Econometrica, Vol. 73, No. 3 (May, 2005), 983–1002

GMM, GEL, SERIAL CORRELATION, AND ASYMPTOTIC BIAS

BY STANISLAV ANATOLYEV[1]

For stationary time series models with serial correlation, we consider generalized method of moments (GMM) estimators that use heteroskedasticity and autocorrelation consistent (HAC) positive definite weight matrices and generalized empirical likelihood (GEL) estimators based on smoothed moment conditions. Following the analysis of Newey and Smith (2004) for independent observations, we derive second order asymptotic biases of these estimators. The inspection of bias expressions reveals that the use of smoothed GEL, in contrast to GMM, removes the bias component associated with the correlation between the moment function and its derivative, while the bias component associated with third moments depends on the employed kernel function. We also analyze the case of no serial correlation, and find that the seemingly unnecessary smoothing and HAC estimation can reduce the bias for some of the estimators.

KEYWORDS: GMM, empirical likelihood, higher order asymptotic expansions, asymptotic bias, serial correlation, HAC estimation, smoothed moment conditions.

1. INTRODUCTION

IN RECENT YEARS, one step generalized empirical likelihood (GEL) estimators (Smith (1997, 2001)) have attracted attention as theoretically appealing alternatives to generalized method of moments (GMM) estimators. These estimators include empirical likelihood (EL) (Qin and Lawless (1994), Imbens (1997)), exponential tilting (Kitamura and Stutzer (1997)), continuously updating (CU) GMM (Hansen, Heaton, and Yaron (1996)), and other members (Imbens, Spady, and Johnson (1998)). It has been established that the first order asymptotic properties of GEL estimators are identical to those of GMM estimators (Smith (1997)), but their higher order asymptotic properties are advantageous. In particular, Newey and Smith (2004) recently found that the second order asymptotic bias of GEL estimators lacks some components that are characteristic of GMM estimators when observations are independent. The EL estimator is most distinctive in that its bias has the fewest components and, moreover, its bias corrected version is second order asymptotically efficient. An important fact is that in an instrumental variables regression the bias of EL estimators does not, in contrast to that of GMM estimators, grow with the number of instruments.

[1]I would like to thank a co-editor and two referees whose comments helped greatly improve the presentation. I also thank Geert Dhaene, Yuichi Kitamura, Grigory Kosenok, and Richard Smith for valuable discussions, Andrey Vasnev for research assistance, and the participants of seminars at the New Economic School, the Catholic University of Leuven, the 2002 Econometric Study Group Annual Conference at the University of Bristol, the 2003 North American Summer Meeting of the Econometric Society in Evanston, IL, and the 2003 European Economic Association Annual Meeting in Stockholm, Sweden.

The latter property would make the class of (appropriately modified) GEL estimators especially attractive in numerous time series models typically estimated by GMM often using big instrument sets. The CU estimator, in particular, was initially proposed in the context of a capital asset pricing model (Hansen, Heaton, and Yaron (1996)) in an attempt to improve finite sample properties of GMM. In this paper, we consider a framework with a stationary moment function that is serially correlated of possibly infinite order. In such situations the GMM weight matrix usually has a heteroskedasticity and autocorrelation consistent (HAC) form that yields positive definiteness (Andrews (1991)), while the GEL estimator should be modified to attain asymptotic efficiency by using kernel smoothing of moment conditions (Imbens (1997), Smith (1997)). We derive second order asymptotic bias expressions for the GMM and GEL estimators under mixing conditions and a sufficiently slow bandwidth growth rate. Some of our results parallel those of Newey and Smith (2004), while others reflect time series specifics. In particular, the use of smoothed GEL removes the bias component that is associated with the correlation between the moment function and its derivative, and has a potential to grow with the degree of overidentification. On the other hand, the bias component associated with third moments can be removed only by a judicious choice of the kernel, even in the case of EL.

In addition, we analyze the important case of no serial correlation, and we draw a surprising conclusion that for the sake of reducing the finite sample bias, it is worthwhile to use the weight matrix in the HAC form in CU (but not in two step GMM!) and undertake smoothing in GEL, even though these are not necessary to do from the point of view of first order asymptotic properties (cf. Donald and Newey (2000)).

2. GMM AND GEL ESTIMATORS FOR TIME SERIES

Suppose we have the system of unconditional moment restrictions

$$E[m(w_t, \theta)] = 0, \tag{1}$$

where w_t is an observable random vector on which data from $t = 1$ to $t = T$ are available, θ is a $k \times 1$ vector of parameters to be estimated, and $m_t \equiv m(w_t, \theta)$ is an $\ell \times 1$ moment function, $\ell \geq k$. Let $\hat{m}$, $\bar{m}$, m^*, etc. refer to m evaluated at $\hat{\theta}$, $\bar{\theta}$, θ^*, etc. and let the θ subscript denote first derivatives with respect to θ. Let $\|A\|$ denote the norm $\mathrm{tr}(A'A)^{1/2}$ for a matrix A. We make the following assumptions about the model and data generation.

ASSUMPTION 1: *The sequence w_t is strictly stationary and strongly mixing with mixing coefficients α_j satisfying $\sum_{j=1}^{\infty} j^2 \alpha_j^{1-1/\nu} < \infty$ for some $\nu > 1$.*

ASSUMPTION 2: *The following regularity conditions hold*:

(a) *The moment restriction* (1) *holds for unique* $\theta \in \text{int}(\Theta)$, *where* $\Theta \subseteq \mathbb{R}^k$ *is compact*.

(b) *The function* $m(w_t, \theta^*)$ *is Borel measurable for all* $\theta^* \in \Theta$ *and is twice continuously differentiable in* θ^* *for all* $\theta^* \in \Theta$ *and for all* w_t *in its support*.

(c) *For some stationary series* d_t *with finite* $E[d_t^8]$, $\sup_{\theta^* \in \Theta} \max\{\|m_t^*\|, \|m_{\theta t}^*\|, \|\partial m_{\theta t}^*/\partial \theta_j\|, \|\partial^2 m_t^*/\partial \theta_j \, \partial \theta'\| \, \forall \, j = 1, \ldots, k\} \leq d_t$ *and* $\max\{\|m_t^* - m_t\|, \|m_{\theta t}^* - m_{\theta t}\|, \|\partial m_{\theta t}^*/\partial \theta_j - \partial m_{\theta t}/\partial \theta_j\| \, \forall \, j = 1, \ldots, k\} \leq d_t \|\theta^* - \theta\|$ *for all* $\theta^* \in \Theta$.

(d) *The matrices* $Q = E[m_{\theta t}]$ *and* $V = \sum_{s=-\infty}^{\infty} E[m_t m_{t-s}']$ *are of full rank*.

Define the matrices

$$\Sigma = (Q'V^{-1}Q)^{-1}, \quad \Xi = \Sigma Q'V^{-1}, \quad \Omega = V^{-1} - V^{-1}Q\Xi.$$

In constructing various estimators, we will be using, explicitly or implicitly, consistent estimates of the long run variance V. Let us choose a kernel with the following properties:

ASSUMPTION 3: *The kernel function* $k(x) : [-b, b] \to [-\bar{k}, \bar{k}]$ *for finite* b *and* $\bar{k}$ *is symmetric, nonzero at* 0, *continuous on* $(-b, b)$, *continuously differentiable on* $(-b, b)$ *except possibly at a finite number of points, and normalized so that* $\int_{-b}^{b} k(x)\, dx = 1$.

