Kernel Weighted GMM for Conditionally Heteroskedastic Models

Guido M. Kuersteiner*
MIT

Preliminary Draft

April, 2001

Abstract

This paper extends kernel weighted GMM estimators recently proposed by the author in the context of homoskedastic processes to a class of models with conditionally heteroskedastic innovations. GMM estimation of such models was previously studied by Kuersteiner (1997, 1999a/b) in the context of ARMA processes and Guo and Phillips (1997) in the context of ARCH processes. Optimal implementation of the GMM estimator requires to include more and more instruments as the sample size grows. The use of kernel weighted moment conditions is a natural way to handle the infinite dimensionality of the instrument space. In addition a higher order asymptotic theory is provided to choose the optimal number of instruments in a finite sample context. The higher order analysis reveals that the GMM implementation proposed in Kuersteiner (1997) does not suffer from the usual bias problems of standard 2SLS procedures.

*Massachusetts Institute of Technology, Dept. of Economics, 50 Memorial Drive E52-251B, Cambridge MA 02142. Email: gkuerste@mit.edu. Web page: http://web.mit.edu/gkuerste/www.
Financial support from NSF grant SES-0095132 is gratefully acknowledged.

1 Introduction

GMM estimators are one of the main tools for analyzing models of financial markets. Estimation of these models is often motivated by exploiting Euler equations for optimal investment and consumption decisions. This leads to a set of conditional moment restrictions that can be used to set up a GMM estimator. Rational expectations hypotheses imply that the estimation equation is augmented by an error that satisfies a martingale difference sequence property or in other words that is orthogonal to the information set of the agent.

While the informational structure of rational choice models implies lack of correlation between the innovation and elements of the information set of the agent it does not imply a particular structure for the correlation pattern of nonlinear functions of the innovations. In fact an enormous empirical literature has documented correlations between squared residuals and between other nonlinear functions of the innovation process. A good summary of this literature is Bollerslev, Chou and Kroner (1992) .

In the context of GMM estimation this lack of restrictions on the higher order moments of data and innovations complicates estimation of the optimal weight matrix and correct standard errors for the parameter estimates. Newey and West (1987), Andrews (1991) and Andrews and Monahan (1992) address the issue of consistent covariance matrix estimation in the presence of conditionally heteroskedastic and autocorrelated errors while White (1980) considers the conditionally heteroskedastic case without autocorrelation. Andrews (1991) and Andrews and Monahan (1992) develop automated bandwidth selection procedures that minimize the asymptotic mean squared error of the estimated covariance matrix. Such bandwidth selection rules are not directly applicable for the estimation of the optimal weight matrix in the GMM problem. Xiao and Phillips (1998) obtain optimal data dependent bandwidth rules for the weight matrix in a regression model with serially correlated but homoskedastic errors.

Efficiency properties of GMM estimators in the context of conditionally heteroskedastic errors have been analyzed by Hansen (1985), Bates and White (1990) and Newey (1991). More recently Kuersteiner (1997) and Guo and Phillips (1997) have constructed GMM estimators based on a linear set of instruments in a time series context that allows for general forms of conditional heteroskedasticity. In Kuersteiner (1997) a fully feasible efficient GMM estimator in a more restricted class of problems is analyzed. In the context of an autoregressive model additional restrictions on the fourth order moments of the innovation sequence are used to construct an estimator that is free of truncation or bandwidth parameters. This is achieved by exploiting the parametric structure and moment restrictions in a way that essentially allows to reduce the infinite dimensional instrument problem to a finite dimensional one.

In Kuersteiner (1999b) consistency and asymptotic normality of an infeasible GMM estimator for ARMA models with martingale difference errors is analyzed. No specific restrictions on the form of conditional heteroskedasticity are imposed. A feasible version of such an estimator requires, unlike in the case of Kuersteiner (1997), a truncation parameter that limits the number of instruments used for estimation. The higher order analysis provided in this paper reveals that the dominant distortion from adding more instruments is to the variance of the estimator. This contrasts with the usual result for 2SLS where adding more instruments primarily affects the bias of the estimator. The reason for this difference lies in the way the IV estimator is implemented here. While the weight matrix is truly infinite dimensional under the more general martingale assumptions there is still a parametric component in the construction of the optimal instrument. The presence of the parametric component is responsible for the good bias properties of these alternative IV procedures.

The problem of selecting the right number of instruments has been analyzed by Donald and Newey (1997) in the context of a cross sectional regression. Kuersteiner (2000) extends their approach to a time series context and proposes a new kernel weighted GMM estimator. It is shown that kernel weighting reduces the asymptotic higher order bias of the estimator. In the context of covariance stationary time series processes kernel weighting has intuitive appeal as it exploits an approximate natural ordering of the instruments implied by the summability properties of the autocovariance function of stationary processes.

In this paper we propose to apply the kernel weighting technique to the problem of efficient GMM estimation of univariate time series models with general martingale difference errors similar to the ones studied in Kuersteiner (1999b). We analyze the second order asymptotic properties of such estimators by means of an approximation to the asymptotic mean squared error. It is shown that the particular implementation strategy used in Kuersteiner (1999b) effectively eliminates the dependence of the higher order bias on the number of instruments. The higher order analysis also confirms findings in Kuersteiner (1999a) that under additional diagonality restrictions on the fourth moments of the innovation process the higher order MSE of the estimator essentially becomes independent of the number of instruments. This leads to an efficient GMM estimator that can be implemented without the need to select the number of instruments.

2 Kernel Weighted GMM

We consider the problem of estimating the parameters of a univariate time series model of the form

$$y_t = \phi(1)\mu + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \varepsilon_t - \theta_1 ... - \theta_q \varepsilon_{t-q} \tag{1}$$

where the innovations ε_t are a martingale difference sequence with $E\left|\varepsilon_t\right|^{12} < \infty$. We define the lag polynomials $\phi(L) = 1 - \phi_1 L - ... - \phi_p L^p$ and $\theta(L) = 1 - \theta_1 L - ... - \theta_q L^q$ where L is the lag operator. The vector of parameters determining (1) is defined as $\beta = (\phi_1, ..., \phi_p, \theta_1, ..., \theta_q)$. We allow for special cases of (1) where $q = 0$. Here we consider estimation of $\phi = (\phi_1, ..., \phi_q)$ while the moving average part of the model is treated as a nuisance parameter. Estimators for this model were proposed by Hayashi and Sims (1983), Stoica, Soderstrom and Friedlander (1985) Hansen and Singleton (1996) and Kuersteiner (2000).

The martingale difference property of ε_t imposes restrictions on the fourth order cumulants. These restrictions can be conveniently summarized by defining the following function

$$\sigma(s,r) = \begin{cases} E\left(\varepsilon_t^2 \varepsilon_{t-|s|} \varepsilon_{t-|r|}\right) & r \neq s \\ E\left(\varepsilon_t^2 \varepsilon_{t-s}^2\right) - \sigma^4 & r = s \end{cases} \quad \text{for } r, s \in \{0, \pm 1, \pm 2, ...\}. \tag{2}$$

It should be emphasized that $\sigma(s,r)$ is equal to the fourth order cumulant for $s, r > 0$. Let

$$\alpha_{s,r} = \begin{cases} \sigma(s,r) & \text{if } s \neq r \\ \sigma(r,r) + \sigma^4 & \text{if } s = r \end{cases} \tag{3}$$

We assume that we have a probability space $(\Omega, \mathcal{F}, P)$ with a filtration $\mathcal{F}_t$ of increasing σ-fields such that $\mathcal{F}_t \subseteq \mathcal{F}_{t+1} \subseteq \mathcal{F}$ $\forall t$. The doubly infinite sequence of random variables $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ generates the filtration $\mathcal{F}_t$ such that $\mathcal{F}_t = \sigma(\varepsilon_t, \varepsilon_{t-1}, ...)$. The assumptions on $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ are summarized as follows:

Condition 1 *(i)* ε_t *is strictly stationary and ergodic, (ii)* $E(\varepsilon_t \mid \mathcal{F}_{t-1}) = 0$ *almost surely, (iii)* $E\left(\varepsilon_t^2\right) = \sigma^2 > 0$*, (iv)* $\sum_{r=1}^{\infty} \sum_{s=1}^{\infty} |\sigma(s,r)| = B < \infty$*, (v)* $E\left(\varepsilon_t^2 \varepsilon_{t-s}^2\right) \neq 0$ *for all* s.

Remark 1 *Assumption 1(iii) rules out degenerate distributions. A consequence of the martingale assumption (i) is that in general terms of the form* $E\left(\varepsilon_t^2 \varepsilon_{t-s}\varepsilon_{t-r}\right)$ *are nonzero for* $s \neq r \neq 0$ *and depend on* s *for* $s = r \neq 0$*. Assumption (iv) limits the dependence in higher moments by imposing a summability condition on the fourth cumulants. The assumption is needed to prove invertibility of the infinite dimensional weight matrix of the optimal GMM estimator. Assumption (v) together with* $\sum_{s=1}^{\infty} |\sigma(s,s)| < \infty$ *implied by (iv) insures that* $E\left(\varepsilon_t^2 \varepsilon_{t-s}^2\right) > \underline{\alpha}$ *uniformly in* s*.*

For the higher order expansions in the next section we need additional moment restrictions in the form of summability conditions on higher order cumulants. Such conditions are common in the literature on optimal weight matrix estimation as for example Andrews (1991). We first define the higher order cumulants. Let $\xi = (\xi_1, ..., \xi_k) \in \mathbb{R}^k$ and $\varepsilon = (\varepsilon_{t_1}, ..., \varepsilon_{t_k})$, then $\phi_{t_1,...,t_k}(\xi) \equiv E\left(e^{i\xi'\varepsilon}\right)$ is the joint characteristic function with corresponding cumulant generating function $\ln \phi_{j_1,...j_k,t_1,...,t_k}(\xi)$. The joint v-th order cross-cumulant is

$$\text{cum}_{v_1,...,v_k}(\varepsilon_{t_1}, ..., \varepsilon_{t_k}) \equiv \frac{\partial^{v_1+...+v_k}}{\partial \xi_1^{v_1} \cdots \partial \xi_k^{v_k}} \ln \phi_{j_1,...j_k,t_1,...,t_k}(\xi)\Big|_{\xi=0} \tag{4}$$

where v_i are nonnegative integers $v_1 + ... + v_k = v$. When $v_1, ..., v_k = 1$ the shorthand notation $\text{cum}(\varepsilon_{t_1}, ..., \varepsilon_{t_k})$ for $\text{cum}_{1,...,1}(\varepsilon_{t_1}, ..., \varepsilon_{t_k})$ is used.

Condition 2 *For* $\text{cum}(\varepsilon_{t_1}, ..., \varepsilon_{t_k})$ *defined in (2) with* $E|\varepsilon_t|^{12} < \infty$

$$\sum_{t_1,...,t_{k-1}=-\infty}^{\infty} |1 + t_k| \left|\text{cum}(\varepsilon_{t_1}, ..., \varepsilon_{t_k})\right| < \infty \text{ for } k = 2, 3, ..., 12$$

In order to guarantee identification of the parameters $\phi = (\phi_1, ..., \phi_p)$ we impose restrictions on the parameter space Θ. These restrictions insure the existence of a stationary solution to (1) and guarantee that the autoregressive and moving average parts of the model do not cancel out.

Condition 3 *Let* $C(\beta, L) = \theta(L)/\phi(L)$*. The parameter space* $\Theta \subset int\Theta_0$ *where* Θ_0 *is a subset of* $\mathbb{R}^d$ *defined by* $\Theta_0 = \{\beta \in \mathbb{R}^d \,|\phi(\zeta) \neq 0 \text{ for } |\zeta| \leq 1,\ \theta(\zeta) \neq 0 \text{ for } |\zeta| \leq 1, \theta(\zeta), \phi(\zeta) \text{ have no common zeros}, \theta_q \neq 0, \phi_p \neq 0\}$*. Assume that* Θ *is compact in* $\mathbb{R}^d$*.*

GMM estimation of (1) exploits moment conditions of the form

$$E(\phi_0(L)(y_{t+q} - \mu)\varepsilon_{t-j}) = 0 \text{ for all } j > 0.$$

Note that this representation is valid for the pure AR(p) case as well as for the case where the moving average part is treated as a nuisance parameter. It is convenient to collect the instruments $\varepsilon_t, \varepsilon_{t-1}, ...$ in a vector $\varepsilon_{t,n}^* = (\varepsilon_{t-1}, \varepsilon_{t-2}, ...\varepsilon_{t-n})'$. In practice such an instrument vector is not available. We introduce the observable instrument $\varepsilon_{t,n} = S_n(t)\varepsilon_{t,n}^*$ where $S_n(t) = \text{diag}(\{t > 1 + p + q\}, ..., \{t > n + p + q\})$ where $\{.\}$ is the indicator function. An efficient GMM estimator with instruments ε_t is then obtained by weighting the moment conditions by a weight matrix Ω_n.

