left\|\hat{\Gamma}_j^{xy} - \check{\Gamma}_j^{xy}\right\| = O_p(n^{-1/2}) \tag{B.2}$$

uniformly in j such that H_{211} is bounded in expectation by $n^{-1} c_1 \sum_{j_1,j_2=1}^{[M]} \left\|\vartheta_{j_1,j_2}^M\right\| = O(M/n)$ for some constant c_1.

Lemma B.13. *Let H_{212} be defined in (A.7). Then $H_{212} = O_p(n^{-1/2})$.*

Proof. From (A.7) H_{212} can be written as

$$H_{212} = -\sum_{j_1,j_2=1}^{[M]} \left(\hat{\Gamma}_{j_1}^{xy} - \check{\Gamma}_{j_1}^{xy}\right) k(j_1/M)\vartheta_{j_1,j_2}^M k(j_2/M)\check{\Gamma}_{-j_2}^{yx} - \sum_{j_1,j_2=1}^{[M]} \check{\Gamma}_{j_1}^{xy} k(j_1/M)\vartheta_{j_1,j_2}^M k(j_2/M)\left(\hat{\Gamma}_{j_2}^{xy} - \check{\Gamma}_{j_2}^{xy}\right)'.$$

First note that

$$\|H_{212}\| \leq \sum_{j_1,j_2=1}^{[M]} \left\|\hat{\Gamma}_{j_1}^{xy} - \check{\Gamma}_{j_1}^{xy}\right\| \|\vartheta_{j_1,j_2}^M\| \left\|\check{\Gamma}_{j_2}^{xy}\right\| + \left\|\check{\Gamma}_{j_1}^{xy}\right\| \|\vartheta_{j_1,j_2}^M\| \left\|\hat{\Gamma}_{j_2}^{xy} - \check{\Gamma}_{j_2}^{xy}\right\|.$$

Now using Lemma (B.7iv)

$$\begin{aligned} E\left\|\check{\Gamma}_j^{xy}\right\|^2 &= n^{-2}\sum\nolimits_{t,s=j+1}^{n-m} \operatorname{tr} E(x_{t+m}-\mu_x)(y_{t-j}-\mu_y)'(y_{s-j}-\mu_y)(x_{s+m}-\mu_x)' \\ &= n^{-2}\sum\nolimits_{t,s=j+1}^{n-m} \operatorname{tr}(\Gamma_j^{xy}\Gamma_{-j}^{yx} + \Gamma_{s-t}^{xx}\gamma_{s-t}^{yy} + \Gamma_{t-s+j}^{xy}\Gamma_{t-s-j}^{yx} + \mathcal{K}_4^4) \\ &= \left\|\Gamma_j^{xy}\right\|^2 + O(n^{-1}) \end{aligned}$$

where $\gamma_{s-t}^{yy} = E(y_{t-j}-\mu_y)'(y_{s-j}-\mu_y)$ and $\mathcal{K}_4^4$ is a matrix containing fourth order cumulants of x_{t+m} and y_{t-j} defined in Lemma (B.7iv). This together with the arguments in the proof of the previous lemma shows that $E\|H_{212}\| \leq n^{-1/2}2\sum_{j_1,j_2=0}^{[M]} \|\vartheta_{j_1,j_2}^M\| \left\|\Gamma_{j_2}^{yx}\right\|^2 + O(M/n^{3/2}) = O(Mn^{-3/2}+n^{-1/2}) = O(n^{-1/2})$. ■

Lemma B.14. *Let H_{221} be defined in (A.8). Then $H_{221} = O_p(M/n)$.*

Proof. From the definition $H_{221} = \sum_{j_1,j_2=1}^{[M]} \left(\check{\Gamma}_{j_1}^{xy} - \Gamma_{j_1}^{xy}\right) k(j_1/M)\vartheta_{j_1,j_2}^M k(j_2/M)\left(\check{\Gamma}_{j_2}^{yx} - \Gamma_{j_2}^{yx}\right)'$. Using Lemma (B.7iv) we consider

$$\begin{aligned} E\left\|\check{\Gamma}_j^{xy} - \Gamma_j^{xy}\right\|^2 \leq\ & \left\|n^{-1}\sum\nolimits_{t=j+1}^{n-m} \check{w}_{t,j}\right\|^2 + \frac{(j+1+m)^2}{n^2}\left\|\Gamma_j^{xy}\right\|^2 \\ & + \frac{j+1+m}{n}\left\|\Gamma_j^{xy}\right\| \left\|n^{-1}\sum\nolimits_{t=j+1}^{n-m} \check{w}_{t,j}\right\| \end{aligned}$$

where $\frac{j+1+m}{n}\left\|\Gamma_j^{xy}\right\| = O(n^{-1})$ uniformly in j and

$$\text{(B.3)} \qquad E\left\|n^{-1}\sum\nolimits_{t=j+1}^{n-m} \check{w}_{t,j}\right\|^2 = n^{-2}\sum\nolimits_{t=j+1}^{n-m} \operatorname{tr}(\Gamma_{s-t}^{xx}\gamma_{s-t}^{yy} + \Gamma_{t-s+j}^{xy}\Gamma_{t-s-j}^{yx} + \mathcal{K}_4^4) = O(n^{-1})$$

uniformly in j. Then

$$\begin{aligned} & E\left\|\sum\nolimits_{j_1,j_2=1}^{[M]} \left(\check{\Gamma}_{j_1}^{xy} - \Gamma_{j_1}^{xy}\right)\vartheta_{j_1,j_2}^M \left(\check{\Gamma}_{-j_2}^{yx} - \Gamma_{-j_2}^{yx}\right)' |k(j_1/M)|\,|k(j_2/M)|\right\| \\ & \qquad \leq \sum\nolimits_{j_1,j_2=1}^{[M]} \left(E\left\|\check{\Gamma}_{j_1}^{xy} - \Gamma_{j_1}^{xy}\right\|^2\right)^{1/2} \left(E\left\|\check{\Gamma}_{-j_2}^{yx} - \Gamma_{-j_2}^{yx}\right\|^2\right)^{1/2} \|\vartheta_{j_1,j_2}^M\| \\ & \qquad \leq Cn^{-1}\sum\nolimits_{j_1,j_2=1}^{[M]} \|\vartheta_{j_1,j_2}\| = O(M/n). \end{aligned}$$

where C is some constant. ■

Lemma B.15. *Let H_{222} be defined in (A.9). Then $H_{222} = O_p(M/n^{1/2})$.*

Proof. Using (A.9) we write

$$\begin{aligned} H_{222} &= -\sum_{j_1,j_2=1}^{[M]} \check{\Gamma}^{xy}_{j_1} k(\frac{j_1}{M}) \vartheta^M_{j_1,j_2} k(\frac{j_2}{M}) \left(\check{\Gamma}^{yx}_{-j_2} - \Gamma^{yx}_{-j_2}\right)' \\ &\quad -\sum_{j_1,j_2=1}^{[M]} \left(\check{\Gamma}^{xy}_{j_1} - \Gamma^{xy}_{j_1}\right) k(\frac{j_1}{M}) \vartheta^M_{j_1,j_2} k(\frac{j_2}{M}) \check{\Gamma}^{yx}_{-j_2}. \end{aligned}$$

We use the same arguments as in the proof of the previous lemma to obtain the following bound

$$\begin{aligned} E\left\|\sum_{j_1,j_2=1}^{[M]} \check{\Gamma}^{xy}_{j_1} \vartheta_{j_1,j_2} \left(\check{\Gamma}^{yx}_{-j_2} - \Gamma^{yx}_{-j_2}\right)' k(\frac{j_1}{M}) k(\frac{j_2}{M})\right\| &\leq Cn^{-1/2} \sum_{j_1,j_2=1}^{[M]} \left(E\left\|\check{\Gamma}^{xy}_{j_2}\right\|^2\right)^{1/2} \left\|\vartheta_{j_1,j_2}\right\| \left|k(\tfrac{j_2}{M})\right| \\ &= O(M/n^{1/2}) \end{aligned}$$

where C is a constant such that $\left(E\left\|\check{\Gamma}^{xy}_{j_1} - \Gamma^{xy}_{j_1}\right\|^2\right)^{1/2} < Cn^{-1/2}$ uniformly in j_1 and $E\left\|\check{\Gamma}^{xy}_{j_2}\right\|^2 = O(1)$ uniformly in j_2. The result then follows from Markov's inequality. ■

Lemma B.16. *Let H_{31} be as defined in (A.10), H_{32} as defined in (A.11) and H_{33} as defined in (A.12). Then $H_{31} = o_p(M/n)$, $H_{32} = o_p(M/n)$ and $H_{33} = o_p(M/n)$.*

Proof. We use the fact that $\left\|\hat{\Gamma}^{xy}_j - \Gamma^{xy}_j\right\| = O_p(n^{-1/2})$ uniformly in j as shown in the proof of Lemmas (B.12) and (B.14) as well as the fact that the blocks of $\Omega_M - \hat{\Omega}_M$ are uniformly $O_p(n^{-1/2})$ and summable over one index as shown in Lemma (B.17). Then

$$\begin{aligned} \|H_{32}\| &\leq \sum_{j_1,\ldots,j_4=1}^{[M]} \left\|\hat{\Gamma}^{xy}_{j_1} - \Gamma^{xy}_{j_1}\right\| |k(j_1/M)| \left\|\vartheta^M_{j_1,j_2}\right\| \left\|\omega_{j_2,j_3} - \hat{\omega}_{j_2,j_3}\right\| \left\|\vartheta^M_{j_3,j_4}\right\| |k(j_4/M)| \left\|\hat{\Gamma}^{yx}_{-j_4} - \Gamma^{xy}_{-j_4}\right\|) \\ &\quad + \sum_{j_2,\ldots,j_4=1}^{[M]} \left\|a^M_{j_2}\right\| \left\|\omega_{j_2,j_3} - \hat{\omega}_{j_2,j_3}\right\| \left\|\vartheta^M_{j_3,j_4}\right\| |k(j_4/M)| \left\|\hat{\Gamma}^{yx}_{-j_4} - \Gamma^{xy}_{-j_4}\right\|. \end{aligned}$$

where $a^M_i = \sum_{j=0}^{[M]} \Gamma^{xy}_j k(j/M) \vartheta^M_{ji}$. The term in the first line is bounded by

$$\sum_{j_1=1}^{[M]} \left\|\hat{\Gamma}^{xy}_{j_1} - \Gamma^{xy}_{j_1}\right\| \sum_{j_1,\ldots,j_4=1}^{[M]} \left\|\vartheta^M_{j_1,j_2}\right\| \left\|\omega_{j_2,j_3} - \hat{\omega}_{j_2,j_3}\right\| \left\|\vartheta^M_{j_3,j_4}\right\| \left\|\hat{\Gamma}^{yx}_{-j_4} - \Gamma^{xy}_{-j_4}\right\|$$

where $\sum_{j_1=1}^{[M]} \left\|\hat{\Gamma}^{xy}_{j_1} - \Gamma^{xy}_{j_1}\right\| = O_p(M/n^{1/2})$ by the arguments in the proof of Lemmas (B.12) and (B.14) and

$$\begin{aligned} &E\sum_{j_1,\ldots,j_4=1}^{[M]} \left\|\vartheta^M_{j_1,j_2}\right\| \left\|\omega_{j_2,j_3} - \hat{\omega}_{j_2,j_3}\right\| \left\|\vartheta^M_{j_3,j_4}\right\| \left\|\hat{\Gamma}^{yx}_{-j_4} - \Gamma^{xy}_{-j_4}\right\| \\ &\leq E\sum_{j_1,\ldots,j_4=1}^{[M]} \left\|\vartheta^M_{j_1,j_2}\right\| \left(E\left\|\omega_{j_2,j_3} - \hat{\omega}_{j_2,j_3}\right\|^2\right)^{1/2} \left\|\vartheta^M_{j_3,j_4}\right\| \left(E\left\|\hat{\Gamma}^{yx}_{-j_4} - \Gamma^{xy}_{-j_4}\right\|^2\right)^{1/2} = O(M/n). \end{aligned}$$