A variety of popular kernels satisfy Assumption 3: truncated, Bartlett, Parzen, Tukey–Hanning (see Andrews (1991)). Define the system of weights $\kappa(s) = \delta_T^{-1} k(\delta_T^{-1} s)$, where δ_T is the bandwidth parameter tending to infinity more slowly than the sample size and chosen so that $\kappa(s)$'s sum to unity over $s = -r_T, -r_T + 1, \ldots, r_T - 1, r_T$, where $r_T = \lfloor \delta_T b \rfloor$. To derive the results related to asymptotic bias we require a sufficiently slow growth rate of the bandwidth:

ASSUMPTION 4: *As* $T \to \infty$, $\delta_T \to \infty$ *and* $\delta_T = o(T^{1/3})$.

Define the smoothed moment function

$$m_t^\kappa = \sum_{s=-r_T}^{r_T} \kappa(s) m_{t-s}$$

and let other objects with κ superscripts (like $m_{\theta t}^\kappa$) be defined analogously. The convention is the following: if some time index of a summand is beyond the sample limits, the entire summand is dropped. Let us denote

$$\rho_2 = \int_{-b}^{b} k(x)^2\, dx, \quad \rho_3 = \int_{-b}^{b} k(x)^3\, dx.$$

Define the *induced kernel* (Smith (2001)) with domain $[-2b, 2b]$ to be proportional to the self-convolution of $k(x)$ and normalized so that $k^*(0) = 1$,

$$k^*(x) = \rho_2^{-1} \int_{-b}^{b} k(x+y)k(y)\,dy,$$

and let the associated system of weights be $\kappa^*(s) = \delta_T^{-1} k^*(\delta_T^{-1} s)$.

In constructing the efficient weight matrix for the *generalized method of moments* estimator, researchers most often use HAC matrices that are positive definite. For comparability with other estimators defined below, we assume that the weight matrix is constructed in the form (cf. Smith (2001, Section 2.4))

$$\rho_2^{-1} \frac{\delta_T}{T} \sum_{t=1-r_T}^{T+r_T} m_t^{\kappa} m_t^{\kappa\prime},$$

which equals, up to a remainder that has no impact on the second order bias,

$$\frac{\delta_T}{T} \sum_{t=1}^{T} \sum_{s=-2r_T}^{2r_T} \kappa^*(s) m_t m'_{t-s},$$

which in turn is a consistent and positive definite estimator of V and has a habitual form for a HAC weight matrix. The two step GMM estimator $\hat{\theta}_{\mathrm{GMM}}$ (Hansen (1982)) is

$$\arg\min_{\theta\in\Theta} \left(\sum_{t=1}^{T} m(w_t, \theta) \right)' \left(\sum_{t=1-r_T}^{T+r_T} \bar{m}_t^{\kappa} \bar{m}_t^{\kappa\prime} \right)^{-1} \left(\sum_{t=1}^{T} m(w_t, \theta) \right), \tag{2}$$

where $\bar{\theta}$ is the first step preliminary (possibly asymptotically inefficient) GMM estimator. Other variants of efficient GMM iterate one more time or to convergence. Let W denote the probability limit of weight matrix used at the preliminary step and let $\Xi_W = (Q'WQ)^{-1}Q'W$. If $\bar{\theta}$ is asymptotically efficient, then $\Xi_W = \Xi$.

The *continuously updating* estimator $\hat{\theta}_{\mathrm{CU}}$ (Hansen, Heaton, and Yaron (1996)) is

$$\arg\min_{\theta\in\Theta} \left(\sum_{t=1}^{T} m(w_t, \theta) \right)' \times \left(\sum_{t=1-r_T}^{T+r_T} m^{\kappa}(w_t, \theta) m^{\kappa}(w_t, \theta)' \right)^{-1} \left(\sum_{t=1}^{T} m(w_t, \theta) \right). \tag{3}$$

The baseline *generalized empirical likelihood* estimator $\hat{\theta}_{\mathrm{GEL}}$ together with the $\ell \times 1$ vector of additional parameters $\hat{\lambda}_{\mathrm{GEL}}$ solves the saddle point problem

$$\min_{\theta \in \Theta} \sup_{\lambda : \lambda' m_t \in \Upsilon} \sum_{t=1}^{T} h(\lambda' m(w_t, \theta)), \tag{4}$$

where the scalar function $h(\varsigma)$ and set Υ index members of the GEL class. When $h(\varsigma) = \log(1 - \varsigma)$ and $\Upsilon = (-\infty, 1)$, it is the empirical likelihood estimator; when $h(\varsigma) = 1 - \exp(\varsigma)$, it is the exponential tilting estimator; when $h(\varsigma) = -\frac{1}{2}\varsigma^2 - \varsigma$, it is the CU estimator with a non-HAC weight matrix (Newey and Smith (2004)). Let $h(\varsigma)$ satisfy the following conditions:

ASSUMPTION 5: *The function $h(\varsigma)$ is concave and three times continuously differentiable on Υ, an open interval containing zero, has bounded Lipschitz third derivative in a neighborhood of zero, and is normalized so that $h_0 = 0$, $h_1 = h_2 = -1$, where $h_r = \partial^r h(0)/\partial \varsigma^r$.*

The GEL estimator is generally inefficient when serial correlation in m_t is present. To construct an efficient estimator when there is serial correlation of unknown order, the problem is modified in the following way. As suggested by Kitamura and Stutzer (1997) and Smith (1997), the moment function is smoothed using the system of weights $\kappa(s)$. The saddle point problem for the *smoothed generalized empirical likelihood* (SGEL) estimator $\hat{\theta}_{\mathrm{SGEL}}$ and the associated $\ell \times 1$ vector of additional parameters $\hat{\lambda}_{\mathrm{SGEL}}$ is

$$\min_{\theta \in \Theta} \sup_{\lambda : \lambda' m_t^{\kappa} \in \Upsilon} \sum_{t=1-r_T}^{T+r_T} h(\lambda' m^{\kappa}(w_t, \theta)). \tag{5}$$

Note that quadratic $h(\varsigma)$ leads to the CU estimator (3).

The GMM, CU, and SGEL estimators have the same asymptotic distribution $N(0, \Sigma)$ (Hansen (1982), Smith (1997)).

3. ASYMPTOTIC BIAS OF GMM AND GEL ESTIMATORS

Nagar (1959) type asymptotic expansions have become a standard tool of analyzing finite sample behavior of first order asymptotically efficient estimators. Aside from Newey and Smith (2004), they have been recently undertaken, for example, in Rilstone, Srivastava, and Ullah (1996) for a variety of nonlinear estimators. In the time series context, early papers were interested in expansions of simple statistics in simple models; for example, Phillips (1977) derived the Edgeworth expansion for the ordinary least squares estimator and associated t statistic in a first order autoregression. Recently, Bao and Ullah (2003) have derived stochastic expansions for various estimators in nonlinear time series

models, although no variance estimators in the HAC form are used. Higher order properties of GMM test statistics that employ HAC variance matrix estimators have been explored in the bootstrap literature (e.g., Götze and Künsch (1996) and Inoue and Shintani (2003)). In particular, it follows from Lemmas A1–A4 in Inoue and Shintani (2003) that the leading asymptotically vanishing term in the Edgeworth expansion for recentered and normalized GMM estimators has the order $T^{-1/2}$ unaffected by HAC estimation, while the order of the next term depends both on the bandwidth and on the characteristic exponent of the kernel function.