When $q = 0$ then $\Omega_M = n^{-1}\sum_{t,s=1}^{n} E\varepsilon_t\varepsilon_s\varepsilon_{t,M}\varepsilon_{s,M}'$. Using the martingale property of the innovations this expression simplifies to

$$\Omega_M = n^{-1}\sum_{t=1}^{n} E\varepsilon_t^2\varepsilon_{t,M}\varepsilon_{t,M}' = n^{-1}\sum_{t=1}^{n} S_M(t)\tilde{\Omega}_M S_M(t)$$

with $\tilde{\Omega}_M = E\varepsilon_t^2\varepsilon_{t,M}^*\varepsilon_{t,M}^{*\prime}$ which is independent of t due to the stationarity of ε_t. In fact

$$\tilde{\Omega}_M = \begin{bmatrix} \sigma(1,1) + \sigma^4 & \cdots & \sigma(1, M) \\ \vdots & \ddots & \vdots \\ \sigma(M, 1) & \cdots & \sigma(M, M) + \sigma^4 \end{bmatrix}$$

and for M fixed $\Omega_M \to \tilde{\Omega}_M$ as $n \to \infty$

For the case where $q > 0$ the instruments need to be adjusted for the potential presence of moving average terms. The optimal weight matrix is now given by

$$\Omega_n = n^{-1}\sum_{j=-q}^{q}\sum_{t=q+1}^{n-q} Eu_{t+q}u_{t+q-j}\varepsilon_{t,n}\varepsilon_{t-j,n}'$$

with $u_t = \theta(L)\varepsilon_t$. Let $\gamma_u(j) = Eu_tu_{t-j}/Eu_t^2 = \sum_{k=0}^{q}\theta_k\theta_{k+j}$ where $\theta_0 = 1$ and $\theta_k = 0$ for $k > q$. It then follows that $Eu_{t+q}u_{t+q-j}\varepsilon_{t,n}\varepsilon_{t-j,n}' = \gamma_u(j)S_n(t)\Omega_n(j)S_n(t-j)$ where $\Omega_n(j)$ is defined as $\Omega_n(j) = E\varepsilon_{t+q}^2\varepsilon_{t,n}^*\varepsilon_{t-j,n}^{*\prime}$.

We now introduce the kernel weighted GMM estimator proposed in Kuersteiner (2000). Define the matrix

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

having kernel weight $k(j/M)$ in its j-th diagonal. M is a bandwidth parameter that controls the number of lagged instruments used in estimation. The kernel function satisfies the following conditions.

Condition 4 *The kernel function* $k(.)$ *satisfies* $k : \mathbb{R} \mapsto [-1,1]$, $k(0) = 1$, $k(x) = 0$ *for* $|x| > 1, k(x) = k(-x) \forall x \in \mathbb{R}$, $\int |k(x)|\, dx < \infty$, $k(.)$ *is continuous at* 0 *and at all but a finite number of points.*

Condition 5 *The kernel function* $k(.)$ *satisfies Assumption (4) 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* $k_q \in (0, \infty)$.

Letting $X = [x_{\max(q,p)+1} - \bar{x}, ..., x_n - \bar{x}]'$ with $x_t = [y_{t-1} - \bar{y}, ..., y_{t-p} - \bar{y}]'$ and $Q_M = \left[\varepsilon_{\max(q,p)+1,M}, ..., \varepsilon_{n,M}\right]'$ the GMM estimator can now be written as a 2SLS estimator

$$\hat{\beta} = \left(X'Q_M K_M \Omega_M^{-1} K_M Q_M' X\right)^{-1} X'Q_M K_M \Omega_M^{-1} K_M Q_M' Y$$

where $Y = \left[y_{\max(q,p)+1} - \bar{y}, ..., y_n - \bar{y}\right]$. The higher order properties of this estimator, which corresponds to standard GMM when the truncated kernel is used, were analyzed in Kuersteiner (2000) under the additional assumption that ε_t is conditionally homoskedastic. It is shown there that using a suitable kernel function reduces the bias of the kernel weighted GMM estimator relative to classical 2SLS.

This bias reduction comes at the cost of an increased second order variance. The reason for this inefficiency lies in the fact that the kernel weighted GMM procedure uses an inefficient weight matrix $K_M \Omega_M^{-1} K_M$ rather than the efficient weight matrix Ω_n^{-1}. This inefficiency however only affects the higher order properties of the estimator while first order efficiency is maintained.

Here we focus on an alternative version of the estimator that was proposed in Kuersteiner (1997). The idea behind the alternative estimator is to replace $n^{-1}X'Q$ by the population moments. These population moments depend on the underlying parameters in a way that allows to estimate them at parametric rates. Note that $En^{-1}X'Q_M \approx P_M$ where $P_M = [b_1, ..., b_M]$ with the l-th element of b_k given by $[b_k]_l = (2\pi)^{-1} \int_{-\pi}^{\pi} C(\beta, L) e^{i\lambda(k-l)} d\lambda$. One can then define the alternative estimator based on $Z_M' = P_M \Omega_M^{-1} K_M Q'$ by

$$\hat{\beta}_{if} = \left(Z_M' X\right)^{-1} Z_M' Y.$$

Using a parametric estimate for P_n essentially removes the higher order bias that is typically associated with standard 2SLS procedures where the bias is explained by the correlation between $X'Q$ and $Q'\varepsilon$. The summability properties of P_n also make it possible to eliminate one of the kernel functions.

In order to arrive at a feasible procedure we need to replace the unobservable quantities Z_M by estimated counterparts. We assume that we have $\sqrt{n}$ consistent first stage estimates $\tilde{\phi}$ which can be obtained from simple OLS regressions when $q = 0$ or from inefficient IV procedures using $y_{t-q}, ..., y_{t-p-q}$ as instruments when $q > 0$. Using the first stage estimates of ϕ we can then obtain corresponding consistent residuals $\tilde{\varepsilon}_t$. In the case where we only estimate the AR part of an ARMA model consistent estimates of $u_t = \theta(L)\varepsilon_t$ are obtained from $\tilde{u}_t = \tilde{\phi}(L)y_t$. Using these estimated $\tilde{u}_t$ the parameters θ can be consistently estimated. Since we only require $\sqrt{n}$ consistency for these parameters a simple procedure such as pseudo maximum likelihood or nonlinear leas squares can be used. Such a procedure is discussed in Kuersteiner (2000).

A $\sqrt{n}$ consistent estimator of P_n is straight forwardly available from

$$\left[\tilde{b}_k\right]_l = (2\pi)^{-1} \int_{-\pi}^{\pi} C(\tilde{\beta}, \lambda) e^{i\lambda(k-l)} d\lambda$$

where $\tilde{\beta}$ is the set of first order estimates $\tilde{\phi}$ and $\tilde{\theta}$. When we only estimate the AR part of an ARMA model an estimate of $C(\beta, \lambda)$ can be obtained from $\theta(e^{i\lambda})/\phi(e^{i\lambda})$. If $q = 0$ then $C(\beta, \lambda) = 1/\phi(e^{i\lambda})$ and the coefficients b_k are the impulse response coefficients of the AR model.

The covariance matrix Ω_n is estimated by $n^{-1} \sum_{t=1}^{n} \tilde{\varepsilon}_t^2 \tilde{\varepsilon}_{t,n} \tilde{\varepsilon}_{t,n}'$ in the case where model (1) is of the restricted form $q = 0$. The innovations $\tilde{\varepsilon}_t$ are easily obtained from $\tilde{\varepsilon}_t = \tilde{\phi}(L)y_t$. In the case of AR estimation in an ARMA model we set

$$\tilde{\Omega}_M = n^{-1} \sum_{t=1}^{n} \sum_{j=-q}^{q} \tilde{u}_{t+q} \tilde{u}_{t+q-j} \tilde{\varepsilon}_{t,M} \tilde{\varepsilon}_{t-j,M}'.$$

Here estimation of $\tilde{u}_t = \tilde{\phi}(L)y_t$ proceeds as described before. The martingale innovations can be estimated recursively by using the consistent estimate $\tilde{\theta}$. The details of this procedure are outlined in the Appendix.

We are now in a position to define the feasible version of $\hat{\beta}_{if}$. Defining $\hat{Z}_{n,M} = \hat{P}_M \hat{\Omega}_M^{-1} K_M \hat{Q}_M$ where $\hat{P}_M = [\hat{b}_1, ..., \hat{b}_M]$ and $\hat{Q}_M = \left[\hat{\varepsilon}_{\max(p+q)+1,M}, ..., \hat{\varepsilon}_{n,M}\right]$ we have

$$\hat{\beta}_{n,M} = (\hat{Z}_{n,M}' X)^{-1} \hat{Z}_{n,M}' Y.$$

It is convenient to let $n^{-1}\hat{Z}_{n,M}' X = \hat{D}_M$ and $n^{-1/2}\hat{Z}_{n,M}' \varepsilon = \hat{d}_M$. Then$\sqrt{n}(\hat{\beta} - \beta_0) = \hat{D}_M^{-1} \hat{d}_M$.

When Ω_n^{-1} is a diagonal matrix, $q = 0$, $k(.)$ is the truncated kernel and $M = n$ then $\hat{\beta}_{n,M}$ corresponds to the estimator analyzed in Kuersteiner (1997). It is shown there that under these conditions the estimator is first order asymptotically equivalent to an infeasible optimal IV estimator. In the more general case when Ω_n^{-1} is not diagonal this result no longer holds and the number M of instruments used can grow at most at rate $o(n)$. This rate however is not optimal as the higher order analysis in this paper shows. The need to estimate a high dimensional weight matrix adds to the higher order variance of the estimator. This additional variance has to be balanced against the efficiency gains from using more instruments.

An alternative to implementing the estimator $\hat{\beta}_{n,M}$ can be obtained from defining an approximation to the optimal weight matrix Ω_n as

$$\Omega_{n,M}^{*} = \begin{bmatrix} \Omega_M & 0 \\ 0 & \sigma^4 I_{n-M} \end{bmatrix}.$$

It is shown in Kuersteiner (1999) that as $n, M \to \infty$ the matrix $\Omega_{n,M}^{*}$ converges to Ω_n in an apropriate operator norm on the space of square summable sequences. The inverse of $\Omega_{n,M}^{*}$ is given by diag $\left(\Omega_M^{-1}, \sigma^{-4} I\right)$. The infeasible estimator β_{if}^{*} is then obtained from $Z_M^{*\prime} = P_n \Omega_{n,M}^{*-1} Q_n^{\prime}$ by

$$\beta_{if}^{*} = \left(Z_M^{*\prime} X\right)^{-1} Z_M^{*\prime} Y.$$

Feasible versions of this estimator are formulated in the same manner as before by replacing Ω_M by an estimate $\hat{\Omega}_M$ and P_n by an estimate $\hat{P}_n$. Note that P_n can be estimated $\sqrt{n}$-consistently independent of the dimension n.

3 Second Order Asymptotic Approximation

Approximations are developed around the optimal infeasible version of the estimators. These can be obtained by letting $\varepsilon_{t,\infty} = (\varepsilon_{t-1}, \varepsilon_{t-2}, ...)$ and defining the optimal weight matrix

$$\Omega(j) = E\varepsilon_{t+q}^2 \varepsilon_{t,\infty} \varepsilon_{t-j,\infty}^{\prime}$$

and $\Omega = \sum_{j=-q}^{q} \gamma_u(j)\Omega(j)$. In the case where $q = 0$ this matrix simplifies to $\Omega = \Omega(0)$. In the terminology developed in Kuersteiner (1999b) Ω defines an invertible operator on an l^2 space. It follows that $P'\Omega^{-1} \in l^2$ for $P' = [b_1, b_2, ...]$ which has rows that are in l^2. This implies that $z_t^{\infty} = P'\Omega^{-1}\varepsilon_{t,\infty}$ exists almost surely, is stationary and has finite second moments.

Letting $D = P'\Omega^{-1}P$ an infeasible efficient estimator for $\hat{\beta}$ can be written as $\sqrt{n}(\hat{\beta} - \beta_0) = D^{-1}d_0$ where $d_0 = n^{-1/2}\sum_t u_{t+q} z_t^{\infty}$ and it follows from results in Kuersteiner (1999b) that $D^{-1}d_0 \to N(0, D^{-1})$.

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

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

The expansion is valid as long as $M/n \to 0$. We decompose the expansion into $\hat{D}_M - D = H_1 + ... + H_k$ and $\hat{d}_n = d_0 + d_1 + ... + d_p$ such that

$$\sqrt{n}(\beta_{n,M} - \beta) = D^{-1} \sum_i d_i + D^{-1} \sum_i \sum_j H_i D^{-1} d_j + o_p(M/n).$$

We now denote by $\sqrt{n}\,(b_{n,M} - \beta)$ all the terms $D^{-1}d_i$ and $D^{-1}H_iD^{-1}d_j$ which have biases and variances of order $O(M/n)$ or terms that are mean zero with variance $O(M^{-q})$. The remaining terms $R_{n,M} = \sqrt{n}(\beta_{n,M} - b_{n,M})$ are of order $o_p(M/n)$. The size of the mean squared error of the estimator is given in the next lemma. Define the approximate mean squared error of $\beta_{n,M}$ 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. The next proposition gives an expression for the asymptotic MSE using the largest in probability terms depending on M and n. For this purpose we define the i,j-th element of the infinite dimensional inverse Ω^{-1} as $\vartheta_{i,j}$.