The term on the second line is

$$\sum\nolimits_{j_2,..,j_4=1}^{[M]} \left\|a_{j_2}^M\right\| \left\|\omega_{j_2,j_3} - \hat{\omega}_{j_2,j_3}\right\| \left\|\vartheta_{j_3,j_4}^M\right\| |k(j_4/M)| \left\|\hat{\Gamma}_{-j_4}^{yx} - \Gamma_{-j_4}^{xy}\right\| = O_p(n^{-1})$$

such that the result follows. The arguments for H_{31} and H_{33} are identical. ■

Lemma B.17. *Let H_{34} be defined in (A.13). Then $H_{34} = O_p(n^{-1/2})$.*

Proof. First note that H_{34} can be written as $\sum_{j_1,j_2=1}^{[M]} a_{j_1}^M(\omega_{j_1,j_2} - \hat{\omega}_{j_1,j_2})a_{j_2}^{\prime M}$ with $a_i^M = \sum_{j=0}^{[M]} \Gamma_j^{xy} k(j/M)\vartheta_{ji}^M$. Note that $\left\|a_i^M\right\| \leq \sum_{j=1}^{\infty} \left\|\Gamma_j \vartheta_{ji}^M\right\|$ such that $\left\|a_i^M\right\|$ is summable $\forall M$. Then

$$\|H_{34}\| \leq \sum_{j_1,j_2=1}^{[M]} \left\|a_{j_1}^M\right\| \left\|a_{j_2}^M\right\| \left\|\omega_{j_1,j_2} - \hat{\omega}_{j_1,j_2}\right\|$$

Furthermore

$$\begin{aligned} \left\|\omega_{j_1,j_2} - \hat{\omega}_{j_1,j_2}\right\| \quad &\leq \quad \sum\nolimits_{l=-m+1}^{m-1} |\gamma_l^\varepsilon - \hat{\gamma}_l^\varepsilon| \left(\left\|\omega_{j_1,j_2}(l) - \hat{\omega}_{j_1,j_2}(l)\right\| + \left\|\omega_{j_1,j_2}(l)\right\|\right) \\ &\quad + \sum\nolimits_{l=-m+1}^{m-1} |\gamma_l^\varepsilon| \left\|\omega_{j_1,j_2}(l) - \hat{\omega}_{j_1,j_2}(l)\right\| \end{aligned}$$

such that by Lemma (B.8) $E\left\|\omega_{j_1,j_2}(l) - \hat{\omega}_{j_1,j_2}(l)\right\| = O(n^{-1/2})$ where the bound holds uniformly in j_1, j_2 and l. The result then follows immediately after standard probability manipulations and an application of the Markov inequality.

Lemma B.18. *Let H_4 be as defined in (A.14). Then $H_4 = o_p(M/n)$*

Proof. We use the matrix valued Taylor expansion $\hat{\Omega}_M^{-1} = \Omega_M^{-1} - \Omega_M^{-1}(\hat{\Omega}_M - \Omega_M)\Omega_M^{-1} + B + R$ where $R = o_p(\|B\|^2)$ and B can be expressed as $B = \Omega_M^{-1}(\Omega_M - \hat{\Omega}_M)\Omega_M^{-1}(\Omega_M - \hat{\Omega}_M)\Omega_M^{-1}$ to write $H_4 = H_{41} + H_{42}$ where

$$\begin{aligned} H_{41} &= \hat{P}_M' K_M B K_M \hat{P}_M \\ H_{42} &= \hat{P}_M' K_M R K_M \hat{P}_M. \end{aligned}$$

Further decompose $H_{41} = H_{411} + H_{412} + H_{413} + H_{414}$ where

$$\begin{aligned} H_{411} &\equiv -\left(\hat{P}_M - \check{P}_M\right)' K_M B K_M (\hat{P}_M - \check{P}_M) \\ H_{412} &\equiv \hat{P}_M' K_M B K_M (\hat{P}_M - \check{P}_M) + (\hat{P}_M - \check{P}_M)' K_M B K_M \hat{P}_M \\ H_{413} &\equiv -\left(\check{P}_M - P_M\right)' K_M B K_M' \left(\check{P}_M - P_M\right) \\ H_{414} &\equiv \check{P}_M' K_M B K_M (\check{P}_M - P_M) + (\check{P}_M - P_M)' K_M B K_M \check{P}_M \\ H_{415} &= P_M' K_M B K_M P_M. \end{aligned}$$

It follows immediately that $H_{411} = o_p(H_{412})$ such that we consider

$$\left\| \hat{P}_M' K_M B K_M (\hat{P}_M - \check{P}_M) \right\| \leq \sum_{j_1,\ldots,j_4}^{[M]} \hat{a}_{j_1} \left\| (\hat{\omega}_{j_1,j_2} - \omega_{j_1,j_2}) \right\| \left\| \vartheta_{j_2,j_3} \right\| \left\| (\hat{\omega}_{j_3,j_4} - \omega_{j_3,j_4}) \right\| \hat{b}_{j_4}'$$

where $\hat{a}_j = \sum_{k=1}^{[M]} \left\| \hat{\Gamma}_k^{xy} \right\| \left\| \vartheta_{k,j}^M \right\|$ and $\hat{b}_j = \sum_{k=1}^{[M]} \left\| \hat{\Gamma}_{-k}^{yx} - \check{\Gamma}_{-k}^{xy} \right\| \left\| \vartheta_{k,j}^M \right\|$. From the proof of Lemma (B.12) it follows that $\hat{b}_j = O_p(n^{-1/2})$ uniformly in j. Also

$$\left\| \hat{\Gamma}_k^{xy} \right\| \leq n^{-1} \sum_{t=1}^{n} \left\| x_t y_{t-k}' \right\| + \|\bar{x}\| \|\bar{y}\| + \|\bar{x}\| n^{-1} \sum_{t=1}^{n} \|y_{t-k}\| + \|\bar{y}\| n^{-1} \sum_{t=1}^{n} \|x_t\|$$

where the bound is uniformly $O_p(1)$ in k implying that $\hat{a}_j = O_p(1)$ uniformly in j. Since

$$\sum_{j_1,\ldots,j_4}^{[M]} \left(E \left\| (\hat{\omega}_{j_1,j_2} - \omega_{j_1,j_2}) \right\|^2 \right)^{1/2} \left\| \vartheta_{j_2,j_3} \right\| \left(E \left\| (\hat{\omega}_{j_3,j_4} - \omega_{j_3,j_4}) \right\|^2 \right)^{1/2} = O(M/n)$$

by the proof of Lemma (B.8) it follows that $H_{412} = O_p(M/n^{3/2}) = o_p(M/n)$.

Since $H_{413} = o_p(H_{414})$ we consider now

$$\left\| \check{P}_M' K_M B K_M (\check{P}_M - P_M) \right\| \leq \sup_j \check{a}_j \sum_k^{[M]} \check{b}_k' \sum_{j_1,\ldots,j_4}^{[M]} \left\| (\hat{\omega}_{j_1,j_2} - \omega_{j_1,j_2}) \right\| \left\| \vartheta_{j_2,j_3} \right\| \left\| (\hat{\omega}_{j_3,j_4} - \omega_{j_3,j_4}) \right\|$$

where $\check{a}_j = \sum_{k=1}^{[M]} \left\| \check{\Gamma}_k^{xy} \right\| \left\| \vartheta_{k,j}^M \right\|$ with $\sup_j \check{a}_j = O_p(1)$ by the same arguments as before and $\check{b}_j' = \sum_{k=1}^{[M]} \left\| \check{\Gamma}_{-k}^{xy} - \Gamma_{-k}^{yx} \right\| \left\| \vartheta_{k,j}^M \right\|$ such that $\sum_k^{[M]} \check{b}_k' = O_p(M/n^{1/2})$ since $E \left\| \check{\Gamma}_{-k}^{xy} - \Gamma_{-k}^{yx} \right\|^2 = O(n^{-1})$ uniformly in k as shown in the proof of Lemma (B.14). It follows that $H_{414} = O_p(M^2/n^{3/2}) = o_p(M/n)$ as long as $M/n^{1/2} \to 0$.

Finally

$$\|H_{415}\| \leq \sum_{j_1,\ldots,j_4}^{[M]} a_{j_1} \left\| (\hat{\omega}_{j_1,j_2} - \omega_{j_1,j_2}) \right\| \left\| \vartheta_{j_2,j_3} \right\| \left\| (\hat{\omega}_{j_3,j_4} - \omega_{j_3,j_4}) \right\| a_{j_4}'$$

with $a_j = \sum_{k=1}^{[M]} \left\| \Gamma_k^{xy} \right\| \left\| \vartheta_{k,j}^M \right\|$ such that a_j is absolutely summable. This implies that $H_{415} = O_p(n^{-1})$.

For H_{42} we can use the fact that $R = o_p\left(\|B\|^2 \right)$ by the matrix version of the Taylor expansion result (Gawronski, 1977). By the same arguments as for H_{41} it then follows that $H_{42} = o_p(M/n)$.