Denote the jth column of the identity matrix by e_j and let

$$B_{m^3}(u,v) = \Xi E[m_t m'_{t-u} \Omega m_{t-v}],$$
$$B_{\partial m \Omega m}(u) = -\Sigma E[m'_{\theta t} \Omega m_{t-u}],$$
$$B_W(u) = \Xi \sum_{j=1}^{k} E\left[\frac{\partial m_t\, m'_{t-u}}{\partial \theta_j} \Omega V \Xi'_W e_j\right],$$
$$B_{\partial m \Xi m} = \Xi \sum_{u=-\infty}^{+\infty} E[m_{\theta t} \Xi m_{t-u}],$$
$$B_{\partial^2 m} = -\Xi \sum_{j=1}^{k} E\left[\frac{\partial m_{\theta t}}{\partial \theta_j} \frac{\Sigma}{2} e_j\right].$$

THEOREM 1: *Under Assumptions* 1–5, *the asymptotic biases of order* T^{-1} *for the GMM, CU, and SGEL estimators are*

$$B_{\mathrm{GMM}} = \sum_{u=-\infty}^{+\infty} \sum_{v=-\infty}^{+\infty} B_{m^3}(u,v) + \sum_{u=-\infty}^{+\infty} B_{\partial m \Omega m}(u) + \sum_{u=-\infty}^{+\infty} B_W(u) + B_{\partial m \Xi m} + B_{\partial^2 m},$$
$$B_{\mathrm{CU}} = \sum_{u=-\infty}^{+\infty} \sum_{v=-\infty}^{+\infty} B_{m^3}(u,v) + B_{\partial m \Xi m} + B_{\partial^2 m},$$
$$B_{\mathrm{SGEL}} = \left(1 + \frac{h_3 \rho_3}{2\rho_2^2}\right) \sum_{u=-\infty}^{+\infty} \sum_{v=-\infty}^{+\infty} B_{m^3}(u,v) + B_{\partial m \Xi m} + B_{\partial^2 m}.$$

When observations are independent, the formulas coincide with their simpler counterparts from Newey and Smith (2004), except for the presence of the m^3 component in B_{SGEL}, even in the case of EL due to smoothing when there is no need to smooth. All bias expressions involve third moments of the moment function. However, in the case of SGEL this term is scaled by a factor that depends on the choice of the kernel and may be manipulated (see below). Note

that because CU is a special case of SGEL with $h_3 = 0$, this factor for CU is unity and does not depend on the kernel.

As in Newey and Smith (2004), there are two common components $\partial m\, \Xi m$ and $\partial^2 m$ in all bias formulas which are present even under exact identification and represent the bias of the method of moments estimator based on the infeasible optimal combination of moment conditions. The W term in B_{GMM} vanishes if the first step estimator $\bar{\theta}$ arrives from the efficient GMM (then $\Omega V \Xi_W' = 0$). The GMM estimator also has the bias component $\partial m\, \Omega m$, the long run covariance between the moment function and its derivative, which is absent from bias expressions of the others. Newey and Smith (2004) establish that in cross sectional models estimated by instrumental variables, this bias component grows linearly with the number of instruments. Use of SGEL estimators removes this term critical for finite sample properties of the GMM estimator.

As noted above, the first bias term in B_{SGEL} can be removed by a judicious choice of the kernel function. Consider empirical likelihood estimation ($h_3 = -2$), in which case that term can be removed by setting $\rho_3 \rho_2^{-2} = 1$ (alternatively, the factor can be tuned to offset other bias components, which, however, does not seem realistic in practice). It turns out that among positive kernels, only the truncated kernel satisfies this condition, due to the Cauchy–Schwarz inequality $(\int k(x)^2\,dx)^2 \le \int k(x)\,dx \cdot \int k(x)^3\,dx$. It is this smoother that was originally proposed, although for the sake of simplicity, in Kitamura and Stutzer (1997). Among nonpositive kernels, there are many that achieve $\rho_3 \rho_2^{-2} = 1$, for example, $k(x) = 0.515x^4 - 1.809x^2 + 1$, $|x| \le 1$.

To better understand the sources of some bias components, we follow Donald and Newey (2000) and look at first order conditions (FOC). The FOC for SGEL (omitting limits of summation) are

$$\left(\sum_t \frac{\partial h(\hat{\lambda}'\hat{m}_t^\kappa)}{\partial \varsigma} \hat{m}_{\theta t}^\kappa e_j\right)' \left(\sum_t \varphi(\hat{\lambda}'\hat{m}_t^\kappa)\hat{m}_t^\kappa \hat{m}_t^{\kappa\prime}\right)^{-1} \sum_t \hat{m}_t^\kappa = 0$$

for all $j = 1, \ldots, k$, where $\varphi(\varsigma) \equiv -(\partial h(\varsigma)/\partial \varsigma + 1)/\varsigma$ (note that $\varphi(0) \equiv \lim_{\varsigma \to 0} \varphi(\varsigma) = 1$). Donald and Newey (2000) explain the absence of the $\partial m\, \Omega m$ bias component for the CU estimator on serially correlated data with an HAC weight matrix by the fact that the Jacobian term in the FOC is a sum of residuals from a projection of moment derivatives on moments and thus is orthogonal to the moment function. Likewise, in the case of SGEL the Jacobian can also be represented as a sum of residuals of the type

$$\begin{aligned} &-\frac{\partial h(\hat{\lambda}'\hat{m}_t^\kappa)}{\partial \varsigma} \hat{m}_{\theta t}^\kappa e_j \\ &\quad = \hat{m}_{\theta t}^\kappa e_j - \sum_\tau \varphi(\hat{\lambda}'\hat{m}_\tau^\kappa)\hat{m}_{\theta\tau}^\kappa e_j \left(\hat{m}_\tau^{\kappa\prime}\left(\sum_s \varphi(\hat{\lambda}'\hat{m}_s^\kappa)\hat{m}_s^\kappa \hat{m}_s^{\kappa\prime}\right)^{-1} \hat{m}_t^\kappa\right). \end{aligned}$$

Here the sample projection coefficients do properly estimate the population projection coefficients when there is serial correlation.

Regarding the m^3 bias component, the situation is complicated by the fact that the second moment FOC term is in the denominator. However, in the neighborhood of $\hat{\lambda} = 0$,

$$\begin{aligned}
&\left(\sum_t \varphi(\hat{\lambda}'\hat{m}_t^\kappa)\hat{m}_t^\kappa \hat{m}_t^{\kappa\prime}\right)^{-1} \\
&= \left(\sum_t \hat{m}_t^\kappa \hat{m}_t^{\kappa\prime} + \sum_t \psi(\hat{\lambda}'\hat{m}_t^\kappa)\hat{m}_t^\kappa \hat{m}_t^{\kappa\prime}\hat{\lambda}'\hat{m}_t^\kappa\right)^{-1} \\
&\approx \left(\sum_t \hat{m}_t^\kappa \hat{m}_t^{\kappa\prime}\right)^{-1}\left(I - \sum_\tau \psi(\hat{\lambda}'\hat{m}_\tau^\kappa)\hat{m}_\tau^\kappa \hat{m}_\tau^{\kappa\prime}(\hat{\lambda}'\hat{m}_\tau^\kappa)\left(\sum_t \hat{m}_t^\kappa \hat{m}_t^{\kappa\prime}\right)^{-1}\right) \\
&= \left(\sum_t \hat{m}_t^\kappa \hat{m}_t^{\kappa\prime}\right)^{-1}\left(\sum_t \zeta_t\right)\left(\sum_t \hat{m}_t^\kappa \hat{m}_t^{\kappa\prime}\right)^{-1},
\end{aligned}$$

where $\psi(\varsigma) \equiv (\varphi(\varsigma) - 1)/\varsigma$ (note that $\psi(0) \equiv \lim_{\varsigma\to 0}\psi(\varsigma) = -h_3/2$) and

$$\zeta_t \equiv \hat{m}_t^\kappa \hat{m}_t^{\kappa\prime} - \sum_\tau \psi(\hat{\lambda}'\hat{m}_\tau^\kappa)\hat{m}_\tau^\kappa \hat{m}_\tau^{\kappa\prime}\left(\hat{m}_\tau^{\kappa\prime}\left(\sum_s \varphi(\hat{\lambda}'\hat{m}_s^\kappa)\hat{m}_s^\kappa \hat{m}_s^{\kappa\prime}\right)^{-1}\hat{m}_t^\kappa\right).$$