Proposition 1 *Suppose Assumptions (1), (2) and (3) hold and $k()$ satisfies Assumptions (4) and (5). If $M \to \infty$ and $M^{2q+1}/n \to \kappa$ then for any $\ell \in \mathbb{R}^d$ with $\ell'\ell = 1$*

$$\lim_n n/M \varphi_n(M, \ell, k(.)) = \mathcal{A} + k_q^2 \mathcal{B}^{(q)}/\kappa$$

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

$$\mathcal{A}_1 = \lim_M M^{-1} \sum_{j_1,...,j_7=1}^{\infty} b_{j_1} \vartheta_{j_1,j_2} \vartheta_{j_3,j_4} k(j_4/M)^2 \vartheta_{j_4,j_2-j_3+j_6} \vartheta_{j_6,j_7} b_{j_7}, \tag{5}$$

$$\mathcal{B}_1^{(q)} = \sum_{l,j=1}^{\infty} b_l \vartheta_{lj} \, |j|^q \, b_j' \tag{6}$$

This result is remarkable in several ways. If we compare the size of the MSE depending positively on M we note that it is of order $O(M/n)$. This contrasts with standard GMM which is of order $O(M^2/n)$ as shown in Kuersteiner (2000). In other words the MSE of the estimators here is comparable to bias corrected GMM. If we analyze the size of the bias we find that it is independent of M altogether. The largest terms appearing in the mean squared error approximation are thus variance terms. Here we find a trade-off between first order efficiency which requires to add more instruments and second order distortion where more instruments increase the variability of the estimator.

If a truncated kernel is used then the efficiency gains from using more instruments are of secondary importance and the second term in the asymptotic mean squared error vanishes such that $\lim_n n/M\varphi_n(M,\ell,k(.)) = \mathcal{A}$. In this case the optimal rate of increase of the number of instruments is $\log(n)$.

We now analyze the IV estimator under the additional assumption that the optimal weight matrix is diagonal with all off diagonal elements equal to zero. This assumption was proposed in Kuersteiner (1997, 1999) and is satisfied for example for GARCH models with symmetric innovation distributions and many stochastic volatility models.

Condition 6 *Assume that* $\sigma(s,r) = 0$ *for* $s \neq r$.

It was shown in Kuersteiner (1997, 1999a/b) that under this assumption there is no need for kernel smoothing of the instruments and that the number of instruments is allowed to grow at the same rate as the sample. This result is confirmed by the second order approximation of the mean squared error which now does no longer depend positively on the number of instruments used.

Corollary 1 *Suppose Assumptions (1), (2), (3) and (6) hold and* $k()$ *satisfies Assumptions (4) and (5). If* $M \to \infty$ *and* $M^{2q+1}/n \to \kappa$ *then for any* $\ell \in \mathbb{R}^d$ *with* $\ell'\ell = 1$

$$\lim_n n/M\varphi_n(M,\ell,k(.)) = k_q^2 \mathcal{B}^{(q)}/\kappa$$

with the constants $\mathcal{B}^{(q)} = 1/2\ell' D^{-1/2}(\mathcal{B}_1^{(q)}\mathcal{B}_1^{(q)'} - \mathcal{B}_1^{(q)} D^{-1}\mathcal{B}_1^{(q)'})D^{-1/2}\ell$ *where* $\mathcal{B}_1^{(q)}$ *is defined as*

$$\mathcal{B}_1^{(q)} = \sum_{l=1}^{\infty} \frac{|l|^q}{\alpha_{l,l}} b_l b_l' \tag{7}$$

Under the Condition (6) it is therefore optimal from a second order point of view to use the truncated kernel for which $k_q = 0$ and to set $M = n$. This is exactly the estimator that was proposed by Kuersteiner (1997). The second order expansion shows that this estimator does not have the usual bias problems associated with GMM procedures.

4 Monte Carlo Simulations

In this section we report a small Monte Carlo experiment to demonstrate the behavior of a number or estimators discussed in the previous sections. We generate samples of size $n = 2^k$ for $k = 7, 8, 9$ from the following model

$$y_t = \phi y_{t-1} + \varepsilon_t \tag{8}$$

where ε_t is generated by the process $\varepsilon_t = u_t h_t^{1/2}$ with $u_t \sim N(0,1)$ and $h_t = .1 + .5\varepsilon_{t-1}^2$.

Starting values are $y_0 = 0$, $h_0 = 0$ and $\varepsilon_0 = 0$. In each sample the first 500 observations are discarded to eliminate dependence on initial conditions. Small sample properties are evaluated for different values of $\phi \in (-1, 1)$.

The parameter ϕ is estimated by using different implementations of

$$\hat{\phi}(M) = \left(X' Q_M K_M \Omega_M^{-1} K_M Q_M' X\right)^{-1} X' Q_M K_M \Omega_M^{-1} K_M Q_M' Y.$$

Define $X = [y_1, ..., y_{n-1}]$ and $Y = [y_2, ..., y_n]$. We use the pseudo maximum likelihood estimator $\tilde{\phi} = (X'X)^{-1} X'Y$ as our first stage consistent estimator leading to residuals $\hat{\varepsilon}_t = y_t - \tilde{\phi} y_{t-1}$. We define the instruments $Q_{\varepsilon,M} = [\hat{\varepsilon}_{n-M,n}, ..., \hat{\varepsilon}_{n,n}]$ and $Q_{y,M} = [y_{n-M,n}, ..., y_{n,n}]$ where $y_{t,M} = S_n(t) [y_1, ..., y_n]'$. We use the Parzen kernel defined by $k_P(x) = (1 - 6x^2 + 6|x|^3)\{|x| \leq 1/2\} + (2(1 - |x|^3)\{1/2 \leq |x| \leq 1\}$ to implement the kernel weighted procedures. The optimal weight matrix $\hat{\Omega}_M$ is estimated as $\hat{\Omega}_M = n^{-1} \sum_{t=1}^n \hat{\varepsilon}_t^2 \hat{\varepsilon}_{t,M} \hat{\varepsilon}_{t,M}'$. Using these definitions we define four different implementations of $\hat{\phi}(M)$ summarized in the following table:

$$\begin{array}{rll}
\hat{\phi}_y(M): & Q_M = Q_{y,M} & K_M = I_M \\
\hat{\phi}_\varepsilon(M): & Q_M = Q_{\varepsilon,M} & K_M = I_M \\
\hat{\phi}_{k,y}(M): & Q_M = Q_{y,M} & K_M = \text{diag}(k_p(1/M), ..., k_p(1)) \\
\hat{\phi}_{k,\varepsilon}(M): & Q_M = Q_{\varepsilon,M} & K_M = \text{diag}(k_p(1/M), ..., k_p(1)).
\end{array}$$

For the implementation of $\hat{\beta}_{n,M}$ we use the first stage estimator $\tilde{\phi}$ to obtain the vector $\tilde{P}_M = \left[\tilde{\phi}, \tilde{\phi}^2, ..., \tilde{\phi}^M\right]$ and construct $\hat{Z}_M = \hat{P}_M \hat{\Omega}_M^{-1} K_M \hat{Q}_{\varepsilon,M}$ such that $\hat{\phi}_{n,M} = \left(\hat{Z}_M' X\right)^{-1} \hat{Z}_M' Y$.

We are also considering an estimator that imposes the diagonality restriction implied by the ARCH residuals on Ω. This estimator was proposed in Kuersteiner 1997. It is defined as

$$\tilde{\phi}_{FD} = \left[\sum_{j=1}^{n-1} I_{n,yy}(\lambda_j) \hat{h}^x(\hat{\phi}, \lambda_j) \right]^{-1} \sum_{j=1}^{n-1} I_{n,yy}(\lambda_j) \hat{h}(\hat{\phi}, \lambda_j)$$

where

$$\begin{aligned} \hat{h}\left(\hat{\phi}, \lambda\right) &= \operatorname{Re}\left[\hat{l}_{\psi}(-\lambda)(1 - \hat{\phi} e^{i\lambda})\right], \\ \hat{l}_{\psi}(\lambda) &= \sum_{j=1}^{n-2} \hat{\alpha}_j^{-1}(\hat{\phi}) \hat{b}_j e^{-i\lambda j} \end{aligned}$$

and $\hat{\alpha}_j(\hat{\phi}) = \frac{1}{n} \sum_{t=p+j+1}^{n} \hat{\varepsilon}_t^2 \hat{\varepsilon}_{t-j}^2$. This estimator uses a frequency domain representation of $\hat{P}_{n-2} \hat{A}_{n-2}^{-1} Q_{\varepsilon,n-2}$ where $\hat{A}_{n-1} = \operatorname{diag}(\hat{\alpha}_1(\hat{\phi}), ..., \hat{\alpha}_{n-2}(\hat{\phi}))$. This estimator in other words imposes the diagonality restriction on Ω and uses as many instruments as observations in the sample.

Tables 1-4 show the sample quantiles from 1,000 simulation replications of the estimators $\hat{\phi}_y(M), \hat{\phi}_{\varepsilon}(M), \hat{\phi}_{k,y}(M), \hat{\phi}_{k,\varepsilon}(M)$ and $\hat{\phi}_{n,M}$ for $M = 1, 2, 3, 4, 5, 10$ and $n = 128$. We also report estimates for $\tilde{\phi}_{FD}$ which does not require the choice of a bandwidth parameter. For ϕ_1 the bias of OLS and instrumental variables procedures using just one instrument is generally very small. It increases as the number of instruments increases.The results also show that kernel weighting reduces the bias especially when M is large. This is true both for implementations using y_t as instruments as well as ε_t. The bias of $\tilde{\phi}_{FD}$ is larger than the OLS bias but still relatively small while the bias for $\hat{\phi}_{n,M}$ is generally larger than for the other implementations of the estimator. In terms of mean absolute error (MAE) and mean squared error (MSE) $\tilde{\phi}_{FD}$ dominates all other estimators. For the non-kernel weighted procedures the MSE generally increases monotonically with M while for the kernel weighted implementations generally slightly lower MAE and MSE values are achieved for $M > 1$. As far as variance is concerned only $\tilde{\phi}_{FD}$ manages to outperform the OLS estimator while $\hat{\phi}_{n,M}$ has a generally a slightly larger variance than the other procedures.

As ϕ_1 increases towards 1 the performance of $\hat{\phi}_{n,M}$ relative to the other estimators starts to improve. For $\phi_1 = .6$ and $\phi_1 = .9$ $\hat{\phi}_{n,M}$ has lowest bias for all reported choices of M. As far as variance and MSE is concerned $\hat{\phi}_{n,M}$ is still dominated by other procedures. The lowest values for MAE, MSE and variance are achieved for $\tilde{\phi}_{FD}$ as long as $\phi_1 \leq .6$. Only for $\phi_1 = .9$ does the implementation $\hat{\phi}_y(M)$ seem to dominate. The differences are however small. Bias reduction through kernel methods seems to be less reliable on the other hand when ϕ_1 is large.

The same simulations for lareger sample size are repeated where n is now 256 and 512 respectively. The results remain essentially unchanged compared to the situation for 128 observations and are therefore not reported.

5 Conclusions

We have provided a higher order asymptotic analysis of estimators for the autoregressive parameters in ARMA models when the innovations are general martingale difference sequences. Instrumental variables estimators which are based on instruments that are linear in the innovations are frequently used to estimate such models. First order asymptotic efficiency requires that the number of instruments tends to infinity. In practice such a rule can not be directly implemented due to limitations of sample size. More importantly standard 2SLS procedures suffer from significant small sample bias. This bias is increasing with the number of instruments.

Here it is shown that parametric estimation of the sample correlation between instruments and regressors essentially eliminates the instrument induced bias. Despite the good behavior of these estimators as far as their bias is concerned they do suffer from increased small sample variance which is increasing with the number of instruments. The source of this variance is the estimate of the optimal weight matrix.

It also turns out that if the weight matrix is sufficiently simple, in particular if it is diagonal then the instrument set can grow at the same rate as the sample size. In this case the efficiency variance trade-off between adding additional instruments disappears and the optimal instrumental variables estimator can essentially be implemented.

A Proofs

Frequent reference will be made to the proofs in Kuersteiner (2000) which will be abbreviated by MRG.