■

Lemma B.19. *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 using Lemma (B.7iii)

$$\begin{aligned}
Ed_0 d_0' &= \frac{1}{n} \sum_{t,s=1}^{n} \sum_{j_1,\ldots,j_4=1}^{\infty} \Gamma_{j_1}^{xy} \vartheta_{j_1,j_2} \gamma_{t-s}^{\varepsilon} \Gamma_{t-s-j_2+j_3}^{yy} \vartheta_{j_3,j_4} \Gamma_{-j_4}^{yx} \\
&= \sum_{l=-m+1}^{m-1} \frac{n-|l|}{n} \sum_{j_1,\ldots,j_4=1}^{\infty} \Gamma_{j_1}^{xy} \vartheta_{j_1,j_2} \gamma_{l}^{\varepsilon} \Gamma_{l+j_3-j_2}^{yy} \vartheta_{j_3,j_4} \Gamma_{-j_4}^{yx} \\
&\to P' \Omega^{-1} P \text{ as } n \to \infty.
\end{aligned}$$

where the second line follows from the fact that $\gamma_l^{\varepsilon} = 0$ for $l \geq m$. ■

Lemma B.20. *Let* d_1 *be defined in (A.16). Then* $Ed_1 d_1' = o(M^{-2s})$*.*

Proof. We write $d_1 = d_{11} + d_{12}$ where $d_{11} = P' \left(\Omega_M^{*-1} - \Omega^{-1} \right) V$ and $d_{12} = P_M' \Omega_M^{-1} V_M - P' \Omega_M^{*-1} V$ where Ω_M^* is defined in (B.1). Then by the same arguments as in the proof of Lemma (B.9) we can bound

$$\begin{aligned}
\|d_{11}\| &\leq \sum_{j_3 \geq [M]} \sum_{j_1,j_4} \left\| \Gamma_{j_1}^{xy} \vartheta_{j_1,j_3} \right\| \|\omega_{j_3,j_3} - I\| \left\| \vartheta_{j_3,j_4}^{M} \right\| \left\| n^{-1/2} \sum_{t=1}^{n} v_{t,j_4} \right\| \\
&\quad + \sum_{j_3 \geq [M]} \sum_{j_1,j_2,j_4} \left\| \Gamma_{j_1}^{xy} \vartheta_{j_1,j_2} \right\| \|\omega_{j_2,j_3}\| \left\| \vartheta_{j_3,j_4}^{M} \right\| \left\| n^{-1/2} \sum_{t=1}^{n} v_{t,j_4} \right\|
\end{aligned}$$

where $\left\| n^{-1/2} \sum_{t=1}^{n} v_{t,j_4} \right\| = O_p(1)$ uniformly in j_4 by Lemma (B.7iii) such that $\|d_{11}\| = o_p(M^{-2s})$ by the same arguments as in the proof of Lemma (B.9). For d_{12} consider

$$\begin{aligned}
E \|d_{12}\|^2 &= n^{-1} \sum_{t,s=1}^{n} \sum_{j_2,j_4 > [M]} \Gamma_{j_2}^{xy} E v_{t,j_2} v_{s,j_4} \Gamma_{-j_4}^{yx} \\
&= \sum_{j_2,j_4 > [M]} \Gamma_{j_2}^{xy} \omega_{j_2,j_4} \Gamma_{-j_4}^{yx} + O(n^{-1}) = o(M^{-2s}).
\end{aligned}$$

where the last equality follows from summability properties of Γ_j^{xy}. ■

Lemma B.21. *Let* d_2 *be defined in (A.17). Then* $Ed_2 d_2' = O(M^{-4q})$*.*

Proof. Consider

$$\begin{aligned}
Ed_2 d_2' &= \frac{1}{n} \sum_{t,s=1}^{[M]} \sum_{j_1,..,j_4=1}^{[M]} \Gamma_{j_1}^{xy} (1 - k(\frac{j_1}{M})) \vartheta_{j_1,j_2}^{M} (1 - k(\frac{j_2}{M})) E(v_{t,j_2} v_{s,j_3}') (1 - k(\frac{j_3}{M})) \vartheta_{j_3,j_4}^{M} (1 - k(\frac{j_4}{M})) \Gamma_{j_4}^{yx} \\
&= M^{-4q} \sum_{j_1,..,j_4=1}^{n} |j_1|^q \, \Gamma_{j_1}^{xy} \frac{(1 - k(\frac{j_1}{M}))}{|j_1/M|^q} \vartheta_{j_1,j_2}^{M} \frac{(1 - k(\frac{j_2}{M}))}{|j_2/M|^q} |j_2|^q \\
&\quad \times \sum_{l=-m+1}^{m-1} \tfrac{n-|l|}{n} \gamma_l^{\varepsilon} \Gamma_{l-j_2+j_3}^{yy} |j_3|^q \frac{(1 - k(\frac{j_3}{M}))}{|j_3/M|^q} \vartheta_{j_3,j_4}^{M} \frac{(1 - k(\frac{j_2}{M}))}{|j_4/M|^q} |j_4|^q \, \Gamma_{j_4}^{yx}.
\end{aligned}$$

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

$$\left\|Ed_2 d_2'\right\| \leq C_1 M^{-4q} \left(\sum_{j_1,j_2=0}^{n} |j_1|^q |j_2|^q \left\|\Gamma_{j_1}^{xy} \vartheta_{j_1,j_2}^M\right\| \right)^2$$

where $\sum_{j_1,j_2=1}^{n} |j_1|^q |j_2|^q \left\|\Gamma_{j_1}^{xy} \vartheta_{j_1,j_2}^M\right\| = O(1)$ by arguments similar to the proof of Lemma 5.2. in Kuersteiner (1999b). Thus $Ed_2 d_2' = O(M^{-4q})$. ■

Lemma B.22. *Let d_3 be defined in (A.18). Then $d_3 = O_p(M^{-q})$.*

Proof. Write $d_3 = d_{31} + d_{32}$ where

$$d_{31} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sum_{j_1,j_2=1}^{[M]} \Gamma_{j_1}^{xy} \left(1 - k(j_1/M)\right) \vartheta_{j_1,j_2}^M k(j_2/M) v_{t,j_2} \tag{B.4}$$

and

$$d_{32} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sum_{j_1,j_2=1}^{[M]} \Gamma_{j_1}^{xy} k(j_1/M) \vartheta_{j_1,j_2}^M \left(1 - k(j_2/M)\right) v_{t,j_2} \tag{B.5}$$

such that

$$M^q E \left\|d_{31}\right\| \leq \sum_{j_1,j_2=1}^{[M]} \left\|\Gamma_{j_1}^{xy}\right\| \left\|\vartheta_{j_1,j_2}^M\right\| \left|\frac{1-k(j_1/M)}{|j_1/M|^q}\right| E \left\|\frac{1}{\sqrt{n}} \sum\nolimits_{t=1}^{n} v_{t,j_2}\right\| = O(1).$$

and the result follows from the Markov inequality. The same arguments apply to d_{32}. ■

Lemma B.23. *Let d_4 be defined in (A.19). Then $d_4 = O_p(M/n)$.*

Proof. First note that $\hat{\Gamma}_{j_1}^{xy} - \check{\Gamma}_{j_1}^{xy} = \frac{n-m+j_1}{n}(\bar{x} - \mu_x)(\bar{y} - \mu_y)' + (\bar{x} - \mu_x) n^{-1} \sum_{t=j+1}^{n-m} (y_t - \bar{y})' + n^{-1} \sum_{t=j+1}^{n-m} (x_t - \bar{x})(\bar{y} - \mu_y)'$ such that d_4 can be analyzed by considering

$$\begin{aligned}
d_{41} &= (\bar{x} - \mu_x)(\bar{y} - \mu_y)' n^{-1/2} \sum\nolimits_{t=1}^{n} \sum\nolimits_{j_1,j_2=1}^{n} k(j_1/M) \vartheta_{j_1,j_2}^M k(j_2/M) v_{t,j_2}, \\
d_{42} &= (\bar{x} - \mu_x) n^{-1} \sum_{t=j+1}^{n-m} (y_t - \bar{y})' n^{-1/2} \sum\nolimits_{t=1}^{n} \sum\nolimits_{j_1,j_2=1}^{n} k(j_1/M) \vartheta_{j_1,j_2}^M k(j_2/M) v_{t,j_2}
\end{aligned}$$

and similarly for d_{43}. Then

$$\left\|d_{41}\right\| \leq \left\|\bar{x} - \mu_x\right\| \left\|\bar{y} - \mu_y\right\| \left\|n^{-1/2} \sum\nolimits_{t=1}^{n} \sum\nolimits_{j_1,j_2=1}^{n} k(j_1/M) \vartheta_{j_1,j_2}^M k(j_2/M) v_{t,j_2}\right\|$$

The third term in the previous display can be bounded in expectation by

$$\frac{1}{n}\sum\nolimits_{t,s=1}^{n}\sum\nolimits_{j_1,..,j_4=1}^{[M]} k(\frac{j_1}{M})\vartheta^M_{j_1,j_2}k(\frac{j_2}{M})E\left(v_{t,j_2}v'_{s,j_3}\right)k(\frac{j_3}{M})\vartheta^M_{j_3,j_4}k(\frac{j_4}{M})$$
$$\leq \sum\nolimits_{j_1,..,j_4=1}^{[M]}\left\|\vartheta^M_{j_1,j_2}\right\|\left\|\omega_{j_2,j_3}\right\|\left\|\vartheta^M_{j_3,j_4}\right\| + o(1) = O(M^2)$$

and using $\|\bar{x}-\mu_x\|\left\|\bar{y}-\mu_y\right\| = O_p(n^{-1})$ shows that $d_{41} = O_p(M/n)$. For d_{42} write

$$\|d_{42}\| \leq \|\bar{x}-\mu_x\|\left(\|\bar{y}\| + n^{-1}\sum_{t=1}^{n}\|y_t\|\right)\left\|n^{-1/2}\sum\nolimits_{t=1}^{n}\sum\nolimits_{j_1,j_2=1}^{[M]}k(j_1/M)\vartheta^M_{j_1,j_2}k(j_2/M)v_{t,j_2}\right\|$$

where we use $\left\|n^{-1}\sum_{t=j_1+1}^{n-m}(y_t-\bar{y})\right\| \leq \|\bar{y}\| + n^{-1}\sum_{t=1}^{n}\|y_t\|$. Then the third term in the previous display is bounded in expectation by

$$n^{-1/2}\sum\nolimits_{t=1}^{n}\sum\nolimits_{j_1,j_2=1}^{[M]}\left\|k(j_1/M)\vartheta^M_{j_1,j_2}k(j_2/M)\right\|\left(E\left\|v_{t,j_2}\right\|^2\right)^{1/2} = O(M/\sqrt{n})$$

such that d_{42} is $O_p(M/n)$. Finally d_{43} can be analyzed in the same way as d_{42}. ■

Lemma B.24. *Let d_5 be defined (A.20). Then $d_5 = O_p(M/\sqrt{n})$.*

Proof. We consider

$$\begin{aligned} E\|d_5\| \quad \leq \quad & \sum\nolimits_{j_1,j_2=1}^{[M]}\left(E\left\|n^{-1}\sum\nolimits_{t=1+j_1}^{n-m}\left(w_{t,j_1}-\Gamma^{xy}_{j_1}\right)k(\tfrac{j_1}{M})\vartheta^M_{j_1,j_2}k(\tfrac{j_2}{M})\right\|^2 E\left\|n^{-1/2}\sum\nolimits_{t=1}^{n}v_{t,i}\right\|^2\right)^{1/2} \\ & + \sum\nolimits_{j_1,j_2=1}^{[M]}\left|\tfrac{j_1+m}{n}\right|\left\|\Gamma^{xy}_{j_1}\vartheta^M_{j_1,j_2}\right\|^2\left(E\left\|n^{-1/2}\sum\nolimits_{t=1}^{n}v_{t,i}\right\|^2\right)^{1/2}. \end{aligned}$$

where the second term is of lower order. Then

$$\begin{aligned} E\left\|n^{-1/2}\sum\nolimits_{t=1}^{n}v_{t,i}\right\|^2 \quad &= \quad \operatorname{tr}E(\frac{1}{n}\sum\nolimits_{s,t=1}^{n}v_{t,i}v'_{s,i}) \\ &= \quad \sum\nolimits_{l=1-m}^{m-1}\tfrac{n-|l|}{n}\gamma^{\varepsilon}_l\operatorname{tr}\left[\Gamma^{yy}_l\right] \leq 2m\sup_j\left|\operatorname{tr}\left[\Gamma^{yy}_j\right]\right|\sup_j\left|\gamma^{\varepsilon}_j\right| \end{aligned}$$

and by Lemma (B.7iv) we have

$$\begin{aligned} E\left\|\left(\check{\Gamma}^{xy}_{j_1}-\Gamma^{xy}_{j_1}\right)\vartheta^M_{j_1,j_2}\right\|^2 &\leq \left\|\vartheta^M_{j_1,j_2}\right\|^2 n^{-2}\operatorname{tr}\sum\nolimits_{t,s=1+j_1}^{n-m}E\left(w_{t,j_1}-\Gamma^{xy}_{j_1}\right)'(w_{s,i}-\Gamma^{xy}_{j_1}) \\ &= \left\|\vartheta^M_{j_1,j_2}\right\|^2 n^{-2}\operatorname{tr}\sum\nolimits_{t,s=1+j_1}^{n-m}\left(\Gamma_{s-t+j_1}\Gamma'_{t-s+j_1}+\gamma_{t-s}\Gamma^{xx'}_{t-s}+\mathcal{K}^4_4\right) \end{aligned}$$

where $n^{-2}\operatorname{tr}\sum_{t,s=1}^{n}\left(\Gamma_{s-t+j_1}\Gamma'_{t-s+j_1}+\gamma_{t-s}\Gamma^{xx'}_{t-s}+\mathcal{K}^4_4\right) = O(n^{-1})$ uniformly in j_1 by the Toeplitz Lemma. Summability of $\left\|\vartheta^M_{j_1,j_2}\right\|^2$ over j_1 shows that $E\|d_5\| = O(M/\sqrt{n})$. ■