The inverted second moment FOC term asymptotically equals a sandwich with the middle matrix having summands that may or may not be residuals from the projection of squares of moments on moments. In the IID setting without smoothing, the sample projection coefficients properly estimate the population counterparts when $\psi(0) = \varphi(0)$ which holds when $h_3 = -2$. This explains why in the IID setting EL estimation removes the m^3 bias component, while GEL with $h_3 \neq -2$ does not. When there is serial correlation, kernel smoothing provides consistent estimation of projection coefficients but also requires additional multipliers creating a new disparity, hence the presence of the m^3 bias component even for the smoothed EL.

4. SERIALLY UNCORRELATED MOMENT FUNCTION

Consider now the important case when m_t is serially uncorrelated, but not IID across time. Then the standard practice is to use a non-HAC form of weight matrices when applying GMM or CU and to use the GEL estimator (4) without smoothing.

THEOREM 2: *Suppose that $E[m_t m_{t-s}'] = 0$ for all $s \neq 0$, a non-HAC weight matrix is used for GMM and CU, and no smoothing is used for GEL. Under*

Assumptions 1, 2, *and* 5, *the asymptotic biases of order* T^{-1} *for the GMM, CU, and GEL estimators are*

$$B_{\mathrm{GMM}} = \sum_{u=-\infty}^{+\infty} B_{m^3}(u,0) + \sum_{u=-\infty}^{+\infty} B_{\partial m\,\Omega m}(u) + B_W(0) + B_{\partial m\,\Xi m} + B_{\partial^2 m},$$

$$B_{\mathrm{CU}} = \sum_{u=-\infty}^{+\infty} B_{m^3}(u,0) + \sum_{u\neq 0} B_{\partial m\,\Omega m}(u) + B_{\partial m\,\Xi m} + B_{\partial^2 m},$$

$$B_{\mathrm{GEL}} = \left(1+\frac{h_3}{2}\right) B_{m^3}(0,0) + \sum_{u\neq 0} B_{m^3}(u,0) + \sum_{u\neq 0} B_{\partial m\,\Omega m}(u) + B_{\partial m\,\Xi m} + B_{\partial^2 m}.$$

If in addition $E[m_t|w_{t-1}, w_{t-2}, \dots] = 0$, *the signs* $\sum_{u=-\infty}^{+\infty}$ *and* $\sum_{u\neq 0}$ *can be replaced, respectively, by* $\sum_{u\geq 0}$ *and* $\sum_{u>0}$.

Without smoothing, the sample projection coefficients do not properly estimate population counterparts when moments and derivatives are correlated, so the summands of the Jacobian and second moment FOC terms do not contain sample residuals any longer and hence are not orthogonal to the moment function. Most importantly, in the absence of smoothing the problematic Jacobian-related bias component can be only partially removed by using GEL. Taking this to be the primary concern, and comparing the results of Theorems 1 and 2, we can draw several important conclusions: when the moment function is serially uncorrelated, but not IID across time, smoothing the moment function in GEL tends to reduce the bias, and so does use of weight matrices in the HAC form in CU (cf. Donald and Newey (2000)), albeit not in GMM.

In addition, it can be concluded that an attempt to turn the problem with serially correlated errors into one with no serial correlation by prewhitening the moment function in order to avoid practical complications arising from smoothing will likely lead to a bigger bias compared to use of smoothing. The prewhitening will not eliminate correlatedness between the moment function and its derivatives, while the smoothing will.

New Economic School, Nakhimovsky Prospekt, 47, Room 1721, Moscow 117418, Russia; sanatoly@nes.ru; http://www.nes.ru/∼sanatoly/.

Manuscript received July, 2003; final revision received October, 2004.

APPENDIX: PROOFS

Denote

$$\zeta_T = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} m_t,$$

$$\Delta_{\partial m} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (m_{\theta t} - Q), \quad \Delta_{mm} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (m_t m_t' - V).$$

In what follows, we omit subscripts referring to estimator types when it does not cause ambiguity. Unless stated explicitly, all summations with indices s, u, v are two-sided infinite. Less essential proofs are omitted to save space.

LEMMA 1: *Under Assumptions* 1 *and* 2, *the following infinite summations converge for all* i, j, k: $\sum_u |E[m_{i,t-u}\, \partial m_{j,t}/\partial \theta_k]|$, $\sum_u u^2 |E[m_{i,t} m'_{j,t-u}]|$, *and* $\sum_u \sum_v (|u| + |v|) |E[m_{i,t} m_{j,t-u} m'_{k,t-v}]|$.

LEMMA 2: *Under Assumptions* 1 *and* 2, *the following are* $o(1)$: $E[\Delta_{\partial m} \Xi \zeta_T] - \sum_s E[m_{\theta t} \Xi m_{t-s}]$, $E[\Delta'_{\partial m} \Omega \zeta_T] - \sum_s E[m'_{\theta t} \Omega m_{t-s}]$, $E[\Delta_{mm} \Omega \zeta_T] - \sum_s E[m_t m'_t \times \Omega m_{t-s}]$, $E[\Xi \zeta_T (\Xi \zeta_T)'] - \Sigma$, $E[\Omega \zeta_T (\Xi \zeta_T)']$, $E[\Omega \zeta_T (\Xi_W \zeta_T)'] - \Omega V \Xi'_W$, *and* $E[\Omega \zeta_T (\Omega \zeta_T)'] - \Omega$.

PROOF: All results are obtained in the same way, using $\Xi V \Xi' = \Sigma$, $\Omega V \Xi' = 0$, and $\Omega V \Omega' = \Omega$, and convergence in Lemma 1. For example, $E[\Delta_{\partial m} \Xi \zeta_T] = T^{-1} \sum_{t=1}^{T} \sum_{\tau=1}^{T} E[m_{\theta t} \Xi m_\tau] = \sum_{s=-(T-1)}^{T-1} (1 - |s|/T) \times E[m_{\theta t} \Xi m_{t-s}] \to \sum_s E[m_{\theta t} \Xi m_{t-s}]$. *Q.E.D.*

LEMMA 3: *Under Assumptions* 1–4, *the following are true*:

(a) $$\frac{1}{T} \sum_{t=1-r_T}^{T+r_T} \frac{\partial m_{\theta t}^{\kappa}}{\partial \theta_j} = E\left[\frac{\partial m_{\theta t}}{\partial \theta_j}\right] + O_p\left(\frac{1}{\sqrt{T}}\right);$$