Proof of Proposition 1:

We note that $D = P'\Omega^{-1}P = \sum_{k,j=1}^{\infty} b_k b_j' \vartheta_{kj}$ where ϑ_{kj} is the k,j-th element of Ω^{-1} and $D = H_{11}+H_{12}+H_{13}$ where $H_{11} = -\sum_{k,j=n+1}^{\infty} b_k b_j' \vartheta_{kj} + 2\sum_{k,j=1}^{n}\sum_{j=n+1}^{\infty} b_k b_l' \vartheta_{kl} = o(n^{-2s})$ for all s by Lemma A.9 in MRG and $H_{12} = \sum_{l,j=1}^{n} b_l(1-k(\frac{l}{M})\vartheta_{kj} b_j' = M^{-q} k_q \sum_{l,j=1}^{n} b_l \left|l\right|^q \vartheta_{lj} b_j' + o(M^{-q})$ by Lemma A.11 in MRG. $H_{13} = \sum_{l,j=1}^{n} b_l k(\frac{l}{M})\vartheta_{kj} b_j'$ can now be decomposed. Let $H_{13} = H_{21} + H_{22}$ where $H_{21} = \sum_{l,j=1}^{n}\left(b_l - \hat{b}_l\right) k(\frac{l}{M})\vartheta_{lj} b_j'$ and $H_{22} = \sum_{l,j=1}^{n} \hat{b}_l k(\frac{j}{M})\vartheta_{lj}(b_j' - \hat{\Gamma}_j^{\hat{\varepsilon}x})$. The term H_{21} is bounded by

$$H_{21} \leq \sup_l \left\|b_l - \hat{b}_l\right\| \sum_{l,j=1}^{n} \left\|\vartheta_{lj}\right\| \left\|b_j'\right\| = O_p(n^{-1/2}).$$

Next for any $\epsilon > 0$ and $M > \sup \left|b_l\right|$

$$\begin{aligned} P(\left\|H_{22}\right\| &> \epsilon) \leq P\left(\sum_{l,j=1}^{n} \left\|\hat{b}_l\right\| \left\|\vartheta_{lj}\right\| \left\|b_j' - \hat{\Gamma}_j^{\hat{\varepsilon}x}\right\| > \epsilon\right) \\ &\leq P\left(\sum_{l,j=1}^{n} \left\|b_l\right\| \left\|\vartheta_{lj}\right\| \left\|b_j' - \hat{\Gamma}_j^{\hat{\varepsilon}x}\right\| > \epsilon\right) + P\left(\left\|\hat{\beta} - \beta\right\| \sum_{l,j=1}^{n} \left\|\frac{\partial b_l}{\partial \beta}\right\| \left\|\vartheta_{lj}\right\| \left\|b_j' - \hat{\Gamma}_j^{\hat{\varepsilon}x}\right\| > \epsilon\right) \end{aligned}$$

A typical element r in $\hat{\Gamma}_j^{\hat{\varepsilon}x}$ is $n^{-1}\sum_{t=1+j+p+q}^{n} \hat{\varepsilon}_{t-j}(y_{t+q-j-r} - \bar{y}) = n^{-1}\sum_{t=1+j+p+q}^{n} \left(\hat{\varepsilon}_{t-j} - \varepsilon_{t-j}\right)(y_{t+q-j-r} - \bar{y}) + n^{-1}\sum_{t=1+j+p+q}^{n} \varepsilon_{t-j}(y_{t+q-j-r} - \bar{y})$. Following the notation in Kreiss (1987), we set $\theta(L)^{-1} = \sum_{k=0}^{\infty} \zeta_k L^k$. It then follows that

$$\hat{\varepsilon}_t = \sum_{k=0}^{t-p} \hat{\zeta}_k (y_{t-k} - \bar{y} - \hat{\phi}_1(y_{t-1-k} - \bar{y}) - \ldots - \hat{\phi}_p(y_{t-p-k} - \bar{y}))$$

and consequently

$$\hat{\varepsilon}_t - \varepsilon_t = \sum_{k=0}^{t-p}\sum_{i=0}^{p}\left(\zeta_k \phi_i - \hat{\zeta}_k \hat{\phi}_i\right)(y_{t-k-i} - \bar{y}) + \sum_{k=t-p+1}^{\infty}\sum_{i=0}^{p} \zeta_k \phi_i (y_{t-k-i} - \bar{y}).$$

Using these results we can write $b_j' - \hat{\Gamma}_j^{\hat{\varepsilon}x} = b_j' - \hat{\Gamma}_j^{\varepsilon x} + R_1^{\varepsilon x} + R_2^{\varepsilon x} + R_3^{\varepsilon x}$ where

$$R_1^{\varepsilon x} = n^{-1} \sum_{t=1+j+p+q}^{n} \sum_{k=0}^{t-p} \sum_{i=0}^{p} \hat{\phi}_i \left(\zeta_k - \hat{\zeta}_k\right) (y_{t-j-k-i} - \bar{y})(x_t - \bar{x})$$

$$R_2^{\varepsilon x} = \sum_{i=0}^{p} (\phi_i - \hat{\phi}_i) n^{-1} \sum_{t=1+j+p+q}^{n} \sum_{k=0}^{t-p} \zeta_k (y_{t-j-k-i} - \bar{y})(x_t - \bar{x})$$

$$R_2^{\varepsilon x} = n^{-1} \sum_{t=1+j+p+q}^{n} \sum_{k=t-p+1}^{\infty} \sum_{i=0}^{p} \phi_i \zeta_k (y_{t-j-k-i} - \bar{y})(x_t - \bar{x}).$$

We further divide $R_1^{\varepsilon x} = R_{11}^{\varepsilon x} + R_{12}^{\varepsilon x} + R_{13}^{\varepsilon x} + R_{14}^{\varepsilon x}$ where

$$R_{11}^{\varepsilon x} = n^{-1} \sum_{t=1+j+p+q}^{n} \sum_{k=0}^{t-p} \sum_{i=0}^{p} \hat{\phi}_i \left(\zeta_k - \hat{\zeta}_k\right) (y_{t-j-k-i} - \mu_y)(x_t - \mu_x) +$$

$$R_{12}^{\varepsilon x} = (\bar{y} - \mu_y) n^{-1} \sum_{t=1+j+p+q}^{n} \sum_{k=0}^{t-p} \sum_{i=0}^{p} \hat{\phi}_i \left(\zeta_k - \hat{\zeta}_k\right) (x_t - \mu_x)$$

$$R_{13}^{\varepsilon x} = (\bar{x} - \mu_x) n^{-1} \sum_{t=1+j+p+q}^{n} \sum_{k=0}^{t-p} \sum_{i=0}^{p} \hat{\phi}_i \left(\zeta_k - \hat{\zeta}_k\right) (y_{t-j-k-i} - \mu_y)$$

$$R_{14}^{\varepsilon x} = (\bar{x} - \mu_x)(\bar{y} - \mu_y) \frac{n-1-j-p-q}{n} \sum_{k=0}^{t-p} \sum_{i=0}^{p} \hat{\phi}_i \left(\zeta_k - \hat{\zeta}_k\right)$$

where $\|\bar{x} - \mu_x\| = O_p(n^{-1/.2})$, $\|\bar{y} - \mu_y\| = O_p(n^{-1/2})$ and $\left\| n^{-1} \sum_{t=1+j+p+q}^{n} \sum_{k=0}^{t-p} \zeta_k (y_{t-j-k-i} - \mu_y) \right\| = o_p(1)$. We split $R_2^{\varepsilon x}$ and $R_3^{\varepsilon x}$ into similar subparts. By the mean value theorem we obtain $\zeta_k - \hat{\zeta}_k = \frac{\partial \zeta_k}{\partial \theta}|_\theta (\hat{\theta} - \theta) + o(\hat{\theta} - \theta)$. Note that $\frac{\partial \zeta_k}{\partial \theta}|_\theta$ is absolutely summable for $\theta(L)$ invertible. We thus bound

$$\|R_{11}\| \leq \left\|\hat{\theta} - \theta\right\| \left(\sum_{i=0}^{p} \left|\phi_i - \hat{\phi}_i\right| \left\| n^{-1} \sum_{t=1+j+p+q}^{n} \sum_{k=0}^{t-p} \frac{\partial \zeta_k}{\partial \theta}|_\theta (y_{t-j-k-i} - \mu_y)(x_t - \mu_y) \right\| \right.$$

$$\left. + \sum_{i=0}^{p} |\phi_i| \left\| n^{-1} \sum_{t=1+j+p+q}^{n} \sum_{k=0}^{t-p} \frac{\partial \zeta_k}{\partial \theta}|_{\theta=\dot{\theta}} (y_{t-j-k-i} - \bar{y})(x_t - \bar{x}) \right\| \right) + o_p(\hat{\theta} - \theta).$$

whit the same type of expansions holding for R_{12}, R_{13}, R_{14}. Next consider

$$E\left\|n^{-1}\sum_{t=1+j+p+q}^{n}\sum_{k=0}^{t-p}\zeta_k\left[(y_{t-j-k-i}-\mu_y)(x_t-\mu_x)-\Gamma^{yx}_{-(j+k+i)}\right]\right\|^2$$

$$=n^{-2}\sum_{t,s=1+j+p+q}^{n}\sum_{k,l=0}^{t-p}\zeta_k\zeta_l E(y_{s-j-l-i}-\mu_y)(y_{t-j-k-i}-\mu_y)(x_t-\mu_x)(x_s-\mu_x)'$$

$$=n^{-2}\sum_{t,s=1+j+p+q}^{n}\sum_{k,l=0}^{t-p}\zeta_k\zeta_l\left[\Gamma^{yy}_{s-l-t+k}\Gamma^{xx}_{t-s}+\Gamma^{xy}_{s-j-l-i-t}\Gamma^{yx}_{t-j-k-i}+\operatorname{cum}(y_{s-j-l-i},y_{t-j-k-i},x_t,x_s')\right]$$

$$=O(n^{-1})$$

which shows that $\left\|n^{-1}\sum_{t=1+j+p+q}^{n}\sum_{k=0}^{t-p}\zeta_k(y_{t-j-k-i}-\mu_y)(x_t-\mu_x)\right\|=O_p(1)$ uniformly in j.and k Similar arguments establish that $\left\|n^{-1}\sum_{t=1+j+p+q}^{n}\sum_{k=0}^{t-p}\frac{\partial\zeta_k}{\partial\theta}|_{\theta=\dot{\theta}}(y_{t-j-k-i}-\bar{y})(x_t-\bar{x})\right\|=O_p(1)$ uniformly in j and k. This implies that $\sum_{l,j=1}^{n}\|b_l\|\|\vartheta_{lj}\|\left\|\hat{\Gamma}^{\varepsilon x}_j-\hat{\Gamma}^{\hat{\varepsilon}x}_j\right\|=O_p(n^{-1/2})$.

Next consider a typical element r of $\hat{\Gamma}^{\varepsilon x}_j$, $\left[\hat{\Gamma}^{\varepsilon x}_j\right]_r=\left[\hat{\Gamma}^{\varepsilon\bar{x}}_j\right]_r=+(\mu_y-\bar{y})n^{-1}\sum_{t=1+j+p+q}^{n}\varepsilon_{t-j}$ where $\hat{\Gamma}^{\varepsilon\bar{x}}_j=n^{-1}\sum_{t=1+j+p+q}^{n}\varepsilon_{t-j}(x_{t+q}-\mu_x)$. Then $n^{-1}\sum_{t=1+j+p+q}^{n}\varepsilon_{t-j}=O_p(n^{-1/2})$ and $E\varepsilon_{t-j}(y_{t+q-r}-\mu_y)=[b_j]_r$. Then $E\left[b_j-\hat{\Gamma}^{\varepsilon\bar{x}}_j\right]=(j+p+q)/nb_j$. We now bound

$$\sum_{l,j=1}^{n}\|b_l\|\|\vartheta_{lj}\|\left\|b_j'-\hat{\Gamma}^{\varepsilon x}_j\right\| \leq \sum_{l,j=1}^{n}\|b_l\|\|\vartheta_{lj}\|\left(\left\|\hat{\Gamma}^{\varepsilon\bar{x}}_j-E\hat{\Gamma}^{\varepsilon\bar{x}}_j\right\|+(j+p+q)/n\|b_j\|\right)$$
$$+\left|\mu_y-\bar{y}\right|\sum_{l,j=1}^{n}\|b_l\|\|\vartheta_{lj}\|\left\|n^{-1}\sum_{t=1+j+p+q}^{n}\varepsilon_{t-j}\right\|=O_p(n^{-1/2})$$

such that

$$E\left\|\hat{\Gamma}^{\varepsilon\bar{x}}_j-E\hat{\Gamma}^{\varepsilon\bar{x}}_j\right\|^2$$
$$=n^{-2}\sum_{t,s=1+j+p+q}^{n}tr(\gamma_\varepsilon(t-s)\Gamma^{xx}_{s-t}+\Gamma^{\varepsilon x}_{t-s-q-j}\Gamma^{\varepsilon x}_{s-t+j+q}+\operatorname{cum}(\varepsilon_{t-j},\varepsilon_{s-j},x_{t+q},x_{s+q}))=O(n^{-1})$$

which establishes the last equality. We have shown that $H_{22}=O_p(n^{-1/2})$.

Next we consider decomposition of $H_3=n^{-1}\hat{P}_n\Omega_n^{-1}K_MQ'X=n^{-1}\hat{P}_n\left[\Omega_n\right]^{-1}(\hat{\Omega}_n-\Omega_n)\left[\Omega_n\right]^{-1}K_MQ'X$. Then

$$\begin{aligned}H_3 &= H_{31}+H_{32}\\ &= \sum_{l,j=1}^{n}\sum_{h,i=1}^{n}\hat{b}_l\vartheta_{lj}(\hat{\omega}_{j,h}-\omega_{j,h})\vartheta_{h,i}k(\frac{i}{M})(\hat{\Gamma}^{\hat{\varepsilon}x}_i-b_i')\\ &\quad+\sum_{l,j=1}^{n}\sum_{h,i=1}^{n}\hat{b}_l\vartheta_{lj}(\hat{\omega}_{j,h}-\omega_{j,h})\vartheta_{h,i}k(\frac{i}{M})b_i'.\end{aligned}$$

Using a Taylor expansion for $\hat{b}_l$ we find that $H_{32} = \sum_{h,j=1}^n a_j^n(\hat{\omega}_{j,h} - \omega_{j,h})k(\frac{h}{M})a_h'^n + \sum_{l,j=1}^n \sum_{h,i=1}^n \frac{\partial b_l}{\partial \beta}(\hat{\beta} - \beta)\vartheta_{lj}(\hat{\omega}_{j,h} - \omega_{j,h})\vartheta_{h,i}k(\frac{i}{M})b_i'$ where $a_j^n = \sum_{l=1}^n b_l \vartheta_{lj}$. The first term is $O_p(n^{-1/2})$ by Lemma A.16 in MRG. The second term is bounded by $\left\|\hat{\beta} - \beta\right\| \sum_{h,j=1}^n \left\|\sum_{l=1}^n \frac{\partial b_l}{\partial \beta}\vartheta_{lj}\right\| \|\hat{\omega}_{j,h} - \omega_{j,h}\| \|a_h^n\| = O_p(n^{-1})$. For H_{31} we note that as before we decompose $\hat{\Gamma}_i^{\hat{\varepsilon}x} - b_i' = \hat{\Gamma}_i^{\varepsilon x} - b_i' + R_1 + R_2 + R_3$ where the remainder terms are of smaller order uniformly in i. We concentrate on the leading term where

$$\sum_{j=1}^n \sum_{h,i=1}^n E\left\|a_j^n\right\| \|\hat{\omega}_{j,h} - \omega_{j,h}\| \|\vartheta_{h,i}\| \left\|\hat{\Gamma}_i^{\varepsilon x} - b_i'\right\|$$
$$\leq M \sum_{j=1}^n \sum_{h,i=1}^n E\left\|a_j^n\right\| \left(E\|\hat{\omega}_{j,h} - \omega_{j,h}\|^2\right)^{1/2} \|\vartheta_{h,i}\| \left|k(i/M)\right|/M \left(E\left\|\hat{\Gamma}_i^{\varepsilon x} - b_i'\right\|^2\right)^{1/2} = O(M/n)$$

by the Cauchy Schwartz inequality. Then $\left(E\|\hat{\omega}_{j,h} - \omega_{j,h}\|^2\right)^{1/2} = O(n^{-1/2})$ and $\left(E\left\|\hat{\Gamma}_i^{\varepsilon x} - b_i'\right\|^2\right)^{1/2} = O(n^{-1/2})$ uniformly in i.