Lemma B.25. *Let d_6 be defined in (A.21). Then $d_6 = O_p(M/n)$.*

Proof. For d_6 we define the terms

$$\begin{aligned} d_{61} &= \left(\hat{P}_M - \check{P}_M\right)' K_M \Omega_M^{-1}(\Omega_M - \hat{\Omega}_M)\Omega_M^{-1} K_M V_M \\ d_{62} &= \left(\check{P}_M - P_M\right)' K_M \Omega_M^{-1}(\Omega_M - \hat{\Omega}_M)\Omega_M^{-1} K_M V_M \end{aligned}$$

such that $d_6 = d_{61} + d_{62}$.

For d_{61} define

$$\begin{aligned} \hat{a}_{j_4} &= \sum_{j_1,..,j_3} \left\| \left(\hat{\Gamma}_{j_1}^{xy} - \check{\Gamma}_{j_1}^{xy}\right) \vartheta_{j_1,j_2}^M \left(\hat{\omega}_{j_2,j_3} - \omega_{j_2,j_3}\right) \vartheta_{j_3,j_4}^M \right\| \\ &\leq \sum_l |\hat{\gamma}_\varepsilon(l)| \sum_{j_1,..,j_3} \left\| \hat{\Gamma}_{j_1}^{xy} - \check{\Gamma}_{j_1}^{xy} \right\| \|\hat{\omega}_{j_2,j_3}(l) - \omega_{j_2,j_3}(l)\| \left\|\vartheta_{j_1,j_2}^M\right\| \left\|\vartheta_{j_3,j_4}^M\right\| = O_p(n^{-1}) \end{aligned}$$

uniformly in j_4 such that

$$\|d_{61}\| \leq \sum_{j_4}^{[M]} a_{j_4} \left\| n^{-1/2} \sum\nolimits_{t=1}^n v_{t,j_4} \right\| = O_p(M/n).$$

For d_{62} we use the bound

$$\|d_{62}\| = \sum_l |\hat{\gamma}_\varepsilon(l)| \sum_{j_1,..,j_4}^{[M]} \left\| \check{\Gamma}_{j_1}^{xy} - \Gamma_{j_1}^{xy} \right\| \|\hat{\omega}_{j_2,j_3}(l) - \omega_{j_2,j_3}(l)\| \left\|\vartheta_{j_1,j_2}^M\right\| \left\|\vartheta_{j_3,j_4}^M\right\| \left\| n^{-1/2} \sum\nolimits_{t=1}^n v_{t,j_4} \right\|$$

where $\left\| \check{\Gamma}_{j_1}^{xy} - \Gamma_{j_1}^{xy} \right\| \leq \left\| n^{-1} \sum_{t=j_1}^{n-m} \check{w}_{t,j_1} \right\| + \left\| \Gamma_{j_1}^{xy} \right\|$ with $\check{w}_{t,j_1} = w_{t,j_1} - \Gamma_{j_1}^{xy}$. The terms involving $\Gamma_{j_1}^{xy}$ are of lower order. By the Markov inequality we consider

(B.6)
$$\sum_{j_1,..,j_4}^{[M]} \left(E \left\| n^{-1} \sum\nolimits_{t=j_1}^{n-m} \check{w}_{t,j_1} \right\|^2 \left\| n^{-1/2} \sum\nolimits_{t=1}^n v_{t,j_4} \right\|^2 \right)^{1/2} \left(E \|\hat{\omega}_{j_2,j_3}(l) - \omega_{j_2,j_3}(l)\|^2 \right)^{1/2} \left\|\vartheta_{j_1,j_2}^M\right\| \left\|\vartheta_{j_3,j_4}^M\right\|.$$

Note that

$$\begin{aligned} & E \left\| n^{-1} \sum\nolimits_{t=j_1}^{n-m} \check{w}_{t,j_1} \right\|^2 \left\| n^{-1/2} \sum\nolimits_{t=1}^n v_{t,j_4} \right\|^2 \\ &= \operatorname{tr} E \left[n^{-2} \sum\nolimits_{t_1,t_2=j_1}^{n-m} \check{w}_{t_1,j_1} \check{w}_{t_2,j_1}' \otimes n^{-1} \sum\nolimits_{s_1,s_2=1}^n v_{s_1,j_4} v_{s_2,j_4}' \right]. \end{aligned} \tag{B.7}$$

Let $\breve{w}^{i_1,i_2}_{t_1,j_1}$ be the i_1,i_2-th element of $\breve{w}^{i_1,i_2}_{t_1,j_1}$ such that a typical element of $E\left[\breve{w}_{t_1,j_1}\breve{w}'_{t_2,j_1}\otimes v_{s_1,j_4}v'_{s_2,j_4}\right]$ can be written as $\sum_{i_2=1}^{p}E\left[\breve{w}^{i_1,i_2}_{t_1,j_1}\breve{w}^{i_2,i_3}_{t_2,j_1}v^{i_4}_{s_1,j_4}v^{i_5}_{s_2,j_4}\right]$. Since $\breve{w}^{i_k,i_l}_{t_1,j_1}$ and $v^{i_r}_{t,j_4}$ have zero mean

$$
\begin{aligned}
E\left[\breve{w}^{i_1,i_2}_{t_1,j_1}\breve{w}^{i_2,i_3}_{t_2,j_1}v^{i_4}_{s_1,j_4}v^{i_5}_{s_2,j_4}\right] &= E\left[\breve{w}^{i_1,i_2}_{t_1,j_1}\breve{w}^{i_2,i_3}_{t_2,j_1}\right]E\left[v^{i_4}_{s_1,j_4}v^{i_5}_{s_2,j_4}\right]+E\left[\breve{w}^{i_1,i_2}_{t_1,j_1}v^{i_5}_{s_2,j_4}\right]E\left[\breve{w}^{i_2,i_3}_{t_2,j_1}v^{i_4}_{s_1,j_4}\right]\\
&\quad +E\left[\breve{w}^{i_2,i_3}_{t_2,j_1}v^{i_5}_{s_2,j_4}\right]E\left[\breve{w}^{i_1,i_2}_{t_1,j_1}v^{i_4}_{s_1,j_4}\right]+\operatorname{cum}^*\left(\breve{w}^{i_2,i_3}_{t_2,j_1},v^{i_5}_{s_2,j_4},\breve{w}^{i_1,i_2}_{t_1,j_1},v^{i_4}_{s_1,j_4}\right).
\end{aligned}
$$

By Lemma (B.7) $E\left[\breve{w}^{i_1,i_2}_{t_1,j_1}\breve{w}^{i_2,i_3}_{t_2,j_1}\right]$ is the i_1,i_3 element of $\gamma^{yy}_{t_1-t_2}\Gamma^{xx}_{t_1-t_2}+\Gamma^{xy}_{t_1-j_1-t_2}\Gamma^{yx}_{t_1-t_2+j_1}+\mathcal{K}^4_4$, $E\left[v^{i_4}_{s_1,j_4}v^{i_5}_{s_2,j_4}\right]$ is the i_4,i_5 element of $\gamma^{\varepsilon}_{s_1-s_2}\Gamma^{yy}_{s_1-s_2}+\mathcal{K}^3_4$ and $E\left[\breve{w}^{i_1,i_2}_{t_1,j_1}v^{i_5}_{s_2,j_4}\right]$ is the i_1,i_5 element of $(e'_{i_2}\Gamma^{\varepsilon x}_{t_1-s_2})\Gamma^{yy}_{t_1-s_2+j_1-j_4}+\Gamma^{\varepsilon y}_{t_1-s_2+j_1}(e'_{i_2}\Gamma^{xy}_{s_2-t_1+j_4})+\mathcal{K}^2_4$ where e_{i_2} is the i_2 unit vector. The largest of these terms are $n^{-1}\sum_{t_1,t_2=j_1}^{n-m}\gamma^{yy}_{t_1-t_2}\Gamma^{xx}_{t_1-t_2}=O(1)$ and $n^{-1}\sum_{s_1,s_2=1}^{n}\gamma^{\varepsilon}_{s_1-s_2}\Gamma^{yy}_{s_1-s_2}=O(1)$ while all the higher order cumulant terms are of lower order by Assumption (B). This shows that (B.7) is $O(n^{-1})$ and (B.6) is $O(M/n)$ which establishes that $d_{62}=O_p(M/n)$. ■

Lemma B.26. *Let d_7 be defined in (A.22). Then $d_7=O_p(n^{-1/2})$.*

Proof. We bound d_7 by

$$
\|d_7\|\leq\sum_l|\hat{\gamma}_\varepsilon(l)|\sum_{j_1,\ldots,j_4=1}^{[M]}\left\|\Gamma^{xy}_{j_1}\vartheta^M_{j_1,j_2}\right\|\left\|\vartheta^M_{j_3,j_4}\right\|\left\|\hat{\omega}_{j_2j_3}(l)-\omega_{j_2j_3}(l)\right\|\left\|n^{-1/2}\sum\nolimits_{t=1}^{n}v_{t,j_4}\right\|.
$$

Since $E\left\|n^{-1/2}\sum_{t=1}^{n}v_{t,j_4}\right\|^2=O(1)$ and $E\left\|\hat{\omega}_{j_2j_3}(l)-\omega_{j_2j_3}(l)\right\|^2=c_{j_3}n^{-1}$ where c_{j_3} is an absolutely summable sequence such that $E\|d_7\|=O(n^{-1/2})$. ■

Lemma B.27. *Let d_8 be defined as in (A.23). Then $d_8=O_p(M/n)$.*

Proof. The matrix B can be expressed more explicitly as $\Omega_M^{-1}(\Omega_M-\hat{\Omega}_M)\Omega_M^{-1}(\Omega_M-\hat{\Omega}_M)\Omega_M^{-1}$ such that by the similar arguments as in the proof of Lemmas (B.18) and (B.25) it follows that $\hat{P}'_MK_MBK_MV_M=O_p(M/n)$. From the Taylor expansion result for matrix valued functions (Gawronski, 1977) it then follows that the remainder term is $o_p(M/n)$. ■