(b) $$\frac{1}{T} \sum_{t=1-r_T}^{T+r_T} m_t^{\kappa} m_t^{\kappa'} - \frac{\rho_2}{\delta_T} V, \quad \frac{1}{T} \sum_{t=1-r_T}^{T+r_T} \frac{\partial m_t^{\kappa}\, m_{it}^{\kappa}}{\partial \theta'} - \frac{\rho_2}{\delta_T} \sum_{u=-\infty}^{+\infty} E\left[\frac{\partial m_t\, m_{i,t-u}}{\partial \theta'}\right],$$

and

$$\frac{1}{T} \sum_{t=1-r_T}^{T+r_T} m_{\theta t}^{\kappa'} (e_i' m_t^{\kappa} + e_i m_t^{\kappa'}) - \frac{\rho_2}{\delta_T} \sum_{u=-\infty}^{+\infty} E[m'_{\theta t} (e_i' m_{t-u} + e_i m'_{t-u})]$$

are

$$O_p\left(\frac{\delta_T}{T}+\frac{1}{\delta_T^2}+\frac{1}{\sqrt{\delta_T T}}\right);$$

(c) $$E\left[\rho_2^{-1}\frac{\delta_T}{\sqrt{T}}\sum_{t=1-r_T}^{T+r_T} m_t^{\kappa} m_t^{\kappa'}\Omega\zeta_T\right]-\sum_{u=-\infty}^{+\infty}\sum_{v=-\infty}^{+\infty}E[m_t m_{t-u}'\Omega m_{t-v}]$$
$$=O\left(\frac{\delta_T^3}{T}+\frac{1}{\delta_T}\right);$$

(d) $$\frac{1}{T}\sum_{t=1-r_T}^{T+r_T} m_{it}^{\kappa} m_t^{\kappa} m_t^{\kappa'}-\frac{\rho_3}{\delta_T^2}\sum_{u=-\infty}^{+\infty}\sum_{v=-\infty}^{+\infty}E[m_{it}m_{t-u}m_{t-v}']$$
$$=O_p\left(\frac{\delta_T}{T}+\frac{1}{\delta_T^3}+\frac{1}{\delta_T\sqrt{T}}\right).$$

PROOF: In (a),

$$\sum_{t=1-r_T}^{T+r_T}\frac{\partial m_{\theta t}^{\kappa}}{\partial\theta_j}=\sum_{t=1}^{T}\frac{\partial m_{\theta t}}{\partial\theta_j}.$$

By the law of large numbers and the central limit theorem for α-mixing sequences,

$$T^{-1}\sum_{t=1}^{T}\frac{\partial m_{\theta t}}{\partial\theta_j}=E\left[\frac{\partial m_{\theta t}}{\partial\theta_j}\right]+O_p\left(\frac{1}{\sqrt{T}}\right).$$

In (b),

$$\sum_{t=1-r_T}^{T+r_T} m_t^{\kappa} m_t^{\kappa'}=\sum_{t=1-r_T}^{T+r_T}\sum_{s=-r_T}^{r_T}\kappa(s)m_{t-s}\sum_{v=-r_T}^{r_T}\kappa(v)m_{t-v}'$$
$$=(\rho_2+O(\delta_T^{-1}))\sum_{t=1-r_T}^{T+r_T}\sum_{u=-2r_T}^{2r_T}\kappa^*(u)m_t m_{t-u}'+O_p(\delta_T).$$

Here, the term $O(\delta_T^{-1})$ is the mismatch between discrete and continuous versions of the kernel, and $O_p(\delta_T)$ is due to boundary effects. From Lemma 1, $\sum_u u^2\|E[m_t m_{t-u}']\|<\infty$. The characteristic exponent (Parzen (1957)) of

$k^*(x)$ is at least 1 (because as $x \to 0$, $|1 - k^*(x)|/|x| = \rho_2^{-1} |\int (k(y) - k(y + x))k(y)\,dy|/|x| = \rho_2^{-1} |\int (x + o(|x|))k'(y)k(y)\,dy|/|x| < \infty$); hence

$$\frac{\delta_T}{T} \sum_{t=1}^{T} \sum_{u=-2r_T}^{2r_T} \kappa^*(u) m_t m'_{t-u} = V + O_p\left(\frac{1}{\delta_T} + \sqrt{\frac{\delta_T}{T}}\right)$$

(Parzen (1957, Theorem 5)), so the first statement holds. Other statements are handled similarly. In (c),

$$\begin{aligned}
&E\left[\sum_{t=1-r_T}^{T+r_T} m_t^{\kappa} m_t^{\kappa'} \sum_{t=1}^{T} m_{it}\right] \\
&\quad = (\rho_2 + O(\delta_T^{-1}))E\left[\sum_{t=1-r_T}^{T+r_T} \sum_{u=-2r_T}^{2r_T} \kappa^*(u) m_t m'_{t-u} \sum_{t=1}^{T} m_{it}\right] + O(\delta_T^2) \\
&\quad = (\rho_2 + O(\delta_T^{-1})) \sum_{u=-2r_T}^{2r_T} \kappa^*(u) \sum_{v=-(T-1)}^{T-1} (T - |v|) E[m_t m'_{t-u} m_{i,t-v}] \\
&\qquad + O(\delta_T^2).
\end{aligned}$$

From Lemma 1, $\sum_u \sum_v \|E[m_t m'_{t-u} m_{i,t-v}]\| < \infty$. Then, because

$$\begin{aligned}
&\sum_{u=-2r_T}^{2r_T} k^*\left(\frac{u}{\delta_T}\right) \sum_{v=-(T-1)}^{T-1} \left(1 - \frac{|v|}{T}\right) E[m_t m'_{t-u} m_{i,t-v}] \\
&\quad \to \sum_{u=-\infty}^{+\infty} \sum_{v=-\infty}^{+\infty} E[m_t m'_{t-u} m_{i,t-v}]
\end{aligned}$$

as $T \to \infty$ and $\delta_T \to \infty$, the result follows. In (d), for any combination of indices $i, j, l = 1, \ldots, \ell$, we have

$$\begin{aligned}
&\sum_{t=1-r_T}^{T+r_T} m_{it}^{\kappa} m_{jt}^{\kappa} m_{lt}^{\kappa} \\
&\quad = \sum_{t=1-r_T}^{T+r_T} \sum_{s=-r_T}^{r_T} \kappa(s) m_{i,t-s} \sum_{u=-r_T}^{r_T} \kappa(u) m_{j,t-u} \sum_{v=-r_T}^{r_T} \kappa(v) m_{l,t-v} \\
&\quad = (\rho_3 + O(\delta_T^{-1})) \sum_{t=1-r_T}^{T+r_T} \sum_{u=-2r_T}^{2r_T} \sum_{v=-2r_T}^{2r_T} \kappa^{**}(u, v) m_{it} m_{j,t-u} m_{l,t-v} \\
&\qquad + O_p(\delta_T),
\end{aligned}$$

where $\kappa^{**}(u,v) = \delta_T^{-2} k^{**}(\delta_T^{-1}u, \delta_T^{-1}v)$ and $k^{**}(y,z) = \rho_3^{-1}\int_{-\infty}^{+\infty} k(x)k(x+y) \times k(x+z)\,dx$ is the *bispectral estimating kernel* (Rosenblatt and Van Ness (1965)), symmetric, continuous in both arguments at $(0,0)$, and normalized so that $k^{**}(0,0) = 1$. Because $\sum_u \sum_v (|u| + |v|)\|E[m_{it}m_{t-u}m'_{t-v}]\| < \infty$ by Lemma 1 and since the order of $k^{**}(y,z)$ is at least 1 (because as $y \to 0$ and $z \to 0$, $|1 - k^{**}(y,z)|/|y+z| = \rho_3^{-1}|\int (k(x)^2 - k(x+y)k(x+z))k(x)\,dx|/|y+z| = \rho_3^{-1}|\int ((y+z) + o(|y+z|))k'(x)k(x)^2\,dx|/|y+z| < \infty$),

$$\frac{\delta_T^2}{T}\sum_{t=1-r_T}^{T+r_T}\sum_{u=-2r_T}^{2r_T}\sum_{v=-2r_T}^{2r_T}\kappa^{**}(u,v)m_{it}m_{t-u}m'_{t-v}$$
$$= \sum_{u=-\infty}^{+\infty}\sum_{v=-\infty}^{+\infty} E[m_{it}m_{t-u}m'_{t-v}] + O_p\left(\frac{1}{\delta_T} + \frac{\delta_T}{\sqrt{T}}\right)$$