Next we turn to the analysis of $\hat{d}_M$. From Lemma A.18 in MRG it follows that $d_0 = n^{-1/2}\sum_{t=1}^n \sum_{j_1,j_2=1}^\infty b_{j_1}\vartheta_{j_1,j_2}v_{t,j_2} = O_p(1)$ with $v_{t,j_2} = \varepsilon_{t+q}\varepsilon_{t-j_2}$ and $\lim_n Ed_0d_0' = D$. From Lemma A.19 in MRG it follows that $d_1 = n^{-1/2}\sum_{t=1}^n \sum_{j_1,j_2=n+1}^\infty b_{j_1}\vartheta_{j_1,j_2}v_{t,j_2} = o_p(n^{-2})$ and from Lemma A.20 it follows that $d_2 = n^{-1/2}\sum_{t,j_2=1}^n \sum_{j_1=n+1}^\infty b_{j_1}\vartheta_{j_1,j_2}v_{t,j_2} = o_p(n^{-1})$. The same result obtains for $d_3 = n^{-1/2}\sum_{t,j_1=1}^n \sum_{j_2=n+1}^\infty b_{j_1}\vartheta_{j_1,j_2}v_{t,j_2}$. Next turn to $d_4 = n^{-1/2}\sum_{t,=1}^n \sum_{j_1,j_2=1}^n b_{j_1}\vartheta_{j_1,j_2}(1 - k(j_2/M))v_{t,j_2}$ for which $Ed_4d_4' = O(M^{-2q})$ by Lemma A.22 in MRG. For

$$d_6 = n^{-1/2}\sum_{t,=1}^n \sum_{j_1,j_2=1}^n \left(\hat{b}_{j_1} - b_{j_1}\right)\vartheta_{j_1,j_2}k(j_2/M)v_{t,j_2}$$

we use the Taylor approximation argument to bound

$$\|d_6\| \leq \left\|\hat{\beta} - \beta\right\| \sum_{j_2=1}^n \left\|\sum_{j_1=1}^n \frac{\partial b_{j_1}}{\partial \beta}\vartheta_{j_1 j_2}\right\| \left\|n^{-1/2}\sum_{t=1}^n v_{t,j_2}\right\| + o_p(n^{-1/2})$$

where $E\left\|n^{-1/2}\sum_{t,=1}^n v_{t,j_2}\right\|^2 = O(1)$ and $\left\|\sum_{j_1=1}^n \frac{\partial b_{j_1}}{\partial \beta}\vartheta_{j_1 j_2}\right\|$ is summable in j_2 such that $\|d_6\| = O_p(n^{-1/2})$. Next consider

$$d_7 = n^{-1/2}\sum_{t=1}^n \sum_{j_1,..,j_4=1}^n \left(\hat{b}_{j_1} - b_{j_1}\right)\vartheta_{j_1,j_2}(\hat{\omega}_{j_2,j_3} - \omega_{j_2,j_3})\vartheta_{j_3,j_4}k(j_4/M)v_{t,j_4}$$

which by the previous argument can be bounded by

$$\|d_7\| \leq M\left\|\hat{\beta} - \beta\right\| \sum_{j_2,..,j_4=1}^n \left\|\sum_{j_1=1}^n \frac{\partial b_{j_1}}{\partial \beta}\vartheta_{j_1 j_2}\right\| \|\hat{\omega}_{j_2,j_3} - \omega_{j_2,j_3}\| \left|\vartheta_{j_3,j_4}\right| \left|k(j_4/M)\right|/M \left\|n^{-1/2}\sum_{t=1}^n v_{t,j_4}\right\| = O_p(M/n)$$

from $E\left\|n^{-1/2}\sum_{t,=1}^{n}v_{t,j}\right\|^{2}=\left(1-\frac{j}{n}\right)E\varepsilon_{t+q}^{2}\varepsilon_{t-j}^{2}=O(1)$ uniformly in j and $E\left\|\hat{\omega}_{j_2,j_3}-\omega_{j_2,j_3}\right\|^2=O(n^{-1})$. The term $d_8=n^{-1/2}\sum_{t=1}^{n}\sum_{j_1,..,j_4=1}^{n}b_{j_1}\vartheta_{j_1,j_2}(\hat{\omega}_{j_2,j_3}-\omega_{j_2,j_3})\vartheta_{j_3,j_4}k(j_4/M)v_{t,j_4}=O_p(M/\sqrt{n})$ by similar arguments.

In the same way as for the denominator we consider the remainder terms of $n^{-1/2}\sum_{t=1}^{n}v_{t,j}-\hat{v}_{t,j}=R_1^{\varepsilon\varepsilon}+R_2^{\varepsilon\varepsilon}+R_3^{\varepsilon\varepsilon}$ with $\hat{v}_{t,j}=\varepsilon_{t+q}\hat{\varepsilon}_{t-j}$ where $R_i^{\varepsilon\varepsilon}$ are defined as before replacing x_t with ε_{t+q}. By the same arguments as before it follows that the remainder terms are of lower order and can be neglected.

We now compute the expectations of products and cross products of the largest terms in probability involving M. These include $d_0,d_4,d_7,d_8,H_{12},H_{22},H_{31}$. We have already established that $\lim_n d_0d_0'=D$. Next consider

$$
\begin{aligned}
Ed_4d_4' &= n^{-1}\sum_{t,s=1}^{n}\sum_{j_1,\ldots,j_4=1}^{n}b_{j_1}\vartheta_{j_1,j_2}(1-k(j_2/M))E\left[v_{t,j_2}v_{s,j_3}\right](1-k(j_3/M))\vartheta_{j_3,j_4}b_{j_4}' \\
&= M^{-2q}k_q\sum_{j_1,\ldots,j_4=1}^{\infty}b_{j_1}\vartheta_{j_1,j_2}\left|j_2\right|^q\omega_{j_2,j_3}\left|j_3\right|^q\vartheta_{j_3,j_4}b_{j_4}'+o(M^{-2q})
\end{aligned}
$$

and

$$
\begin{aligned}
Ed_0d_4' &= n^{-1}\sum_{t,s=1}^{n}\sum_{j_1,\ldots,j_4=1}^{n}b_{j_1}\vartheta_{j_1,j_2}E\left[v_{t,j_2}v_{s,j_3}\right](1-k(j_3/M))\vartheta_{j_3,j_4}b_{j_4}' \\
&= n^{-1}M^{-q}\sum_{t,s=1}^{n}\sum_{j_1,\ldots,j_4=1}^{n}b_{j_1}\vartheta_{j_1,j_2}E\left[v_{t,j_2}v_{s,j_3}\right]\left|j_3\right|^q\vartheta_{j_3,j_4}b_{j_4}' \\
&= M^{-q}\sum_{j_1,j_2=1}^{\infty}b_{j_1}\left|j_1\right|^q\vartheta_{j_1,j_2}b_{j_2}'+o(M^{-q}).
\end{aligned}
$$

$$
Ed_8d_8'=n^{-1}\sum_{t,s=1}^{n}\sum_{j_1,..,j_8=1}^{n}b_{j_1}\vartheta_{j_1,j_2}(\hat{\omega}_{j_2,j_3}-\omega_{j_2,j_3})\vartheta_{j_3,j_4}k(j_4/M)v_{t,j_4}v_{s,j_5}\vartheta_{j_5,j_6}k(j_5/M)(\hat{\omega}_{j_6,j_7}-\omega_{j_6,j_7})\vartheta_{j_7,j_8}b_{j_8}
$$

where we focus on the random terms $n^{-1}\sum_{t,s=1}^{n}(\hat{\omega}_{j_2,j_3}-\omega_{j_2,j_3})v_{t,j_4}v_{s,j_5}(\hat{\omega}_{j_6,j_7}-\omega_{j_6,j_7})$ noting that $\hat{\omega}_{j_2,j_3}-\omega_{j_2,j_3}=n^{-1}\sum_t(\varepsilon_{t+q}^2\varepsilon_{t-j_2}\varepsilon_{t-j_3}-E\varepsilon_{t+q}^2\varepsilon_{t-j_2}\varepsilon_{t-j_3})+\max(j_2,j_3)/n\omega_{j_2,j_3}$. Letting

$w_t(j_2,j_3) = \varepsilon_{t+q}^2 \varepsilon_{t-j_2} \varepsilon_{t-j_3} - E\varepsilon_{t+q}^2 \varepsilon_{t-j_2} \varepsilon_{t-j_3}$ we consider

$$
\begin{aligned}
& n^{-1} \sum_{t,s=1}^{n} E(\hat{\omega}_{j_2,j_3} - \omega_{j_2,j_3}) v_{t,j_4} v_{s,j_5} (\hat{\omega}_{j_6,j_7} - \omega_{j_6,j_7}) \\
= \; & n^{-3} \sum_{t,s,q,r}^{n} E w_q(j_2,j_3) v_{t,j_4} v_{s,j_5} w_r(j_6,j_7) \\
= \; & n^{-3} \sum_{t,s,q,r}^{n} E w_q(j_2,j_3) v_{t,j_4} E v_{s,j_5} w_r(j_6,j_7) + E w_q(j_2,j_3) w_r(j_6,j_7) E v_{t,j_4} v_{s,j_5} \\
& + E w_q(j_2,j_3) v_{s,j_5} E v_{t,j_4} w_r(j_6,j_7) + \operatorname{cum}(w_q(j_2,j_3), v_{t,j_4}, v_{s,j_5}, w_r(j_6,j_7)).
\end{aligned}
$$

The last expression is in terms of eights order moments taking the form of products of moments and cumulants. Each of the elements depends on 10 arguments $q, r, s, t, j_2, ..., j_7$. We analyze each individual term. First $E w_t(j_2,j_3) v_{s,j_4} = \operatorname{Cov}(w_t(j_2,j_3), v_{s,j_4})$ which by Brillinger (1980, Theorem 2.3.2) can be written as $\sum_\tau \operatorname{cum}(X_{i,j}; i,j \in \tau_1) ... \operatorname{cum}(X_{i,j}; i,j \in \tau_p)$ where the sum is over all indecomposable partitions $\tau = \tau_1 \cup ... \cup \tau_p$ of the table

$$X = \begin{bmatrix} \varepsilon_{t+q} & \varepsilon_{t+q} & \varepsilon_{t-j_2} & \varepsilon_{t-j_3} \\ \varepsilon_{s+q} & \varepsilon_{s-j_4} & & \end{bmatrix}.$$

In this case there are 12 partitions involving second order cumulants $\operatorname{Cov}(\varepsilon_u, \varepsilon_v)$ only. Because of the martingale structure these products of second order cumulants take the form $\sigma^6 \{t-s\} \{t-s+q+j_4\} \{j_3-j_2\}$ where $\{.\}$ is the indicator function. A term that depends on three indicator functions is summable over 3 of its arguments. The largest terms are therefore the ones that only have two indicator functions. They are $\operatorname{Var}(\varepsilon_{t+q}) \operatorname{Cov}(\varepsilon_{t-j_2}, \varepsilon_{s+q}) \operatorname{Cov}(\varepsilon_{t-j_3}, \varepsilon_{s-j_4}) = \sigma^6 \{t-s-j_2-q\} \{t-s-j_3+j_4\}$ and similarly for $\operatorname{Var}(\varepsilon_{t+q}) \operatorname{Cov}(\varepsilon_{t-j_3}, \varepsilon_{s+q}) \operatorname{Cov}(\varepsilon_{t-j_2}, \varepsilon_{s-j_4})$. Due to the summability assumptions on higher order cumulants all the terms involving cumulants of order larger than two are summable to a larger degree (over more of their arguments). These latter terms can therefore be neglected asymptotically. Next consider $E v_{t,j_4} v_{s,j_5} = \sigma^4 \{t = s\} \{j_4 = j_5\} + \operatorname{cum}(\varepsilon_{t+q}, \varepsilon_{s+q}, \varepsilon_{t-j_4}, \varepsilon_{s-j_5})$ where the first term is summable over t or s and j_4 or j_5 and the cumulant term is summable over t or s, j_4 <u>and</u> j_5. Finally consider $E w_t(j_2,j_3) w_s(j_6,j_7)$ which can be represented as a sum over cumulants of indecomposable partitions of the table

$$X = \begin{bmatrix} \varepsilon_{t+q} & \varepsilon_{t+q} & \varepsilon_{t-j_2} & \varepsilon_{t-j_3} \\ \varepsilon_{s+q} & \varepsilon_{s+q} & \varepsilon_{s-j_6} & \varepsilon_{s-j_7} \end{bmatrix}.$$