Lemma B.28. *Let d_9 be defined in (A.24). Then $d_9=O_p(M/n^{3/2})$.*

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

$$
\begin{aligned}
d_{91} &= n^{-1/2}\sum_t\varepsilon_t\left(\hat{P}-P_M\right)K_M\Omega_M^{-1}(\Omega_M-\hat{\Omega}_M)\Omega_M^{-1}K_M\left[\mathbf{1}_{[M]}\otimes(\bar{y}-\mu_y)\right]\\
d_{92} &= n^{-1/2}\sum_t\varepsilon_t\left(\hat{P}-P_M\right)K_MBK_M\left[\mathbf{1}_{[M]}\otimes(\bar{y}-\mu_y)\right]+o_p(M/n^2)\\
d_{93} &= n^{-1/2}\sum_t\varepsilon_tP_MK_M\Omega_M^{-1}(\Omega_M-\hat{\Omega}_M)\Omega_M^{-1}K_M\left[\mathbf{1}_{[M]}\otimes(\bar{y}-\mu_y)\right]\\
d_{94} &= n^{-1/2}\sum_t\varepsilon_tP_MK_MBK_M\left[\mathbf{1}_{[M]}\otimes(\bar{y}-\mu_y)\right]+o_p(n^{-1})
\end{aligned}
$$

such that $d_9 = d_{91} + d_{92} + d_{93} + d_{94}$. By the same arguments as in the proof of Lemma (B.25) it follows that $d_{91} = O_p(M/n^{3/2})$ which implies that $d_{92} = o_p(M/n^{3/2})$. Similarly it follows that $d_{93} = O_p(n^{-1})$ by similar arguments as in the proof of Lemma (B.26). Consequently, $d_{94} = o_p(n^{-1})$. ■

Lemma B.29. *Let d_0 be defined in (A.15) and H_{12} as defined in (A.3). Then $Ed_0d_0'D^{-1}H_{12} = M^{-2q}k_q^2\mathcal{B}_0^{(q)} + o(M^{-2q})$*

Proof. The result follows immediately from Lemmas (B.10) and (B.19). ■

Lemma B.30. *Let d_0 be as defined in (A.15) and d_2 as defined in (A.17). Then $Ed_0d_2 = M^{-2q}k_q^2\mathcal{B}_0^{(q)} + o(M^{-2q})$.*

Proof. Directly evaluate

$$\begin{aligned}
Ed_0d_2' &= \frac{1}{n}\sum\nolimits_{t,s=1}^{n}\sum\nolimits_{j_1,j_2=1}^{\infty}\sum\nolimits_{j_3,j_4=1}^{[M]}\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}E(v_{t,j_2}v_{s,j_3}')(1-k(\frac{j_3}{M}))\vartheta_{j_3,j_4}^{M}(1-k(\frac{j_4}{M}))\Gamma_{-j_4}^{yx} \\
&= M^{-2q}k_q^2\sum\nolimits_{j_1,j_2=1}^{\infty}\sum\nolimits_{j_3,j_4=1}^{[M]}\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}\omega_{j_2,j_3}\left|j_3\right|^q\vartheta_{j_3,j_4}^{M}\left|j_4\right|^q\Gamma_{-j_4}^{xy} + o(n^{-1}) \\
&= M^{-2q}k_q^2\sum\nolimits_{j_1,j_2=1}^{\infty}\Gamma_{j_1}^{xy}\left|j_1\right|^q\vartheta_{j_1,j_2}\left|j_2\right|^q\Gamma_{-j_2}^{xy} + o(M^{-2q}) \\
&= M^{-2q}k_q^2\mathcal{B}_0^{(q)} + o(M^{-2q}).
\end{aligned}$$

where $\mathcal{B}_0^{(q)}$ is defined in the proof of Lemma (B.10) and we have used the Toeplitz Lemma for the second equality and dominated convergence for the third equality. ■

Lemma B.31. *Let d_0 be as defined in (A.15) and d_3 as defined in (A.18). Then $Ed_0d_3 = -M^{-q}k_q\mathcal{B}_1^{(q)} + o(M^{-q})$.*

Proof. Directly evaluate

$$\begin{aligned}
Ed_0d_3' &= -\frac{1}{n}\sum\nolimits_{t,s=1}^{n}\sum\nolimits_{j_1,j_2=1}^{\infty}\sum\nolimits_{j_3,j_4=1}^{[M]}\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}E(v_{t,j_2}v_{s,j_3}') \\
&\quad\times\left[(1-k(\frac{j_3}{M}))\vartheta_{j_3,j_4}^{M}k(\frac{j_4}{M}) + k(\frac{j_3}{M}))\vartheta_{j_3,j_4}^{M}(1-k(\frac{j_4}{M}))\right]\Gamma_{-j_4}^{yx} \\
&= -M^{-q}k_q\sum\nolimits_{j_1,j_2,j_3,j_4=1}^{\infty}\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}\omega_{j_2,j_3}\left[\left|j_3\right|^q\vartheta_{j_3,j_4} + \vartheta_{j_3,j_4}\left|j_4\right|^q\right]\Gamma_{-j_4}^{yx} + o(M^{-q}) \\
&= -M^{-q}k_q\sum_{j_1,j_2=1}^{\infty}\left(\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}\left|j_2\right|^q\Gamma_{-j_2}^{yx} + \Gamma_{j_1}^{xy}\left|j_1\right|^q\vartheta_{j_1,j_2}\Gamma_{-j_2}^{yx}\right) + o(M^{-q}).
\end{aligned}$$

The second equality uses Lemma (B.21) to replace $k(\frac{j_4}{M})$ by 1. ■

Lemma B.32. *Let H_{222} be defined in (A.9). Then*

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

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

$$\text{(B.8)} \qquad n^{-1} \sum\nolimits_{j_1,j_2=1}^{[M]} \sum\nolimits_{t=1+j_1}^{n-m} \breve{w}_{t,j_1} k(j_1/M) \vartheta^M_{j_1,j_2} k(j_2/M) \Gamma^{yx}_{-j_2} + o_p(M/n^{-1/2}).$$

Next define $a_j = \sum_{i=1}^{\infty} \Gamma^{xy}_i \vartheta_{i,j}$. Then, using only the dominant term in (B.8),

$$\begin{aligned} &EH_{222} D^{-1} d_0 d_0 D^{-1} H'_{222} \\ &\qquad = n^{-3} \sum_{t_1,t_2,s_1,s_2,j_1,\ldots,j_4} \prod_{l=1}^{4} k(\tfrac{j_l}{M}) E\left[\breve{w}_{s_1,j_1} a'_{j_1} D^{-1} a_{j_2} v_{t_1,j_2} v'_{t_2,j_3} a'_{j_3} D^{-1} a_{j_4} \breve{w}_{s_2,j_4}\right] + o(n^{-1}) \end{aligned}$$

Using the same arguments as in the proof of Lemma (B.36) it follows that the leading term in $E\ell' H_{222} D^{-1} d_0 d_0 D^{-1} H'_{222} \ell$ depends on

$$\mathcal{A}_2 = \sum\nolimits_{j_1 j_2=1}^{\infty} \sum\nolimits_{h=-\infty}^{\infty} vec(a'_{j_2} D^{-1} a_{j_1})' \left[\left(vec\Gamma^{yy}_{h+j_1-j_2} \otimes \Gamma^{\varepsilon x\prime}_h\right) + K_{pp}(\Gamma^{\varepsilon y}_{h+j_1} \otimes \Gamma^{yx}_{h-j_2})\right] (I \otimes \ell)$$

where $\mathcal{A}_2$ is well defined due to the summability properties of a_j. The result follows. ■

Lemma B.33. *Let H_{222} be as defined in Lemma (A.9) and d_5 as defined in Lemma (A.20). Then $E\ell' H_{222} D^{-1} d_0 d'_5 \ell = O(M/n)$.*

Proof. We again replace H_{222} by $n^{-1} \sum_{j_1,j_2=1}^{[M]} \sum_{t=1+j_1}^{n-m} \breve{w}_{t,j_1} k(j_1/M) \vartheta^M_{j_1,j_2} k(j_2/M) \Gamma^{yx}_{-j_2}$ and consider

$$E\ell' H_{222} D^{-1} d_0 d'_5 \ell = n^{-3} \sum_{t_1,t_2,s_1,s_2,j_1,\ldots,j_4} E \prod_{l=1}^{4} k(\tfrac{j_l}{M}) \ell' \bar{\omega}_{s_1,j_1} a'_{j_1} D^{-1} a_{j_2} v_{t_1,j_2} v'_{t_2,j_3} \vartheta^M_{j_3,j_4} \bar{\omega}'_{s_2,j_4} \ell$$

where a_j is defined in Lemma (B.32). 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 (B.32) and $\mathcal{A}_1$ is defined in (3.4). ■

B.3. Lemmas for Proposition 3.3

Lemma B.34. *Let $H_{13}$and H_{14} be defined in (A.4) and (A.5). Then* $H_{13}+H_{14} = -M^{-q}k_q\mathcal{B}_1^{(q)} + o(M^{-q})$ *where* $\mathcal{B}_1^{(q)} = \sum_{j_1,j_2=1}^{\infty}\left(\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}\left|j_2\right|^q\Gamma_{-j_2}^{yx} + \Gamma_{j_1}^{xy}\left|j_1\right|^q\vartheta_{j_1,j_2}\Gamma_{-j_2}^{yx}\right)$.

Proof. Write $H_{13} = -M^{-q}\sum_{j_1,j_2=1}^{[M]}\Gamma_{j_1}^{xy}k(j_1/M)\vartheta_{j_1,j_2}^{M}\frac{(1-k(j_2/M))}{|j_2/M|^q}\left|j_2\right|^q\Gamma_{-j_2}^{yx}$ such that

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

where in the second equality we use Assumption (F) about the kernel function. ■

Lemma B.35. *Let d_3 be defined in (A.18). Then* $Ed_3d_3 = k_q^2\mathcal{B}_2^{(q)} + o(M^{-2q})$.