(Rosenblatt and Van Ness (1965, Theorems 4 and 5)), and the result follows. *Q.E.D.*

LEMMA 4: *Under Assumptions* 1–5, *for* $(\theta^{*\prime}\lambda^{*\prime})'$ *lying between* $(\hat{\theta}'_{\mathrm{SGEL}}\hat{\lambda}'_{\mathrm{SGEL}})'$ *and* $(\theta'\ 0')'$,

$$T^{-1}\sum_{t=1-r_T}^{T+r_T}\frac{\partial^2(h_1(\lambda^{*\prime}m_t^{*\kappa})m_{\theta t}^{*\kappa\prime}\lambda^*)}{\partial\theta'\partial\theta_j} = O_p\left(\frac{\delta_T}{\sqrt{T}}\right);$$

$$T^{-1}\sum_{t=1-r_T}^{T+r_T}\frac{\partial^2(h_1(\lambda^{*\prime}\hat{m}_t^{*\kappa})\hat{m}_t^{*\kappa})}{\partial\theta'\partial\theta_j} = -E\left[\frac{\partial m_{\theta t}}{\partial\theta_j}\right] + O_p\left(\frac{\delta_T}{T} + \frac{1}{\sqrt{T}}\right);$$

$$T^{-1}\sum_{t=1-r_T}^{T+r_T}\frac{\partial^2(h_1(\lambda^{*\prime}m_t^{*\kappa})m_{\theta t}^{*\kappa\prime}\lambda^*)}{\partial\lambda'\partial\theta_j} = -E\left[\frac{\partial m'_{\theta t}}{\partial\theta_j}\right] + O_p\left(\frac{\delta_T}{T} + \frac{1}{\sqrt{T}}\right);$$

$$T^{-1}\sum_{t=1-r_T}^{T+r_T}\frac{\partial^2(h_1(\lambda^{*\prime}m_t^{*\kappa})m_t^{*\kappa})}{\partial\theta'\partial\lambda_i}$$
$$= -\rho_2\delta_T^{-1}\sum_u E\left[\frac{\partial m_t m_{i,t-u}}{\partial\theta'}\right] + O_p\left(\frac{\delta_T}{T} + \frac{1}{\delta_T^2} + \frac{1}{\sqrt{\delta_T T}}\right);$$

$$T^{-1}\sum_{t=1-r_T}^{T+r_T}\frac{\partial^2(h_1(\lambda^{*\prime}m_t^{*\kappa})m_{\theta t}^{*\kappa\prime}\lambda^*)}{\partial\lambda'\partial\lambda_i}$$
$$= -\rho_2\delta_T^{-1}\sum_u E[m'_{\theta t}(e_i'm_{t-u} + e_i m'_{t-u})] + O_p\left(\frac{\delta_T}{T} + \frac{1}{\delta_T^2} + \frac{1}{\sqrt{\delta_T T}}\right);$$

$$T^{-1}\sum_{t=1-r_T}^{T+r_T}\frac{\partial^2(h_1(\lambda^{*\prime}m_t^{*\kappa})m_t^{*\kappa})}{\partial\lambda'\partial\lambda_i}$$
$$=h_3\rho_3\delta_T^{-2}\sum_u\sum_v E[m_{it}m_{t-u}m'_{t-v}]+O_p\left(\frac{\delta_T}{T}+\frac{1}{\delta_T^3}+\frac{1}{\delta_T\sqrt{T}}\right).$$

PROOF: Note that $\|f_t^{*\kappa}\|=\|\sum_{s=-r_T}^{r_T}\kappa(s)f_{t-s}^*\|\le\sum_{s=-r_T}^{r_T}|\kappa(s)|\|f_{t-s}^*\|\le d_t^{|\kappa|}$, and $\|f_t^{*\kappa}-f_t^{\kappa}\|=\|\sum_{s=-r_T}^{r_T}\kappa(s)(f_{t-s}^*-f_{t-s})\|\le\sum_{s=-r_T}^{r_T}|\kappa(s)|\|f_{t-s}^*-f_{t-s}\|\le d_t^{|\kappa|}\|\theta^*-\theta\|$ for f_t equal m_t, $m_{\theta t}$, $\partial m_{\theta t}/\partial\theta_j$, and $\partial^2 m_t/\partial\theta_j\partial\theta'\,\forall\, j=1,\dots,k$, where $d_t^{|\kappa|}\equiv\sum_{s=-r_T}^{r_T}|\kappa(s)|d_{t-s}$.

A norm of any quantity in the lemma can be bounded by a sum of norms of three types. The first type includes terms $\|T^{-1}\sum_{t=1-r_T}^{T+r_T}f_t^{\kappa}-\rho_{\xi_m}\delta_T^{1-\xi_m}E[f_t]\|$ for f_t equal $\partial m_{\theta t}/\partial\theta_j$, $m_{\theta t}m_{it}+m_tm_{\theta it}$, $m'_{\theta t}m_{it}+m'_{\theta t}e_im'_t$, or $m_{it}m_tm'_t$, and κ superscripts all ξ_m inclusions ($\xi_m\in\{1,2,3\}$) of m (we define $\rho_1=1$). The orders of magnitude for these terms are given in Lemma 3(a), (b), (d). The second type includes terms $\|T^{-1}\sum_{t=1-r_T}^{T+r_T}(f_t^{*\kappa}\,\partial^\mu h(\lambda^{*\prime}m_t^{*\kappa})/\partial\varsigma^\mu-f_t^{\kappa}h_\mu)\|$ for f_t^{κ} equal $\partial m_{\theta t}^{\kappa}/\partial\theta_j$, $m_{\theta t}^{\kappa}m_{it}^{\kappa}$, $m_t^{\kappa}m_{\theta it}^{\kappa}$, $m_{\theta t}^{\kappa\prime}e_im_t^{\kappa\prime}$, or $m_{it}^{\kappa}m_t^{\kappa}m_t^{\kappa\prime}$ with ξ_m inclusions ($\xi_m\in\{1,2,3\}$) of m, and $\mu\in\{1,2,3\}$. Such terms do not exceed, by Assumption 2, the triangular and Cauchy–Schwarz inequalities, and the Lipschitz property of $\partial^\mu h(\varsigma)/\partial\varsigma^\mu$,

$$h_\mu\frac{1}{T}\sum_{t=1-r_T}^{T+r_T}\|f_t^{*\kappa}-f_t^{\kappa}\|+\frac{1}{T}\sum_{t=1-r_T}^{T+r_T}\left\|\left(\frac{\partial^\mu h(\lambda^{*\prime}m_t^{*\kappa})}{\partial\varsigma^\mu}-h_\mu\right)f_t^{*\kappa}\right\|$$
$$\le h_\mu\xi_m\frac{1}{T}\sum_{t=1-r_T}^{T+r_T}\left(d_t^{|\kappa|}\right)^{\xi_m}\|\theta^*-\theta\|+C\frac{1}{T}\sum_{t=1-r_T}^{T+r_T}\left(d_t^{|\kappa|}\right)^{\xi_m+1}\|\lambda^*\|$$
$$=O_p\left(\frac{\delta_T^{1-\xi_m}}{\sqrt{T}}\right),$$