The least summable elements in this sum are products of second order cumulants involving $\operatorname{Var}(\varepsilon_{t+q})$ and $\operatorname{Var}(\varepsilon_{s+q})$. The only two candidates are $\sigma^8 \{t-s-j_2+j_7\} \{t-s-j_3+j_6\}$ and

$\sigma^8\{t-s-j_2+j_6\}\{t-s-j_3+j_7\}$. We also note that $\operatorname{cum}(w_v(j_2,j_3),v_{t,j_4},v_{s,j_5},w_r(j_6,j_7))$ is a sum of products of lower order cumulants over indecomposable partitions of the table

$$X=\begin{bmatrix} \varepsilon_{v+q} & \varepsilon_{v+q} & \varepsilon_{v-j_2} & \varepsilon_{v-j_3} \\ \varepsilon_{r+q} & \varepsilon_{r+q} & \varepsilon_{r-j_6} & \varepsilon_{r-j_7} \\ \varepsilon_{t+q} & \varepsilon_{t-j_4} & & \\ \varepsilon_{s+q} & \varepsilon_{s-j_5} & & \end{bmatrix}$$

where the least summable terms are again products of second order cumulants. We now consider individual terms of Ed_8d_8' where we only take into account the least summable moments. For $n^{-3}\sum_{t,s,v,r}^{n}Ew_v(j_2,j_3)v_{t,j_4}Ev_{s,j_5}w_r(j_6,j_7)$ we have

$$n^{-3}\sigma^{12}\sum_{t,v=1}^{n}\sum_{j_1,j_2,j_4=1}^{n}b_{j_1}\vartheta_{j_1,j_2}\vartheta_{j_2+j_4+q,j_4}k(j_4/M)\sum_{j_8,j_7,j_5=1}^{n}k(j_5/M)\vartheta_{j_5,j_7+j_5+q}\vartheta_{j_7,j_8}b_{j_8}.$$

Invertibility of Ω implies that the noncentral diagonals of Ω^{-1} are summable such that

$$\sum_{j_1,j_2,j_4=1}^{\infty}|b_{j_1}\vartheta_{j_1,j_2}\vartheta_{j_2+j_4+q,j_4}|<\infty$$

and the above expression is of order $O(n^{-1})$. For the second case involving $\operatorname{Cov}(\varepsilon_{t-j_3},\varepsilon_{s+q})\operatorname{Cov}(\varepsilon_{t-j_2},\varepsilon_{s-j_4})$ note that $j_3=j_4$ only occurs if $j_2=0$ which is not possible. Therefore, also terms involving this second case are $O(n^{-1})$.

For $Ew_t(j_2,j_3)v_{s,j_5}$ note that the implied restrictions are $[t-s-j_2+q=0,t-s+j_5-j_3=0]$ and $[t-s-j_3+q=0,t-s+j_5-j_2=0]$ in combination with corresponding restrictions implied by $Ev_{t,j_4}w_r(j_6,j_7)$. We find

$$\begin{aligned} &n^{-3}\sigma^{12}\sum_{t,v=1}^{n}\sum_{j_1,j_2,j_4,j_5,j_7,j_8=1}^{n}b_{j_1}\vartheta_{j_1,j_2}\vartheta_{j_2+j_5-q,j_4}k(j_4/M)k(j_5/M)\vartheta_{j_5,j_7+j_4+q}\vartheta_{j_7,j_8}b_{j_8}\\ &\qquad\le n^{-1}\sigma^{12}\sum_{j_1,j_2,j_7,j_8=1}^{n}|b_{j_1}\vartheta_{j_1,j_2}|\,|\vartheta_{j_7,j_8}b_{j_8}|\sum_{j_5}^{\infty}\sum_{j_4}^{\infty}|\vartheta_{j_2+j_5-q,j_4}|\,|\vartheta_{j_5,j_7+j_4+q}|=O(n^{-1})\end{aligned}$$

since ϑ_{x,j_4} is summable over j_4 uniformly in x. Similar arguments apply to the remaining terms of $Ew_t(j_2,j_3)v_{s,j_5}Ev_{t,j_4}w_r(j_6,j_7)$.

For terms involving $Ew_v(j_2,j_3)w_r(j_6,j_7)Ev_{t,j_4}v_{s,j_5}$ we note that for $Ev_{t,j_4}v_{s,j_5}$ the least summable terms imply restrictions of the form $t=s$ and $j_4=j_5$. The least summable terms in $Ew_v(j_2,j_3)w_r(j_6,j_7)$ imply restrictions of the form $v=r-j_6+j_2$ and $j_6=j_2-j_3+j_7$. We thus have

$$n^{-3}\sigma^{12}\sum_{t,r=1}^{n}\sum_{j_1,j_2,j_4,j_7,j_8=1}^{n}b_{j_1}\vartheta_{j_1,j_2}\vartheta_{j_3,j_4}k(j_4/M)^2\vartheta_{j_4,j_2-j_3+j_7}\vartheta_{j_7,j_8}b_{j_8}=O(M/n)$$

because $\vartheta_{j_5,x}$ is summable over j_5 uniformly in x and $\sum_{j_3,j_4}^{n} \vartheta_{j_3,j_4} k(j_4/M)^2 = O(M)$.

Finally we consider $\mathrm{cum}(w_q(j_2,j_3), v_{t,j_4}, v_{s,j_5}, w_r(j_6,j_7))$ where the dominant terms are of the form $\sigma^{12}\{t=s\}\{j_4=j_5\}\{v=r+j_2-j_6\}\{j_6=j_7+j_2-j_3\}$ which as we have seen before implies that the weighted sum over these terms is $O(M/n)$.

For

$$\|d_7\| \leq \sum_{j_1=1}^{n} \left| \frac{\partial b_{j_1}}{\partial \beta}(\hat{\beta}-\beta) \right| \left\| n^{-1/2} \sum_{t=1}^{n} \sum_{j_2,..,j_4=1}^{n} \vartheta_{j_1 j_2} \left(\hat{\omega}_{j_2,j_3} - \omega_{j_2,j_3}\right) \vartheta_{j_3,j_4} k(j_4/M) v_{t,j_4} \right\|$$

where $(\hat{\beta}-\beta) = O_p(n^{-1/2})$ and $\left\| n^{-1/2} \sum_{t=1}^{n} \sum_{j_2,..,j_4=1}^{n} \vartheta_{j_1 j_2} \left(\hat{\omega}_{j_2,j_3} - \omega_{j_2,j_3}\right) \vartheta_{j_3,j_4} k(j_4/M) v_{t,j_4} \right\|^2$ is uniformly $O(M/n)$ by the same argument as for d_8. This shows that d_7 is $O_p(\sqrt{M}/n)$.

The cross term $Ed_0 d_8'$ is given by

$$Ed_0 d_8' = n^{-1} \sum_{t,s=1}^{n} \sum_{j_1,..,j_6=1}^{n} E b_{j_1} \vartheta_{j_1,j_2} (\hat{\omega}_{j_2,j_3} - \omega_{j_2,j_3}) \vartheta_{j_3,j_4} k(j_4/M) v_{t,j_4} v_{s,j_5} \vartheta_{j_5,j_6} b_{j6} = O(n^{-1})$$

where $E\left(\hat{\omega}_{j_2,j_3} - \omega_{j_2,j_3}\right) v_{t,j_4} v_{s,j_5} = n^{-1} \sum_r E w_r(j_2,j_3) v_{t,j_4} v_{s,j_5}$ and $E w_r(j_2,j_3) v_{t,j_4} v_{s,j_5} = \mathrm{Cov}(w_r(j_2,j_3), v_{t,j_4} v_{s,j_5})$ and the covariance term can be written as the sum over products of lower order cumulants over the indecomposable partitions of the table

$$X = \begin{bmatrix} \varepsilon_{r+q} & \varepsilon_{r+q} & \varepsilon_{r-j_2} & \varepsilon_{r-j_3} \\ \varepsilon_{t+q} & \varepsilon_{t-j_4} & & \\ \varepsilon_{s+q} & \varepsilon_{s-j_5} & & \end{bmatrix}$$

such that the largest terms imply restrictions of the from

$$[t=s, j_4=j_5, j_2=j_3]$$

and

$$[t=s, r=t+j_2-j_4, j_4=j_5-j_3+j_2].$$

Summing over these terms leads to the results.

Next,

$$\begin{aligned} &EH_{22}D^{-1}d_0 d_0' D^{-1} H_{22} \\ &= n^{-1} \sum_{l,j=1}^{n} \sum_{t,s=1}^{n} \sum_{j_0,\ldots,j_5=1}^{\infty} E b_l \vartheta_{l j_0} k(\frac{j_0}{M}) \left(b_{j_0}' - \hat{\Gamma}_{j_0}^{\varepsilon x}\right) D^{-1} b_{j_1} \vartheta_{j_1,j_2} v_{t,j_2} v_{s,j_3} \vartheta_{j_3,j_4} b_{j_4} D^{-1} \left(b_{j_5}' - \hat{\Gamma}_{j_5}^{\varepsilon x}\right) k(\frac{j_5}{M}) \vartheta_{j_5,m} b_m. \end{aligned}$$

Let the elements of the vector $D^{-1}b_{j_1}$ be denoted by d_{k,j_1} and the elements of $b'_{j0} - \hat{\Gamma}^{\varepsilon x}_{j0}$ by $\hat{\gamma}^{\varepsilon x}_{k,j0}$. Then we have

$$\begin{aligned}
& EH_{22}D^{-1}d_0d_0'D^{-1}H_{22} \\
&= n^{-1}\sum_{l,j=1}^{n}\sum_{t,s=1}^{n}\sum_{j_0,\dots,j_5=1}^{n}\sum_{k_1,k_2}^{p} b_l\vartheta_{lj_0}d_{k_1,j_1}\vartheta_{j_1,j_2}E\left[\hat{\gamma}^{\varepsilon x}_{k_1j0}\hat{\gamma}^{\varepsilon x}_{k_2j5}v_{t,j_2}v_{s,j_3}\right]\vartheta_{j_3,j_4}d_{k_2,j_4}\vartheta_{j_5,m}b_m \\
&\qquad \le Cn^{-1}\sum_{l,j=1}^{n}\sum_{j_0,\dots,j_5=1}^{n}\sum_{k_1,k_2}^{p}\|b_l\vartheta_{lj_0}\|\,|d_{k_1,j_1}\vartheta_{j_1,j_2}|\,|\vartheta_{j_3,j_4}d_{k_2,j_4}|\,\|\vartheta_{j_5,m}b_m\| = O(n^{-1})
\end{aligned}$$

where $\hat{\gamma}^{\varepsilon x}_{k,j0} = n^{-1}\sum_t\left(y_{t+q-k+1}\varepsilon_{t-j_0} - b_{j0}\right)$ and letting $w^x_t(j_0,k) = \left(y_{t+q-k+1}\varepsilon_{t-j_0} - b_{j0}\right)$ we can write

$$\begin{aligned}
E\left[\hat{\gamma}^{\varepsilon x}_{k_1j0}\hat{\gamma}^{\varepsilon x}_{k_2j5}v_{t,j_2}v_{s,j_3}\right] &= n^{-2}\sum_{v,r}Ew^x_v(j_0,k_1)w^x_r(j_5,k_2)v_{t,j_2}v_{s,j_3} \\
&= n^{-2}\sum_{v,r}\left[Ew^x_v(j_0,k_1)w^x_r(j_5,k_2)Ev_{t,j_2}v_{s,j_3} + Ew^x_r(j_5,k_2)v_{t,j_2}Ew^x_v(j_0,k_1)v_{s,j_3}\right. \\
&\quad \left. + Ew^x_r(j_5,k_2)v_{t,j_2}Ew^x_v(j_0,k_1)v_{s,j_3} + \operatorname{cum}(w^x_v(j_0,k_1),w^x_r(j_5,k_2),v_{t,j_2},v_{s,j_3})\right].
\end{aligned}$$

Then

$$\begin{aligned}
Ew^x_v(j_0,k_1)w^x_r(j_5,k_2) &= Ey_{v+q-k_1+1}\varepsilon_{r-j_5}Ey_{r+q-k_2+1}\varepsilon_{v-j_0} \\
&\quad + Ey_{v+q-k_1+1}y_{r+q-k_2+1}E\varepsilon_{r-j_5}\varepsilon_{v-j_0} \\
&\quad + \operatorname{cum}(y_{v+q-k_1+1},y_{r+q-k_2+1},\varepsilon_{r-j_5},\varepsilon_{t-j_0})
\end{aligned}$$

and

$$\begin{aligned}
Ew^x_r(j_5,k_2)v_{t,j_2} &= Ey_{r+q-k_2+1}\varepsilon_{t-j_2}E\varepsilon_{r-j_5}\varepsilon_{t+q} \\
&\quad + Ey_{r+q-k_2+1}\varepsilon_{t+q}E\varepsilon_{r+q-k_1+1}\varepsilon_{t-j_2} \\
&\quad + \operatorname{cum}(y_{r+q-k_2+1},\varepsilon_{t+q},\varepsilon_{r+q-k_1+1},\varepsilon_{t-j_2})
\end{aligned}$$

while $\operatorname{cum}(w^x_v(j_0,k_1),w^x_r(j_5,k_2),v_{t,j_2},v_{s,j_3})$ can be represented as a sum of products of lower order cumulants from indecomposable partitions of the table

$$X = \begin{bmatrix} y_{v+q-k_1+1} & \varepsilon_{v-j_0} \\ y_{r+q-k_2+1} & \varepsilon_{r-j_5} \\ \varepsilon_{t+q} & \varepsilon_{t-j_2} \\ \varepsilon_{s+q} & \varepsilon_{s-j_3} \end{bmatrix}$$

which shows that $E\left[\hat{\gamma}^{\varepsilon x}_{k_1 j0}\hat{\gamma}^{\varepsilon x}_{k_2 j5} v_{t,j_2} v_{s,j_3}\right] = O(n^{-1})$ because all the moment restrictions imply at least one restriction of the form $r = v$ and $t = s$.