Proof. Let d_{31} and d_{32} be defined in Equations (B.4) and (B.5). Then

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

and

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

Finally we consider the cross-product

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

■

Lemma B.36. *Let d_5 as defined in (A.20). Then* $E\ell'd_5d_5'\ell = M^2/n\left(\int_{-\infty}^{\infty}k(x)^2dx\right)^2\mathcal{A}_1\ell\ell'\mathcal{A}_1' + o(M^2/n)$ *where $\mathcal{A}_1$ is defined in (3.4).*

Proof. Using the arguments of the proof of Lemma (B.24) we only consider

$$d_{51} = n^{-3/2}\sum\nolimits_{j_1,j_2=1}^{[M]}\sum\nolimits_{t=1+j_1}^{n-m}\check{w}_{t,j_1}k(\tfrac{j_1}{M})\vartheta_{j_1,j_2}^{M}k(\tfrac{j_2}{M})\sum\nolimits_{t=1}^{n}v_{t,i}$$

where $\breve{w}_{t,j_1} = w_{t,j_1} - \Gamma^{xy}_{j_1}$. By Lemma (B.5) we have

$$E\ell' d_{51} d_{51}' \ell = \frac{1}{n^3} \sum_{t_1,t_2} \sum_{j_1,..,j_4}^{[M]} \sum_{s_1,s_2}^{n-m} \prod_{l=1}^{4} k(\tfrac{j_l}{M}) \left\{ \left(\operatorname{vec} \vartheta^{M\prime}_{j_1,j_2}\right)' E(v_{t_1,j_2} \otimes \breve{w}'_{s_1,j_1})(\ell' \otimes \ell) E(\breve{w}_{s_2,j_4} \otimes v'_{t_2,j_3}) \operatorname{vec} \vartheta^M_{j_3,j_4} \right.$$
$$+ \operatorname{tr}\left[\vartheta^M_{j_1,j_2} E(v_{t_1,j_2} \ell' \breve{w}_{s_2,j_4}) \vartheta^M_{j_3,j_4} E(v_{t_2,j_3} \ell' \breve{w}_{s_1,j_1})\right]$$
$$\left. + \operatorname{tr}\left[\vartheta^M_{j_1,j_2} E(v_{t_1,j_2} v_{t_2,j_3}) \vartheta^M_{j_3,j_4} E(\breve{w}'_{s_1,j_1} \ell\ell' \breve{w}_{s_2,j_4})\right] \right\} + \ell' \mathcal{K}_8 \ell$$

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

$$\operatorname{cum}^* \left(\breve{w}^{i_1,i_2}_{t_1,j_1}, \breve{w}^{i_2,i_3}_{t_2,j_1}, v^{i_5}_{s_2,j_4}, v^{i_4}_{s_1,j_4} \right)$$

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

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

where

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

by Lemma (B.7i) such that

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

Using arguments based on Parzen (1957) it can be shown that

$$\frac{1}{nM} \sum_{t_1,s_1} \sum_{j_1,j_2}^{[M]} \prod_{l=1}^{2} k(\tfrac{j_l}{M}) \left(\operatorname{vec} \vartheta^{M\prime}_{j_1,j_2}\right)' E(v_{t_1,j_2} \otimes \breve{w}'_{s_1,j_1})$$
$$\to (2\pi)^2 \int k(x)^2 dx \int \left(\operatorname{vec} f_\Omega^{-1}(\lambda)\right)' \left(\operatorname{vec}\left(f_{yy}(\lambda)'\right) \otimes f_{\varepsilon x}(\lambda)'\right) d\lambda \text{ as } n, M \to \infty.$$

which establishes the first part of (3.4). Next turn to

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

which follows from Lemma (B.7ii) where for a typical term in this product we have

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

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

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

Similar arguments show that the second and third terms of $E\ell d_5 d_5' \ell$ are both $O(M/n)$. ■

B.4. Lemmas for Theorem 4.2

Lemma B.37. *Let d_{10} be defined in (A.31). Then $d_{10} = o_p(1)$.*

Proof. Write $d_{10} = \sqrt{n/M^*}(d_{101} + d_{102} + d_{103})$ where

$$\begin{aligned}
d_{101} &= \left(\hat{P}_{n-m} - \check{P}_{n-m}\right)' \tilde{K}_{M^*}\Omega^{*-1}_{M^*}(\hat{\Omega}^*_{M^*} - \hat{\Omega}^*_{\hat{M}^*})\Omega^{*-1}_{M^*}\tilde{K}_{M^*}\frac{Z'_{n-m}\varepsilon}{\sqrt{n}} \\
d_{102} &= \left(\check{P}_{n-m} - P_{n-m}\right)' \tilde{K}_{M^*}\Omega^{*-1}_{M^*}(\hat{\Omega}^*_{M^*} - \hat{\Omega}^*_{\hat{M}^*})\Omega^{*-1}_{M^*}\tilde{K}_{M^*}\frac{Z'_{n-m}\varepsilon}{\sqrt{n}} \\
d_{103} &= P'_{n-m}\tilde{K}_{M^*}\Omega^{*-1}_{M^*}(\hat{\Omega}^*_{M^*} - \hat{\Omega}^*_{\hat{M}^*})\Omega^{*-1}_{M^*}\tilde{K}_{M^*}\frac{Z'_{n-m}\varepsilon}{\sqrt{n}}.
\end{aligned}$$

Note that $\tilde{K}_{M^*}\Omega^{*-1}_{M^*}(\hat{\Omega}^*_{M^*} - \hat{\Omega}^*_{\hat{M}^*})\Omega^{*-1}_{M^*}\tilde{K}_{M^*} = 0$ for $\left[\hat{M}^*\right] \geq [M^*]$. Therefore assume without loss of generality that $\left[\hat{M}^*\right] < [M^*]$ and let $\hat{a}_{j_2} = \sum_{j_1=1}^{[M^*]}\left\|\left(\hat{\Gamma}^{xy}_{j_1} - \check{\Gamma}^{xy}_{j_1}\right)\vartheta^*_{j_1,j_2}\right\|$ and $\hat{b}_{j_3} = \sum_{j_4=1}^{[M^*]}\left\|\vartheta^*_{j_3,j_4}n^{-1/2}\sum_{t=1}^{n}v_{t,j_4}\right\|$. Note that by (B.2) and the fact that

$$\hat{a}_{j_2} \leq \sup_{j_1}\left\|\hat{\Gamma}^{xy}_{j_1} - \check{\Gamma}^{xy}_{j_1}\right\|\sum_{j_1=1}^{[M^*]}\left\|\vartheta^*_{j_1,j_2}\right\|$$

it follows that $\hat{a}_{j_2} = O_p(n^{-1/2})$ uniformly in j_2. Then,

$$\begin{aligned}
\sqrt{n/M^*}\|d_{101}\| \leq\ & \sqrt{n/M^*}\sum_{l=-m+1}^{m+1}|\hat{\gamma}_l|\left(\sum_{j_2=1,j_3=\left[\hat{M}^*\right]}^{[M^*]}\|\hat{\omega}_{j_2,j_3}(l)\|\hat{a}_{j_2}\hat{b}_{j_4} + \sum_{j_2=\left[\hat{M}^*\right],j_3=1}^{[M^*]}\|\hat{\omega}_{j_2,j_3}(l)\|\hat{a}_{j_2}\hat{b}_{j_3}\right. \\
&\left. + \sum_{j_2,j_3=\left[\hat{M}^*\right]}^{[M^*]}\|\hat{\omega}_{j_2,j_3}(l)\|\hat{a}_{j_2}\hat{b}_{j_4}\right)
\end{aligned}$$

Then consider the largest term in the previous display

$$\begin{aligned}
&\sqrt{n/M^*}\sum_{j_2=1,j_3=\left[\hat{M}^*\right]}^{[M^*]}\|\hat{\omega}_{j_2,j_3}(l)\|\,\hat{a}_{j_2}\hat{b}_{j_3} \\
&\qquad \leq \sqrt{n/M^*}\sum_{j_2=1,j_3=\left[\hat{M}^*\right]}^{[M^*]}\left(\|\hat{\omega}_{j_2,j_3}(l) - \omega_{j_2,j_3}(l)\| + \|\omega_{j_2,j_3}(l)\|\right)\hat{a}_{j_2}\hat{b}_{j_3} \\
&\quad \leq \sqrt{n/M^*}([M^*] - \left[\hat{M}^*\right])\sum_{j_2=1,j_3=1}^{[M^*]}|j_3|\left(\|\hat{\omega}_{j_2,j_3}(l) - \omega_{j_2,j_3}(l)\| + \|\omega_{j_2,j_3}(l)\|\right)\hat{a}_{j_2}\hat{b}_{j_3}.
\end{aligned}$$

Note that

$$\left[\hat{M}^*\right] = [M^*] \text{ iff } [M^*] - M^* + 1 > \hat{M}^* - M^* \geq [M^*] - M^*$$

with $[M^*] - M^* \le -\varepsilon_M$ by the definition of M^*. Then

$$P\left(1 - \varepsilon_M > \hat{M}^* - M^* \ge -\varepsilon_M\right) \le P\left(\left|[M^*] - \left[\hat{M}^*\right]\right| < \eta n^{-\alpha}\right)$$

for any $\alpha > 0$ and any $\eta > 0$. But since $\hat{M}^* - M^* = O_p(n^{-q/(2q+2)})$ the first probability goes to one. This shows that $[M^*] - \left[\hat{M}^*\right]$ converge at arbitrary fast rates such that $\sqrt{n/M^*}([M^*] - \left[\hat{M}^*\right]) = o_p(1)$. The reason for this result is that [.] is piecewise constant so that $[M^*]$ and $\left[\hat{M}^*\right]$ have the same values as soon as M^* and $\hat{M}^*$ are within a certain distance.

By the same arguments as in the proof of Lemma (B.25) it follows that $\sum_{j_2,j_3}^{[M^*]} |j_3| \|\omega_{j_2,j_3}(l)\| \hat{a}_{j_2}\hat{b}_{j_3} = O_p(M^*/n^{1/2}) = O_p(n^{\frac{-q}{2q+2}})$. Next

$$\sum_{j_2=1,j_3=1}^{[M^*]} |j_3| \|\hat{\omega}_{j_2,j_3}(l) - \omega_{j_2,j_3}(l)\| \hat{a}_{j_2}\hat{b}_{j_3} \le \sup_{j_2} \hat{a}_{j_2} \sum_{j_2=1,j_3=1}^{[M^*]} \|\hat{\omega}_{j_2,j_3}(l) - \omega_{j_2,j_3}(l)\| \hat{b}_{j_3}$$

where $\hat{a}_{j_2} = O_p(n^{-1/2})$ uniformly in j_2. We use the Markov and Cauchy-Schwarz inequalities and consider

$$\sum_{j_2=1,j_3=1}^{[M^*]} |j_3| \left(E\|\hat{\omega}_{j_2,j_3}(l) - \omega_{j_2,j_3}(l)\|^2\right)^{1/2} \left(E\hat{b}_{j_3}^2\right)^{1/2} = O(M^*/n^{1/2}) = O(n^{-q/(2q+2)})$$

where $|j_3|\, E\|\hat{\omega}_{j_2,j_3}(l) - \omega_{j_2,j_3}(l)\|^2 \le c_{j_3} n^{-1}$ for some summable sequence c_{j_3} by the proof of Lemma (B.8) and

$$\begin{aligned} Eb_{j_3}^2 &\le \sum_{j_4,j_5=1}^{[M^*]} |j_3|^2 \left\|\vartheta_{j_3,j_4}^*\right\| \left\|\vartheta_{j_3,j_5}^*\right\| \left(E\left\|n^{-1/2}\sum_{t=1}^n v_{t,j_4}\right\|^2 E\left\|n^{-1/2}\sum_{t=1}^n v_{t,j_5}\right\|^2\right)^{1/2} \\ &= O(1). \end{aligned}$$

It follows that $\sqrt{n/M^*}\|d_{101}\| = o_p(n^{\frac{-q}{2q+2}})$.