because the first term is $O_p(\delta_T^{1-\xi_m}\cdot 1/\sqrt{T})$ and the last term is $O_p(\delta_T^{-\xi_m}\cdot\delta_T/\sqrt{T})$ by arguments similar to the proof of Lemma 3. Finally, the third type includes terms $\|T^{-1}\sum_{t=1-r_T}^{T+r_T}f_t^{*\kappa}\,\partial^\mu h(\lambda^{*\prime}\hat{m}_t^{*\kappa})/\partial\varsigma^\mu\|$ for f_t^* equal to a product of ξ_m copies ($\xi_m\in\{1,2,3\}$) of m_t^*, $m_{\theta t}^*$, $\partial m_{\theta t}^*/\partial\theta_j$, or $\sum_{i=1}^{\ell}\partial^2 m_{\theta it}^{*\prime}/\partial\theta'\partial\theta_j$, and λ^* in $\xi_\lambda\le\xi_m$ copies. Such terms do not exceed, by Assumption 2, the triangular and Cauchy–Schwarz inequalities, the continuity and boundedness of $\partial^\mu h(\varsigma)/\partial\varsigma^\mu$,

$$C\frac{1}{T}\sum_{t=1-r_T}^{T+r_T}\|f_t^{*\kappa}\|\|\lambda^*\|^{\xi_\lambda}=O_p\left(\delta_T^{1-\xi_m}(\delta_T/\sqrt{T})^{\xi_\lambda}\right).$$

When adding up, we take into account that some remainders' orders swallow orders of others. *Q.E.D.*

PROOF OF THEOREM 1: The first order asymptotics for the first step estimator $\bar{\theta}$ is $\sqrt{T}(\bar{\theta}-\theta) = -\Xi_W \zeta_T + O_p(1/\sqrt{T})$. The second step GMM estimator $\hat{\theta}$ has the FOC

$$0 = \left(\frac{1}{T}\sum_{t=1}^{T}\hat{m}_{\theta t}\right)'\left(\rho_2^{-1}\frac{\delta_T}{T}\sum_{t=1}^{T}\bar{m}_t^{\kappa}\bar{m}_t^{\kappa'}\right)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{m}_t.$$

The FOC has the expansion

$$\begin{aligned}
0 = &\left(\frac{1}{T}\sum_{t=1}^{T}m_{\theta t} + \frac{1}{\sqrt{T}}\sum_{j=1}^{k}\frac{1}{T}\sum_{t=1}^{T}\frac{\partial m_{\theta t}^{a}}{\partial\theta_j}\sqrt{T}(\hat{\theta}_j-\theta_j)\right)' \\
&\times\left(\rho_2^{-1}\frac{\delta_T}{T}\sum_{t=1-r_T}^{T+r_T}m_t^{\kappa}m_t^{\kappa'}\right. \\
&\quad\left.+\frac{1}{\sqrt{T}}\sum_{j=1}^{k}\rho_2^{-1}\frac{\delta_T}{T}\sum_{t=1-r_T}^{T+r_T}\frac{\partial m_t^{\kappa\dagger}m_t^{\dagger\kappa'}}{\partial\theta_j}\sqrt{T}(\bar{\theta}_j-\theta_j)\right)^{-1} \\
&\times\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}m_t + \frac{1}{T}\sum_{t=1}^{T}m_{\theta t}\sqrt{T}(\hat{\theta}-\theta)\right. \\
&\quad\left.+\frac{1}{2\sqrt{T}}\sum_{j=1}^{k}\frac{1}{T}\sum_{t=1}^{T}\frac{\partial m_{\theta t}^{b}}{\partial\theta_j}\sqrt{T}(\hat{\theta}_j-\theta_j)\sqrt{T}(\hat{\theta}-\theta)\right),
\end{aligned}$$

where θ^a and θ^b lie componentwise between $\hat{\theta}$ and θ, and $\theta^{\dagger}$ lies between $\bar{\theta}$ and θ. Because the first order asymptotics for $\hat{\theta}$ is $\sqrt{T}(\hat{\theta}-\theta) = -\Xi\zeta_T + O_p(1/\sqrt{T})$, the expansion can be simplified to

$$\begin{aligned}
0 = &\left(Q + \frac{1}{\sqrt{T}}\left(\Delta_{\partial m} - \sum_{j=1}^{k}E\left[\frac{\partial m_{\theta t}}{\partial\theta_j}\right]e_j'\Xi\zeta_T\right) + O_p\left(\frac{1}{T}\right)\right)'V^{-1} \\
&\times\left(I - \left(\rho_2^{-1}\frac{\delta_T}{T}\sum_{t=1-r_T}^{T+r_T}m_t^{\kappa}m_t^{\kappa'} - V\right.\right. \\
&\qquad\left.\left.-\frac{1}{\sqrt{T}}\sum_{j=1}^{k}\sum_{u=-\infty}^{+\infty}E\left[\frac{\partial m_t m_{t-u}'}{\partial\theta_j}\right]e_j'\Xi_W\zeta_T\right)V^{-1}\right.
\end{aligned}$$

$$
\begin{aligned}
&\quad + \frac{1}{\sqrt{T}} O_p\left(\frac{\delta_T^2}{T} + \frac{1}{\delta_T} + \sqrt{\frac{\delta_T}{T}} \right) \Bigg) \\
&\times \Bigg(\zeta_T + Q\sqrt{T}(\hat\theta - \theta) \\
&\quad + \frac{1}{\sqrt{T}} \left(-\Delta_{\partial m} \Xi \zeta_T + \frac{1}{2} \sum_{j=1}^{k} E\left[\frac{\partial m_{\theta t}}{\partial \theta_j} \right] \Xi \zeta_T \Xi \zeta_T e_j \right) \\
&\quad + O_p\left(\frac{1}{T} \right) \Bigg),
\end{aligned}
$$

where the order of the remainder in the middle follows by Lemma 3(b). Simplifying,

$$
\begin{aligned}
&o_p\left(\frac{1}{\sqrt{T}} \right) \\
&\quad = Q' V^{-1} \big(\zeta_T + Q\sqrt{T}(\hat\theta - \theta) \big) - \frac{1}{\sqrt{T}} Q' V^{-1} \Delta_{\partial m} \Xi \zeta_T + \frac{1}{\sqrt{T}} \Delta'_{\partial m} \Omega \zeta_T \\
&\qquad - \frac{1}{\sqrt{T}} Q' V^{-1} \rho_2^{-1} \frac{\delta_T}{\sqrt{T}} \sum_{t=1-r_T}^{T+r_T} m_t^\kappa m_t^{\kappa\prime} \Omega \zeta_T + Q' \Omega \zeta_T \\
&\qquad + \frac{1}{2\sqrt{T}} Q' V^{-1} \sum_{j=1}^{k} E\left[\frac{\partial m_{\theta t}}{\partial \theta_j} \right] \Xi \zeta_T \Xi \zeta_T e_j \\
&\qquad - \frac{1}{\sqrt{T}} \sum_{j=1}^{k} E\left[\frac{\partial m_{\theta t}}{\partial \theta_j} \right] \Omega \zeta_T e_j' \Xi \zeta_T \\
&\qquad - \frac{1}{\sqrt{T}} Q' V^{-1} \sum_{j=1}^{k} \sum_{u=-\infty}^{+\infty} E\left[\frac{\partial m_t m_{t-u}'}{\partial \theta_j} \right] \Omega \zeta_T e_j' \Xi_W \zeta_T .
\end{aligned}
$$

Premultiplying by $-\Sigma$, expressing out $\sqrt{T}(\hat\theta - \theta)$, and taking expectations getting rid of terms of higher order, we obtain the first result using Lemmas 2 and 3.