Next consider

$$EH_{22}D^{-1}d_0 d_0' = n^{-1}\sum_{l,j=1}^{n}\sum_{t,s=1}^{n}\sum_{j_0,\dots,j_5=1}^{\infty} Eb_l \vartheta_{lj_0} k(\frac{j_0}{M})\left(b_{j_0}' - \hat{\Gamma}^{\varepsilon x}_{j_0}\right) D^{-1} b_{j_1}\vartheta_{j_1,j_2} v_{t,j_2} v_{s,j_3} \vartheta_{j_3,j_4} b_{j_4}.$$

which depends on $E\left[\hat{\gamma}^{\varepsilon x}_{kj0} v_{t,j_2} v_{s,j_3}\right] = n^{-1}\sum_v \operatorname{Cov}(w^x_v(j_0,k_1), v_{t,j_2} v_{s,j_3})$ where the covariance depends on partitions of the table

$$X = \begin{bmatrix} y_{v+q-k_1+1} & \varepsilon_{v-j_0} & & \\ \varepsilon_{t+q} & \varepsilon_{s+q} & \varepsilon_{t-j_2} & \varepsilon_{s-j_3} \end{bmatrix}$$

which imply restrictions on t and v such that $EH_{22}D^{-1}d_0 d_0' = O(n^{-1})$.

The remaining cross products involving $H_{22}D^{-1}d_0$ are of smaller order. For H_{12} we consider

$$\begin{aligned} & EH_{12}D^{-1}d_0 d_0' \\ = \;& n^{-1}M^{-q}k_q \sum_{l,j=1}^{n}\sum_{t,s=1}^{n}\sum_{j_1,\dots,j_4=1}^{\infty} b_l \left|l\right|^q \vartheta_{lj} b_j' D^{-1} b_{j_1}\vartheta_{j_1,j_2} E\left[v_{t,j_2} v_{s,j_3}\right]\vartheta_{j_3,j_4} b_{j_4} + o(M^{-q}) \\ = \;& M^{-q}k_q \sum_{l,j=1}^{\infty} b_l \left|l\right|^q \vartheta_{lj} b_j' + o(M^{-q}) \end{aligned}$$

while

$$H_{12}D^{-1}d_0 d_0' D^{-1}H_{12} = M^{-2q}k_q^2 \left(\sum_{l,j=1}^{\infty} b_l \vartheta_{lj}\left|j\right|^q b_j'\right)\left(\sum_{l,j=1}^{\infty} b_l \left|l\right|^q \vartheta_{lj} b_j'\right) + o(M^{-2q}).$$

Cross products with d_7, d_8 are of lower order. Finally, examine $EH_{22}D^{-1}d_0 d_4'$ which is

$$\begin{aligned} EH_{22}D^{-1}d_0 d_4' &= n^{-1}M^{-2q}k_q^2 \sum_{l,j=1}^{n}\sum_{t,s=1}^{n}\sum_{j_1,\dots,j_4=1}^{\infty} b_l \left|l\right|^q \vartheta_{lj} b_j' D^{-1} b_{j_1}\vartheta_{j_1,j_2} E\left[v_{t,j_2} v_{s,j_3}\right]\left|j_3\right|^q \vartheta_{j_3,j_4} b_{j_4}' \\ &= M^{-2q}k_q^2 \left(\sum_{l,j=1}^{\infty} b_l \left|l\right|^q \vartheta_{lj} b_j'\right) D^{-1}\left(\sum_{l,j=1}^{\infty} b_l \vartheta_{lj}\left|j\right|^q b_j'\right) + o(M^{-2q}) \end{aligned}$$

Now consider $H_{31}D^{-1}d_0 d_0$ where

$$EH_{31}D^{-1}d_0 d_0 = n^{-3}\sum_{t,s,v,r}\sum_{j_1,\dots j_8=1}^{n}\sum_{k}^{p} a^n_{j_1} Ew_r(j_1,j_2)\vartheta_{j_2,j_3} k(\frac{j_3}{M}) w^x_v(j_3,k) d_{k,j_4}\vartheta_{j_4,j_5} v_{t,j_5} v_{s,j_6}\vartheta_{j_6,j_7} b_{j_7}$$

and

$$E\left[w_r(j_1,j_2)w_v^x(j_3,k)v_{t,j_5}v_{s,j_6}\right] = Ew_r(j_1,j_2)w_v^x(j_3,k)Ev_{t,j_5}v_{s,j_6} + Ew_r(j_1,j_2)v_{s,j_6}Ev_{t,j_5}w_v^x(j_3,k) + Ew_r(j_1,j_2)v_{t,j_5}Ev_{s,j_6}w_v^x(j_3,k) + \text{cum}(w_r(j_1,j_2),v_{t,j_5},v_{s,j_6},w_v^x(j_3,k))$$

where from previous results the least summable cumulants of $Ew_r(j_2,j_3)v_{s,j_6}$ imply a restriction between j_6 and j_2 which makes this term of lower order. The same argument holds for $Ew_r(j_1,j_2)v_{t,j_5}$. For $Ew_r(j_1,j_2)w_v^x(j_3,k) = \text{Cov}(w_r(j_1,j_2),w_v^x(j_3,k))$ we consider the table

$$X = \begin{bmatrix} y_{v+q-k+1} & \varepsilon_{v-j_3} & & \\ \varepsilon_{t+q} & \varepsilon_{t+q} & \varepsilon_{t-j_1} & \varepsilon_{t-j_2} \end{bmatrix}$$

where the least summable cumulants of second order are $\sigma^6\left\{v-j_3-t+j_1\right\}b_{v-t+j_2+q-k+1}$ which is summable over j_2. Finally consider $\text{cum}(w_r(j_1,j_2),v_{t,j_5},v_{s,j_6},w_v^x(j_3,k))$ which depends on the table

$$X = \begin{bmatrix} y_{v+q-k+1} & \varepsilon_{v-j_3} & & \\ \varepsilon_{r+q} & \varepsilon_{r+q} & \varepsilon_{r-j_1} & \varepsilon_{r-j_2} \\ \varepsilon_{t+q} & \varepsilon_{t-j_5} & & \\ \varepsilon_{s+q} & \varepsilon_{s-j_6} & & \end{bmatrix}$$

where the least summable terms imply restrictions on s,r and j_3 which implies that the terms involving $\text{cum}(w_r(j_1,j_2),v_{t,j_5},v_{s,j_6},w_v^x(j_3,k))$ are summable. This implies that $EH_{31}D^{-1}d_0d_0 = O(n^{-1})$. Consequently, $H_{31}D^{-1}d_0d_0'D^{-1}H_{31} = o_p(n^{-1})$, $H_{31}D^{-1}d_0d_7' = o_p(n^{-1})$ and $H_{31}D^{-1}d_0d_8' = o_p(n^{-1})$.

To summarize these results, the largest term depending positively on M is $Ed_8d_8' = O(M/n)$ while the largest terms depending inversely on M are d_4 and $H_{12}D^{-1}d_0$ where

$$E(d_4+H_{13}D^{-1}d_0)(d_4+H_{13}D^{-1}d_0)' = M^{-2q}k_q^2\left(\sum_{l,j=1}^{\infty} b_l\vartheta_{lj}\left|j\right|^q b_j'\right)(I+D^{-1})\left(\sum_{l,j=1}^{\infty} b_l\left|l\right|^q\vartheta_{lj}b_j'\right).$$

■

Proof of Corollary 1: Here we only need to consider the largest terms in M of the previous result, i.e.

$$Ed_8d_8' = n^{-1}\sum_{t,s=1}^{n}\sum_{j_1,j_2=1}^{n} Eb_{j_1}\vartheta_{j_1,j_1}^2(\hat{\omega}_{j_1,j_1}-\omega_{j_1,j_1})k(j_1/M)v_{t,j_1}v_{s,j_2}k(j_2/M)(\hat{\omega}_{j_2,j_2}-\omega_{j_2,j_2})\vartheta_{j_2,j_2}^2 b_{j_2}$$

where now $\sum_{j_1,j_2=1}^{n} b_{j_1}\vartheta^2_{j_1,j_1}k(j_1/M)k(j_2/M)\vartheta^2_{j_2,j_2}b_{j_2} < \left(\sum_{j_1=1}^{n}\left|b_{j_1}\vartheta^2_{j_1,j_1}\right|\right)^2 < \infty$ is absolutely summable. This holds even though $\vartheta^2_{j_1,j_1}$ itself is not summable. Using the results of the previous proofs this implies that $Ed_8d_8 = O(n^{-1})$.■

References

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

ANDREWS, D. W., AND J. MONAHAN (1992): "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, 60(4), 953–966.

BATES, C., AND H. WHITE (1990): "Efficient instrumental variables estimation of systems of implicit heterogeneous nonlinear dynamic equations with nonspherical errors.," in *Dynamic Econometric Modeling*, ed. by E. B. W. Barnett, and H. White, pp. 3–25. Cambridge University Press.

BOLLERSLEV, T., R. Y. CHOU, AND K. KRONER (1992): "ARCH modeling in finance," *Journal of Econometrics*, 52, 5–59.

DONALD, S. G., AND W. K. NEWEY (1997): "Choosing the Number of Instruments," *Draft, MIT.*

GUO, B., AND P. C. PHILLIPS (1997): "Efficient Estimation of Secon Moment Parameters in the ARCH Model," *Manuscript Yale University.*

HANSEN, L. P. (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., AND K. J. SINGLETON (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 Heterogeneity," *MIT Working Paper No. 99-08.*

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

——— (2000): "Mean Squared Error Reduction for GMM Estimators of Linear Time Series Models," *unpublished manuscript.*

Newey, W. K. (1990): "Efficient Instrumental Variables Estimation of Nonlinear Models," *Econometrica*, 58, 809–837.

Newey, W. K., and K. D. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(3), 703–708.

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.

White, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator And a Direct Test for Heteroskedasticity," *Econometrica*, 48(4), 817–838.

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

Table 1

	Quantiles							
	Sample Size: 128, $\phi_1 = 0.0$							
	.01	.1	.5	.9	.99	MAE	MSE	Var
OLS	-0.3418	-0.1778	-0.0004	0.1809	0.3126	0.1113	0.0198	0.0198
IVy 1	-0.3418	-0.1778	-0.0004	0.1809	0.3126	0.1113	0.0198	0.0198
IVe 1	-0.3469	-0.1804	0.0009	0.1767	0.3434	0.1130	0.0206	0.0206
IVky 1	-0.3418	-0.1778	-0.0004	0.1809	0.3126	0.1113	0.0198	0.0198
IVke 1	-0.3469	-0.1804	0.0009	0.1767	0.3434	0.1130	0.0206	0.0206
IVp 1	-0.3469	-0.1804	0.0009	0.1767	0.3434	0.1130	0.0206	0.0206
IVy 2	-0.3403	-0.1833	-0.0008	0.1780	0.3208	0.1121	0.0201	0.0201
IVe 2	-0.3286	-0.1868	-0.0008	0.1731	0.3407	0.1124	0.0203	0.0203
IVky 2	-0.3391	-0.1757	0.0005	0.1764	0.3120	0.1112	0.0197	0.0197
IVke 2	-0.3447	-0.1828	0.0016	0.1750	0.3434	0.1128	0.0205	0.0205
IVpe 2	-0.3915	-0.1939	-0.0002	0.1706	0.3504	0.1128	0.0209	0.0208
IVy 3	-0.3527	-0.1874	-0.0028	0.1714	0.3226	0.1118	0.0202	0.0202
IVe 3	-0.3359	-0.1898	-0.0007	0.1672	0.3379	0.1123	0.0204	0.0204
IVky 3	-0.3416	-0.1800	0.0001	0.1759	0.3120	0.1120	0.0200	0.0200
IVke 3	-0.3394	-0.1860	-0.0013	0.1725	0.3389	0.1129	0.0204	0.0204
IVpe 3	-0.3865	-0.1869	-0.0025	0.1723	0.3500	0.1136	0.0215	0.0215
IVy 4	-0.3522	-0.1880	-0.0017	0.1714	0.3195	0.1124	0.0204	0.0204
IVe 4	-0.3313	-0.1942	-0.0007	0.1714	0.3440	0.1132	0.0206	0.0206
IVky 4	-0.3484	-0.1804	0.0000	0.1775	0.3178	0.1122	0.0202	0.0202
IVke 4	-0.3376	-0.1880	-0.0016	0.1731	0.3388	0.1127	0.0204	0.0204
IVpe 4	-0.3897	-0.1911	-0.0035	0.1724	0.3481	0.1150	0.0221	0.0220
IVy 5	-0.3488	-0.1866	-0.0023	0.1708	0.3451	0.1143	0.0208	0.0208
IVe 5	-0.3319	-0.1931	-0.0027	0.1743	0.3632	0.1150	0.0211	0.0210
IVky 5	-0.3524	-0.1837	0.0006	0.1762	0.3252	0.1122	0.0202	0.0203
IVke 5	-0.3370	-0.1876	-0.0013	0.1732	0.3431	0.1126	0.0204	0.0204
IVpe 5	-0.3950	-0.1919	-0.0055	0.1781	0.3551	0.1178	0.0233	0.0232
IVy 10	-0.3371	-0.1894	-0.0042	0.1820	0.3294	0.1157	0.0213	0.0213
IVe 10	-0.3510	-0.1920	-0.0040	0.1783	0.3657	0.1157	0.0216	0.0216
IVky 10	-0.3482	-0.1845	-0.0009	0.1786	0.3303	0.1121	0.0201	0.0201
IVke 10	-0.3340	-0.1869	-0.0013	0.1718	0.3465	0.1120	0.0202	0.0202
IVpe 10	-0.4318	-0.1960	-0.0069	0.1771	0.3978	0.1213	0.0247	0.0247
IV FD	-0.3415	-0.1743	0.0022	0.1750	0.3090	0.1094	0.0192	0.0192