For d_{102} we use the same bounds as for d_{101} where now $\hat{a}_{j_2}$ is redefined as

$$\check{a}_{j_2} = \sum_{j_1=1}^{[M^*]} |j_1| \left\|\left(\check{\Gamma}_{j_1}^{xy} - \Gamma_{j_1}^{xy}\right)\vartheta_{j_1,j_2}^*\right\|.$$

Consider

$$\begin{aligned} &\sum_{j_2=1,j_3=1}^{[M^*]} |j_3| \|\hat{\omega}_{j_2,j_3}(l) - \omega_{j_2,j_3}(l)\| \check{a}_{j_2}\hat{b}_{j_3} \\ \le\ & \left(\sum_{j_2=1,j_3=1}^{[M^*]} |j_3| \|\hat{\omega}_{j_2,j_3}(l) - \omega_{j_2,j_3}(l)\|^2\right)^{1/2} \left(\sum_{j_2=1,j_3=1}^{[M^*]} \check{a}_{j_2}^2\hat{b}_{j_3}^2\right)^{1/2} \\ \le\ & \left(\sum_{j_2=1,j_3=1}^{[M^*]} |j_3| \|\hat{\omega}_{j_2,j_3}(l) - \omega_{j_2,j_3}(l)\|^2\right)^{1/2} \sum_{j_2=1}^{[M^*]} \check{a}_{j_2} \sum_{j_3=1}^{[M^*]} \hat{b}_{j_3} = O_p(n^{\frac{-4q+1}{4q+4}}) \end{aligned} \tag{B.9}$$

where the first inequality uses the Cauchy-Schwarz inequality and the second inequality follows from $\hat{a}_{j_2} \geq 0, \hat{b}_{j_3} \geq 0$. By the proof of Lemma (B.8) it follows that $\sum_{j_2=1,j_3=1}^{[M^*]} |j_3| \left\|\hat{\omega}_{j_2,j_3}(l) - \omega_{j_2,j_3}(l)\right\|^2 = O_p(M^*/n)$ and

$$E\hat{a}_{j_2}^2 \leq \sum_{j_1,j_5} |j_1| |j_5| \left\|\vartheta_{j_1,j_2}^*\right\| \left\|\vartheta_{j_5,j_2}^*\right\| \left(E\left\|\check{\Gamma}_{j_1}^{xy} - \Gamma_{j_1}^{xy}\right\|^2 E\left\|\check{\Gamma}_{j_5}^{xy} - \Gamma_{j_5}^{xy}\right\|^2\right)^{1/2} = O(n^{-1})$$

uniformly in j_2 by the proof of Lemma (B.24). Thus, $\sum_{j_2=1}^{[M^*]} \hat{a}_{j_2} = O_p(M^*/n^{1/2})$ while $\sum_{j_3=1}^{[M^*]} \hat{b}_{j_3} = O_p(M^*)$ as shown before. This establishes the last equality of (B.9) since $M^{*5/2}/n = O(n^{-(4q+1)/(4q+4)}) = o(1)$ as long as $q > 1/4$. Since $\sqrt{n/M^*}([M^*] - \left[\hat{M}^*\right])$ converges at arbitrarily fast rates it always follows that $\sqrt{n/M^*}([M^*] - \left[\hat{M}^*\right]) \sum_{j_2=1,j_3=1}^{[M^*]} |j_3| \left\|\hat{\omega}_{j_2,j_3}(l) - \omega_{j_2,j_3}(l)\right\| \check{a}_{j_2}\hat{b}_{j_3} = o_p(1)$. Also note that

$$E\sum_{j_2=1,j_3=1}^{[M^*]} |j_3| \left\|\omega_{j_2,j_3}(l)\right\| \check{a}_{j_2}\hat{b}_{j_3} = O(M^*/n^{1/2}).$$

This shows that $\sqrt{n/M^*}\|d_{102}\| = o_p(1)$.

Finally, consider d_{103} where now $a_{j_2} = \sum_{j_1=1}^{[M^*]} |j_1| \left\|\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}^*\right\|$ such that $\sum_{j_2} |a_{j_2}| < \infty$. This implies that $\left(\sum_{j_2=1,j_3=1}^{[M^*]} a_{j_2}^2 \hat{b}_{j_3}^2\right)^{1/2} = O_p(M^{*1/2})$. Consider

$$\begin{aligned}
\sqrt{n/M^*}\sum_{j_2=1,j_3=\left[\hat{M}^*\right]}^{[M^*]} \left\|\omega_{j_2,j_3}\right\| a_{j_2}\hat{b}_{j_3} &\leq \sqrt{n/M^*}\sum_{j_2=1,j_3=\left[\hat{M}^*\right]}^{[M^*]} |j_3|^{1+\delta} \left\|\omega_{j_2,j_3}(l)\right\| a_{j_2}\hat{b}_{j_3} \\
&\leq \sqrt{n/M^*}\left[\hat{M}^*\right]^{-\delta}([M^*] - \left[\hat{M}^*\right]) \sup_{j_2,j_3\leq[M^*],l} |j_3|^{1+\delta} \left\|\omega_{j_2,j_3}\right\| \hat{a}_{j_2}\hat{b}_{j_3}
\end{aligned}$$

where $\sqrt{n/M^*}\left[\hat{M}^*\right]^{-\delta}([M^*] - \left[\hat{M}^*\right]) = o_p(n^{-\delta/(2q+2)})$ while $\sup_{j_2,j_3\leq[M^*],l} |j_3|^{1+\delta} \left\|\omega_{j_2,j_3}(l)\right\| a_{j_2}\hat{b}_{j_3} = O_p(1)$. Next,

$$\sum_{j_2=1,j_3=1}^{[M^*]} |j_3| \left\|\hat{\omega}_{j_2,j_3}(l) - \omega_{j_2,j_3}(l)\right\| a_{j_2}\hat{b}_{j_3} = O_p(n^{-1/2})$$

such that $\sqrt{n}/M^*\|d_{103}\| = o_p(n^{-\delta/(2q+2)}) = o_p(1)$. ■

Lemma B.38. *Let d_{11} be defined in (A.32). Then $d_{11} = o_p(1)$.*

Proof. Write $d_{11} = \sqrt{n/M^*}(d_{111} + d_{112} + d_{113})$ where

$$\begin{aligned}
d_{111} &= \left(\hat{P}_{n-m} - \check{P}_{n-m}\right)' \tilde{K}_{M^*}\Omega_{M^*}^{*-1}(\Omega_{M^*}^* - \hat{\Omega}_{M^*}^*)\Omega_{M^*}^{*-1}(\hat{\Omega}_{M^*}^* - \hat{\Omega}_{\hat{M}^*}^*)\Omega_{M^*}^{*-1}\tilde{K}_{M^*}\frac{Z_{n-m}'\varepsilon}{\sqrt{n}} \\
d_{112} &= \left(\check{P}_{n-m} - P_{n-m}\right)' \tilde{K}_{M^*}\Omega_{M^*}^{*-1}(\Omega_{M^*}^* - \hat{\Omega}_{M^*}^*)\Omega_{M^*}^{*-1}(\hat{\Omega}_{M^*}^* - \hat{\Omega}_{\hat{M}^*}^*)\Omega_{M^*}^{*-1}\tilde{K}_{M^*}\frac{Z_{n-m}'\varepsilon}{\sqrt{n}} \\
d_{113} &= P_{n-m}'\tilde{K}_{M^*}\Omega_{M^*}^{*-1}(\Omega_{M^*}^* - \hat{\Omega}_{M^*}^*)\Omega_{M^*}^{*-1}(\hat{\Omega}_{M^*}^* - \hat{\Omega}_{\hat{M}^*}^*)\Omega_{M^*}^{*-1}\tilde{K}_{M^*}\frac{Z_{n-m}'\varepsilon}{\sqrt{n}}.
\end{aligned}$$

For d_{111} define $\hat{a}_{j_4} = \sum_{j_1,\ldots,j_3} \left\| \left(\hat{\Gamma}^{xy}_{j_1} - \check{\Gamma}^{xy}_{j_1} \right) \vartheta^*_{j_1,j_2} \left(\hat{\omega}_{j_2,j_3} - \omega_{j_2,j_3} \right) \vartheta^*_{j_3,j_4} \right\|$ and

$$\hat{b}_{j_5} = \sum_{j_6=1}^{[M^*]} \left\| \vartheta^*_{j_5,j_6} n^{-1/2} \sum\nolimits_{t=1}^{n} v_{t,j_6} \right\|$$

such that

$$\begin{aligned} \sqrt{n}/M^* \|d_{111}\| &\leq \sqrt{n/M^*} \sum_{l=-m+1}^{m+1} |\hat{\gamma}_l| \left(\sum\nolimits_{j_4,j_5=[\hat{M}^*]}^{[M^*]} \|\hat{\omega}_{j_4,j_5}(l)\| \hat{a}_{j_4} \hat{b}_{j_5} + \sum\nolimits_{j_4=1,j_5=[\hat{M}^*]}^{[M^*]} \|\hat{\omega}_{j_4,j_5}(l)\| \hat{a}_{j_4} \hat{b}_{j_5} \right. \\ &\quad \left. + \sum\nolimits_{j_5=1,j_4=[\hat{M}^*]}^{[M^*]} \|\hat{\omega}_{j_4,j_5}(l)\| \hat{a}_{j_4} \hat{b}_{j_5} \right). \end{aligned}$$

Note that

$$\hat{a}_{j_4} \leq \sum_{l=-m+1}^{m-1} |\hat{\gamma}(l)| \sum_{j_1,\ldots,j_3} \left\| \hat{\Gamma}^{xy}_{j_1} - \check{\Gamma}^{xy}_{j_1} \right\| \left\| \vartheta^*_{j_1,j_2} \right\| \|\hat{\omega}_{j_2,j_3}(l) - \omega_{j_2,j_3}(l)\| \left\| \vartheta^*_{j_3,j_4} \right\|$$

with

$$\sum_{j_1,\ldots,j_3} \left(E \left\| \hat{\Gamma}^{xy}_{j_1} - \check{\Gamma}^{xy}_{j_1} \right\|^2 \right)^{1/2} \left\| \vartheta^*_{j_1,j_2} \right\| \left(E \|\hat{\omega}_{j_2,j_3}(l) - \omega_{j_2,j_3}(l)\|^2 \right)^{1/2} \left\| \vartheta^*_{j_3,j_4} \right\| = O(n^{-1})$$

uniformly in j_4. It then follows by the arguments in the proof of Lemma (B.37) that

$$\sum\nolimits_{j_4=1,j_5=1}^{[M^*]} |j_4| \, \|\hat{\omega}_{j_4,j_5}(l) - \omega_{j_4,j_5}(l)\| \, \hat{a}_{j_4} \hat{b}_{j_5} = O_p(M^{*5/2}/n^{3/2}) = O_p(n^{-\frac{6q+1}{4q+4}}).$$

For

$$\sqrt{n/M^*} \sum\nolimits_{j_4=[\hat{M}^*],j_5=1}^{[M^*]} \|\omega_{j_2,j_3}\| \, \hat{a}_{j_4} \hat{b}_{j_5} \leq \sqrt{n/M^*}([M^*] - \left[\hat{M}^*\right]) \sum\nolimits_{j_4=1,j_5=1}^{[M^*]} |j_4| \, \|\omega_{j_4,j_5}\| \, \hat{a}_{j_4} \hat{b}_{j_5}$$

note that $\sum_{j_4,j_5=1}^{[M^*]} |j_4| \, \|\omega_{j_4,j_5}\| \, \hat{b}_{j_5} = O_p(M^*)$ and $\sum_{j_4=1}^{[M^*]} \hat{a}_{j_4} = O_p(M^*/n)$ such that

$$\sqrt{n/M^*} \sum\nolimits_{j_4=1,j_5=[\hat{M}^*]}^{[M^*]} \|\omega_{j_4,j_5}\| \, \hat{a}_{j_4} \hat{b}_{j_5} = O_p(M^{*2}/n) = O_p(n^{-\frac{2q}{2q+2}}).$$

For d_{112} replace $\check{a}_{j_4}$ by

$$\check{a}_{j_4} = \sum_{j_1,\ldots,j_3} \left\| \left(\check{\Gamma}^{xy}_{j_1} - \Gamma^{xy}_{j_1} \right) \vartheta^*_{j_1,j_2} \left(\hat{\omega}_{j_2,j_3} - \omega_{j_2,j_3} \right) \vartheta^*_{j_3,j_4} \right\| = O_p(n^{-1})$$

such that $\sum_{j_4} \check{a}_{j_4} = O_p(M^*/n)$ leading to $\sum_{j_4,j_5=1}^{[M^*]} |j_4| \, \|\omega_{j_4,j_5}\| \, \check{a}_{j_4} \hat{b}_{j_5} = O_p({M^*}^2/n)$ and $\sqrt{n/M^*} d_{112} = o_p(n^{-2q/(2q+2)})$ by the same arguments as before.