Table 2

	Quantiles							
	Sample Size: 128, $\phi_1 = 0.3$							
	.01	.1	.5	.9	.99	MAE	MSE	Var
OLS	-0.3090	-0.1704	-0.0025	0.1460	0.2996	0.0987	0.0157	0.0157
IVy 1	-0.3090	-0.1704	-0.0025	0.1460	0.2996	0.0987	0.0157	0.0157
IVe 1	-0.3192	-0.1741	0.0007	0.1708	0.3405	0.1076	0.0185	0.0185
IV ky 1	-0.3090	-0.1704	-0.0025	0.1460	0.2996	0.0987	0.0157	0.0157
IV ke 1	-0.3192	-0.1741	0.0007	0.1708	0.3405	0.1076	0.0185	0.0185
IV pe 1	-0.3192	-0.1741	0.0007	0.1708	0.3405	0.1076	0.0185	0.0185
IVy 2	-0.3111	-0.1690	-0.0013	0.1462	0.2581	0.0984	0.0155	0.0154
IVe 2	-0.3297	-0.1732	-0.0003	0.1641	0.2873	0.1047	0.0172	0.0172
IVky 2	-0.3056	-0.1677	-0.0016	0.1462	0.2968	0.0987	0.0156	0.0156
IVke 2	-0.3214	-0.1702	-0.0002	0.1699	0.3330	0.1067	0.0182	0.0182
IVpe 2	-0.3305	-0.1777	-0.0016	0.1682	0.3122	0.1069	0.0181	0.0181
IVy 3	-0.3273	-0.1661	-0.0037	0.1469	0.2707	0.0998	0.0158	0.0158
IVe 3	-0.3188	-0.1698	0.0041	0.1625	0.2813	0.1053	0.0172	0.0172
IVky 3	-0.3061	-0.1664	-0.0007	0.1468	0.2866	0.0990	0.0156	0.0156
IVke 3	-0.3182	-0.1720	-0.0017	0.1664	0.3120	0.1058	0.0177	0.0177
IVpe 3	-0.3280	-0.1814	-0.0053	0.1712	0.3022	0.1111	0.0276	0.0276
IVy 4	-0.3277	-0.1664	-0.0011	0.1493	0.2681	0.0995	0.0158	0.0158
IVe 4	-0.3328	-0.1748	0.0012	0.1617	0.2784	0.1051	0.0173	0.0173
IVky 4	-0.3043	-0.1668	-0.0002	0.1484	0.2798	0.0991	0.0156	0.0156
IVke 4	-0.3172	-0.1708	0.0003	0.1646	0.3132	0.1052	0.0174	0.0174
IVpe 4	-0.3530	-0.1853	-0.0067	0.1742	0.3189	0.1166	0.0783	0.0782
IVy 5	-0.3291	-0.1691	-0.0017	0.1489	0.2702	0.1001	0.0160	0.0160
IVe 5	-0.3326	-0.1724	-0.0012	0.1642	0.2784	0.1054	0.0174	0.0174
IVky 5	-0.3114	-0.1687	0.0001	0.1479	0.2813	0.0993	0.0157	0.0157
IVke 5	-0.3211	-0.1719	0.0004	0.1646	0.3078	0.1049	0.0174	0.0174
IVpe 5	-0.3577	-0.1765	-0.0062	0.1733	0.3095	0.1119	0.0223	0.0223
IVy 10	-0.3246	-0.1687	-0.0009	0.1449	0.2743	0.1005	0.0161	0.0161
IVe 10	-0.3235	-0.1674	-0.0015	0.1586	0.2743	0.1046	0.0171	0.0171
IVky 10	-0.3149	-0.1684	0.0043	0.1494	0.2698	0.0994	0.0157	0.0157
IVke 10	-0.3268	-0.1691	-0.0000	0.1632	0.2758	0.1046	0.0171	0.0171
IVpe 10	-0.3472	-0.1848	0.0004	0.1768	0.3085	0.1166	0.0318	0.0318
IV FD	-0.3199	-0.1725	-0.0079	0.1393	0.2870	0.0977	0.0154	0.0152

Table 3

	Quantiles							
	Sample Size: 128, $\phi_1 = 0.6$							
	.01	.1	.5	.9	.99	MAE	MSE	Var
OLS	-0.2994	-0.1597	-0.0183	0.0933	0.1698	0.0790	0.0108	0.0102
IV y 1	-0.2994	-0.1597	-0.0183	0.0933	0.1698	0.0790	0.0108	0.0102
IV e 1	-0.3416	-0.1823	-0.0122	0.1583	0.2756	0.1056	0.0184	0.0182
IV ky 1	-0.2994	-0.1597	-0.0183	0.0933	0.1698	0.0790	0.0108	0.0102
IV ke 1	-0.3416	-0.1823	-0.0122	0.1583	0.2756	0.1056	0.0184	0.0182
IV pe1	-0.3416	-0.1823	-0.0122	0.1583	0.2756	0.1056	0.0184	0.0182
IVy 2	-0.2984	-0.1557	-0.0194	0.0950	0.1590	0.0781	0.0107	0.0100
IVe 2	-0.3451	-0.1729	-0.0163	0.1211	0.2173	0.0914	0.0144	0.0139
IVky 2	-0.3075	-0.1581	-0.0184	0.0928	0.1736	0.0797	0.0110	0.0104
IVke 2	-0.3363	-0.1815	-0.0121	0.1503	0.2581	0.1031	0.0177	0.0174
IVpe 2	-0.3389	-0.1709	-0.0196	0.1265	0.2486	0.0958	0.0154	0.0149
IVy 3	-0.2953	-0.1543	-0.0182	0.0934	0.1661	0.0779	0.0106	0.0100
IVe 3	-0.3089	-0.1679	-0.0155	0.1205	0.2076	0.0889	0.0135	0.0130
IVky 3	-0.2984	-0.1560	-0.0169	0.0941	0.1670	0.0783	0.0107	0.0101
IVke 3	-0.3312	-0.1770	-0.0122	0.1387	0.2286	0.0970	0.0158	0.0155
IVpe 3	-0.3308	-0.1724	-0.0162	0.1245	0.2370	0.0940	0.0147	0.0143
IVy 4	-0.2918	-0.1527	-0.0167	0.0919	0.1696	0.0779	0.0105	0.0100
IVe 4	-0.3158	-0.1625	-0.0176	0.1156	0.1972	0.0867	0.0128	0.0123
IVky 4	-0.2967	-0.1542	-0.0130	0.0932	0.1654	0.0777	0.0105	0.0100
IVke 4	-0.3293	-0.1707	-0.0124	0.1274	0.2215	0.0933	0.0148	0.0144
IVpe 4	-0.3168	-0.1667	-0.0135	0.1277	0.2236	0.0922	0.0140	0.0137
IVy 5	-0.2947	-0.1530	-0.0162	0.0921	0.1724	0.0788	0.0108	0.0102
IVe 5	-0.3212	-0.1640	-0.0168	0.1125	0.1929	0.0871	0.0130	0.0125
IVky 5	-0.2950	-0.1522	-0.0134	0.0957	0.1629	0.0776	0.0105	0.0100
IVke 5	-0.3437	-0.1682	-0.0147	0.1252	0.2165	0.0910	0.0142	0.0138
IVpe 5	-0.3440	-0.1684	-0.0110	0.1237	0.2149	0.0924	0.0144	0.0141
IVy 10	-0.2929	-0.1591	-0.0154	0.1010	0.1735	0.0807	0.0111	0.0105
IVe 10	-0.3127	-0.1700	-0.0175	0.1123	0.2048	0.0874	0.0128	0.0122
IVky 10	-0.2838	-0.1539	-0.0138	0.0967	0.1664	0.0781	0.0105	0.0100
IVke 10	-0.3004	-0.1629	-0.0167	0.1189	0.2023	0.0879	0.0131	0.0126
IVpe 10	-0.3817	-0.1735	-0.0149	0.1220	0.2332	0.0962	0.0159	0.0155
IV FD	-0.2943	-0.1575	-0.0234	0.0838	0.1586	0.0781	0.0106	0.0097

Table 4

	Quantiles							
	Sample Size: 128, $\phi_1 = 0.9$							
	.01	.1	.5	.9	.99	MAE	MSE	Var
OLS	-0.1687	-0.0877	-0.0058	0.0390	0.0637	0.0403	0.0031	0.0028
IV y 1	-0.1687	-0.0877	-0.0058	0.0390	0.0637	0.0403	0.0031	0.0028
IV e 1	-0.3146	-0.1806	-0.0054	0.1670	0.3405	0.1101	0.0197	0.0197
IV ky 1	-0.1687	-0.0877	-0.0058	0.0390	0.0637	0.0403	0.0031	0.0028
IV ke 1	-0.3146	-0.1806	-0.0054	0.1670	0.3405	0.1101	0.0197	0.0197
IV pe 1	-0.3146	-0.1806	-0.0054	0.1670	0.3405	0.1101	0.0197	0.0197
IVy 2	-0.1654	-0.0842	-0.0067	0.0370	0.0605	0.0377	0.0027	0.0024
IVe 2	-0.2556	-0.1280	-0.0098	0.0896	0.1674	0.0691	0.0080	0.0078
IVky 2	-0.1682	-0.0869	-0.0070	0.0403	0.0693	0.0417	0.0032	0.0029
IVke 2	-0.3113	-0.1675	-0.0052	0.1534	0.2771	0.1008	0.0166	0.0165
IVpe 2	-0.2524	-0.1203	-0.0068	0.1005	0.1846	0.0700	0.0081	0.0081
IVy 3	-0.1676	-0.0797	-0.0072	0.0365	0.0618	0.0372	0.0026	0.0024
IVe 3	-0.1976	-0.1069	-0.0077	0.0779	0.1462	0.0581	0.0056	0.0055
IVky 3	-0.1607	-0.0858	-0.0087	0.0350	0.0605	0.0381	0.0028	0.0025
IVke 3	-0.2663	-0.1408	-0.0071	0.1103	0.2064	0.0805	0.0108	0.0107
IVpe 3	-0.1968	-0.1037	-0.0036	0.0831	0.1552	0.0583	0.0057	0.0057
IVy 4	-0.1703	-0.0817	-0.0060	0.0370	0.0631	0.0374	0.0027	0.0024
IVe 4	-0.1770	-0.1002	-0.0059	0.0634	0.1235	0.0511	0.0045	0.0044
IVky 4	-0.1526	-0.0840	-0.0076	0.0364	0.0623	0.0374	0.0027	0.0024
IVke 4	-0.2335	-0.1260	-0.0085	0.0952	0.1830	0.0711	0.0083	0.0082
IVpe 4	-0.1870	-0.0943	-0.0020	0.0692	0.1317	0.0522	0.0071	0.0070
IVy 5	-0.1674	-0.0796	-0.0067	0.0375	0.0649	0.0378	0.0027	0.0024
IVe 5	-0.1898	-0.0951	-0.0064	0.0612	0.1158	0.0487	0.0042	0.0040
IVky 5	-0.1515	-0.0815	-0.0078	0.0375	0.0632	0.0369	0.0026	0.0023
IVke 5	-0.2363	-0.1154	-0.0076	0.0868	0.1645	0.0653	0.0071	0.0069
IVpe 5	-0.1854	-0.0886	-0.0006	0.0677	0.1272	0.0500	0.0094	0.0094
IVy 10	-0.1609	-0.0800	-0.0068	0.0391	0.0684	0.0378	0.0026	0.0024
IVe 10	-0.1593	-0.0881	-0.0049	0.0468	0.0856	0.0421	0.0033	0.0031
IVky 10	-0.1587	-0.0770	-0.0034	0.0395	0.0632	0.0362	0.0025	0.0023
IVke 10	-0.1877	-0.0972	-0.0064	0.0669	0.1252	0.0528	0.0048	0.0047
IVpe 10	-0.1640	-0.0770	0.0011	0.0524	0.0914	0.0416	0.0030	0.0030
IV FD	-0.1829	-0.0910	-0.0151	0.0289	0.0552	0.0403	0.0031	0.0026