For d_{113} replace a_{j_4} by

$$a_{j_4} = \sum_{j_1,\ldots,j_3} \left\| \Gamma^{xy}_{j_1} \vartheta^*_{j_1,j_2} \left(\hat{\omega}_{j_2,j_3} - \omega_{j_2,j_3}\right) \vartheta^*_{j_3,j_4} \right\| = O_p(n^{-1/2})$$

where a_{j_4} is now summable over j_4 such that

$$\sum\nolimits_{j_4=1,j_5=1}^{[M^*]} |j_5| \left\|\hat{\omega}_{j_4,j_5}(l) - \omega_{j_4,j_5}(l)\right\| a_{j_4} \hat{b}_{j_5} = O_p(n^{-1})$$

and

$$\sqrt{n/M^*}([M^*] - \left[\hat{M}^*\right]) \sum\nolimits_{j_4=1,j_5=1}^{[M^*]} |j_5| \left\|\omega_{j_4,j_5}\right\| a_{j_4} \hat{b}_{j_5} = O_p(n^{-1/2}).$$

It follows that $\sqrt{n}/M^* d_{113} = o_p(n^{-1/2})$. ■

Lemma B.39. *Let d_{12} be defined in (A.33). Then $d_{12} = o_p(1)$.*

Proof. Let $\mathcal{O}^{\Delta}_{M^*} = \Omega^{*-1}_{M^*}(\Omega^*_{M^*} - \hat{\Omega}^*_{M^*})\Omega^{*-1}_{M^*}$. Write $d_{12} = \sqrt{n}/M^*(d_{121} + d_{122} + d_{123})$ where

$$\begin{aligned}
d_{121} &= \left(\hat{P}_{n-m} - \check{P}_{n-m}\right)' \tilde{K}_{M^*} \mathcal{O}^{\Delta}_{M^*} (\hat{\Omega}^*_{M^*} - \hat{\Omega}^*_{\hat{M}^*}) \mathcal{O}^{\Delta}_{M^*} \tilde{K}_{M^*} \frac{Z'_{n-m}\varepsilon}{\sqrt{n}} \\
d_{122} &= \left(\check{P}_{n-m} - P_{n-m}\right)' \tilde{K}_{M^*} \mathcal{O}^{\Delta}_{M^*} (\hat{\Omega}^*_{M^*} - \hat{\Omega}^*_{\hat{M}^*}) \mathcal{O}^{\Delta}_{M^*} \tilde{K}_{M^*} \frac{Z'_{n-m}\varepsilon}{\sqrt{n}} \\
d_{123} &= P'_{n-m} \tilde{K}_{M^*} \mathcal{O}^{\Delta}_{M^*} (\hat{\Omega}^*_{M^*} - \hat{\Omega}^*_{\hat{M}^*}) \mathcal{O}^{\Delta}_{M^*} \tilde{K}_{M^*} \frac{Z'_{n-m}\varepsilon}{\sqrt{n}}.
\end{aligned}$$

Define

$$\begin{aligned}
\hat{a}_j &= \sum_{j_1,\ldots,j_3} \left\| \left(\hat{\Gamma}^{xy}_{j_1} - \check{\Gamma}^{xy}_{j_1}\right) \vartheta^*_{j_1,j_2} \left(\hat{\omega}_{j_2,j_3} - \omega_{j_2,j_3}\right) \vartheta^*_{j_3,j} \right\| \\
\check{a}_j &= \sum_{j_1,\ldots,j_3} \left\| \left(\hat{\Gamma}^{xy}_{j_1} - \check{\Gamma}^{xy}_{j_1}\right) \vartheta^*_{j_1,j_2} \left(\hat{\omega}_{j_2,j_3} - \omega_{j_2,j_3}\right) \vartheta^*_{j_3,j} \right\| \\
a_j &= \sum_{j_1,\ldots,j_3} \left\| \Gamma^{xy}_{j_1} \vartheta^*_{j_1,j_2} \left(\hat{\omega}_{j_2,j_3} - \omega_{j_2,j_3}\right) \vartheta^*_{j_3,j} \right\|
\end{aligned}$$

where $\hat{a}_j = O_p(n^{-1}), \check{a}_j = O_p(n^{-1})$ and $\sum_j a_j = O_p(n^{-1/2})$ as shown in Lemma (B.38). Next define

$$\hat{b}_j = \sum_{j_1,j_2 j_3=1}^{[M^*]} \left\| \vartheta^*_{j,j_1} \left(\hat{\omega}_{j_1,j_2} - \omega_{j_1,j_2}\right) \vartheta^*_{j_2,j_3} n^{-1/2} \sum\nolimits_{t=1}^{n} v_{t,j_3} \right\| = O_p(n^{-1/2})$$

uniformly in j. By the proof of Lemma (B.38) it follows that $d_{12} = o_p(1)$. ■

Lemma B.40. *Let H_5 be defined in (A.34). Then $H_5 = o_p(1)$.*

Proof. Let $H_5 = H_{51} + H_{52} + H_{53}$ where

$$\begin{aligned} H_{51} &= \sqrt{n/M^*}\left(\hat{P}_{n-m} - \check{P}_{n-m}\right)' \tilde{K}_{M^*}\Omega_{M^*}^{*-1}(\hat{\Omega}_{M^*}^* - \hat{\Omega}_{\hat{M}^*}^*)\Omega_{M^*}^{*-1}\tilde{K}_{M^*}\hat{P}_{n-m} \\ H_{52} &= \sqrt{n/M^*}\left(\check{P}_{n-m} - P_{n-m}\right)' \tilde{K}_{M^*}\Omega_{M^*}^{*-1}(\hat{\Omega}_{M^*}^* - \hat{\Omega}_{\hat{M}^*}^*)\Omega_{M^*}^{*-1}\tilde{K}_{M^*}\hat{P}_{n-m} \\ H_{53} &= \sqrt{n/M^*}P_{n-m}'\tilde{K}_{M^*}\Omega_{M^*}^{*-1}(\hat{\Omega}_{M^*}^* - \hat{\Omega}_{\hat{M}^*}^*)\Omega_{M^*}^{*-1}\tilde{K}_{M^*}\hat{P}_{n-m} \end{aligned}$$

and define $\hat{a}_j = \sum_{j_1}\left\|\left(\hat{\Gamma}_{j_1}^{xy} - \check{\Gamma}_{j_1}^{xy}\right)\vartheta_{j_1,j}^*\right\|$, $\check{a}_j = \sum_{j_1}\left\|\left(\hat{\Gamma}_{j_1}^{xy} - \check{\Gamma}_{j_1}^{xy}\right)\vartheta_{j_1,j}^*\right\|$ and $a_j = \sum_{j_1}\left\|\Gamma_{j_1}^{xy}\vartheta_{j_1,j_2}^*\right\|$ where $\hat{a}_j, \check{a}_j = O_p(n^{-1/2})$ uniformly in j and $\sum_j a_j = O(1)$. Also define $b_j = \sum_{j_1}\left\|\vartheta_{j,j_1}^*\hat{\Gamma}_{j_1}^{yx}\right\| = O_p(1)$ uniformly in j. The result then follows by the same arguments as in the proof of Lemma (B.37). ■

Lemma B.41. *Let H_6 be defined in (A.35). Then $H_6 = o_p(1)$.*

Proof. Arguing in the same way as in the proof of Lemma (B.40) the proof follows along the same lines as the proof of Lemma (B.38). ■

Lemma B.42. *Let H_7 be defined in (A.35). Then $H_7 = o_p(1)$.*

Proof. Arguing in the same way as in the proof of Lemma (B.40) the proof follows along the same lines as the proof of Lemma (B.39). ■

References

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

Andrews, D. W. K. (1994): "Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity," *Econometrica*, 62, 43–72.

Brillinger, D. R. (1981): *Time Series, Data Analysis and Theory, Expanded Edition*. Holden-Day, Inc.

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

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

Donald, S. G., and W. K. Newey (1997): "Choosing the Number of Instruments," *Draft, MIT*.

Gawronski, W. (1977): "Matrix Functions: Taylor Expansion, Sensitivity and Error Analysis," *Applicationes Mathematicae*, XVI, 73–89.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50(4), 1029–1053.

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

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

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

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

Hayashi, F., and C. Sims (1983): "Nearly Efficient Estimation of Time Series Models with Predetermined, but not Exogenous, Instruments," *Econometrica*, 51(3), 783–798.

Kuersteiner, G. M. (1997): "Efficient Inference in Time Series Models with Conditional Heterogeneity," Ph.D. thesis, Yale University.

——— (1999a): "Efficient IV Estimation for Autoregressive Models with Conditional Heteroskedasticity," *MIT Working Paper No. 99-08*, forthcoming in Econometric Theory.

——— (1999b): "Optimal Instrumental Variables Estimation for ARMA Models," *MIT Working Paper No. 99-07*, forthcoming in Journal of Econometrics.

Linton, O. (1995): "Seocnd Order Approximation in the Partially Linear Regression Model.," *Econometrica*, pp. 1079–1112.

Linton, O. B. (1997): "Second Order Approximation for Semiparametric Intrumental Variable Estimators and Test Statistics," *mimeo Yale University*.

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

Newey, W. K. (1988): "Adaptive Estimation of Regression Models via Moment Restrictions," *Journal of Econometrics*, 38, 301–339.

——— (1994): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, pp. 1349–1382.

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

Reinsel, G. C. (1995): *Elements of Multivariate Time Series*. Springer-Verlag.

Stoica, P., T. SŤderstrŤm, and B. Friedlander (1985): "Optimal Instrumental Variable Estimates of the AR Parameters of an ARMA Process," *IEEE Transactions on Automatic Control*, 30, 1066–1074.

Whittle, P. (1951): *Hypothesis Testing in Time Series Analysis*. Almqvist and Wiksells boktr, Uppsala.

Xiao, Z., and P. C. Phillips (1998): "Higher Order Approximations for Frequency Domain Time Series Regression," *Journal of Econometrics*, pp. 297–336.

.