The FOC for the SGEL estimator are

$$
0 = \sum_{t=1-r_T}^{T+r_T} h_1(\lambda' \hat m_t^\kappa) \hat m_t^\kappa, \quad 0 = \sum_{t=1-r_T}^{T+r_T} h_1(\lambda' \hat m_t^\kappa) \hat m_{\theta t}^{\kappa\prime} \hat\lambda .
$$

Using Lemma 3, we have the expansion of the FOC to $O_p(1/\sqrt{T})$:

$$O_p(1/\sqrt{T}) = \zeta_T + Q\sqrt{T}(\hat{\theta}-\theta) + V\rho_2\delta_T^{-1}\sqrt{T}\hat{\lambda},$$
$$O_p(1/\sqrt{T}) = Q'\rho_2\delta_T^{-1}\sqrt{T}\hat{\lambda}.$$

Premultiplying the first equation by $Q'V^{-1}$, adding the second, and expressing out $\sqrt{T}(\hat{\theta}-\theta)$ and $\rho_2\delta_T^{-1}\sqrt{T}\hat{\lambda}$, we get that $\sqrt{T}(\hat{\theta}-\theta) = -\Xi\zeta_T + O_p(1/\sqrt{T})$ and $\rho_2\delta_T^{-1}\sqrt{T}\hat{\lambda} = -\Omega\zeta_T + O_p(1/\sqrt{T})$. Moreover, the leading term in an expansion of the product of these has expectation of order $o(1)$ according to Lemma 2. Taking the second order expansion,

$$\begin{aligned}
0 = &-\frac{1}{\sqrt{T}}\sum_{t=1-r_T}^{T+r_T} m_t^\kappa - \left(Q + \frac{1}{\sqrt{T}}\frac{1}{\sqrt{T}}\sum_{t=1-r_T}^{T+r_T}(m_{\theta t}^\kappa - Q)\right)\sqrt{T}(\hat{\theta}-\theta)\\
&- \left(V + \left(\rho_2^{-1}\frac{\delta_T}{T}\sum_{t=1-r_T}^{T+r_T} m_t^\kappa m_t^{\kappa\prime} - V\right)\right)\rho_2\sqrt{T}\delta_T^{-1}\hat{\lambda}\\
&+ \frac{1}{2\sqrt{T}}\sum_{j=1}^{k}\left[\frac{1}{T}\sum_{t=1-r_T}^{T+r_T}\frac{\partial^2 h_1(\lambda^{*\prime}\hat{m}_t^{*\kappa})\hat{m}_t^{*\kappa}}{\partial\theta'\,\partial\theta_j}\right]\sqrt{T}(\hat{\theta}-\theta)\sqrt{T}(\hat{\theta}_j-\theta_j)\\
&+ \frac{1}{\sqrt{T}}\sum_{i=1}^{\ell}\left[\frac{1}{T}\sum_{t=1-r_T}^{T+r_T}\frac{\partial^2 h_1(\lambda^{*\prime}\hat{m}_t^{*\kappa})\hat{m}_t^{*\kappa}}{\partial\theta'\,\partial\lambda_i}\right]\sqrt{T}(\hat{\theta}-\theta)\sqrt{T}\hat{\lambda}_i\\
&+ \frac{1}{2\sqrt{T}}\sum_{i=1}^{\ell}\left[\frac{1}{T}\sum_{t=1-r_T}^{T+r_T}\frac{\partial^2 h_1(\lambda^{*\prime}\hat{m}_t^{*\kappa})\hat{m}_t^{*\kappa}}{\partial\lambda'\,\partial\lambda_i}\right]\sqrt{T}\hat{\lambda}\sqrt{T}\hat{\lambda}_i,\\
0 = &-\left(Q + \frac{1}{\sqrt{T}}\frac{1}{\sqrt{T}}\sum_{t=1-r_T}^{T+r_T}(m_{\theta t}^\kappa - Q)\right)'\sqrt{T}\hat{\lambda}\\
&+ \frac{1}{2\sqrt{T}}\sum_{j=1}^{k}\left[\frac{1}{T}\sum_{t=1-r_T}^{T+r_T}\frac{\partial^2 h_1(\lambda^{*\prime}m_t^{*\kappa})m_\theta^{*\kappa\prime}\lambda^*}{\partial\theta'\,\partial\theta_j}\right]\\
&\qquad\times\sqrt{T}(\hat{\theta}-\theta)\sqrt{T}(\hat{\theta}_j-\theta_j)\\
&+ \frac{1}{\sqrt{T}}\sum_{j=1}^{k}\left[\frac{1}{T}\sum_{t=1-r_T}^{T+r_T}\frac{\partial^2 h_1(\lambda^{*\prime}m_t^{*\kappa})m_\theta^{*\kappa\prime}\lambda^*}{\partial\lambda'\,\partial\theta_j}\right]\sqrt{T}\hat{\lambda}\sqrt{T}(\hat{\theta}_j-\theta_j)\\
&+ \frac{1}{2\sqrt{T}}\sum_{i=1}^{\ell}\left[\frac{1}{T}\sum_{t=1-r_T}^{T+r_T}\frac{\partial^2 h_1(\lambda^{*\prime}m_t^{*\kappa})m_\theta^{*\kappa\prime}\lambda^*}{\partial\lambda'\,\partial\lambda_i}\right]\sqrt{T}\hat{\lambda}\sqrt{T}\hat{\lambda}_i,
\end{aligned}$$

where $(\theta^{*\prime}\lambda^{*\prime})'$ lies between $(\hat\theta'\hat\lambda')'$ and $(\theta'\ 0')'$ componentwise. By the first order asymptotics and Lemma 4, after simplifications we get

$$
\begin{aligned}
0 = {} & -\zeta_T - V\Omega\zeta_T - Q\sqrt{T}(\hat\theta-\theta) - V\rho_2\delta_T^{-1}\sqrt{T}\hat\lambda + \frac{1}{\sqrt{T}}\Delta_{\partial m}\Xi\zeta_T \\
& + \frac{1}{\sqrt{T}}\rho_2^{-1}\frac{\delta_T}{\sqrt{T}}\sum_{t=1-r_T}^{T+r_T} m_t^{\kappa} m_t^{\kappa\prime}\Omega\zeta_T - \frac{1}{2\sqrt{T}}\sum_{j=1}^{k} E\left[\frac{\partial m_{\theta t}}{\partial\theta_j}\right]\Xi\zeta_T\zeta_T'\Xi' e_j \\
& - \frac{1}{2\sqrt{T}}\sum_{i=1}^{\ell}\sum_{u=-\infty}^{+\infty} E\left[\frac{\partial m_t\, m_{i,t-u}}{\partial\theta'}\right]\sqrt{T}(\hat\theta-\theta)\rho_2\delta_T^{-1}\sqrt{T}\hat\lambda_i \\
& + \frac{1}{\sqrt{T}}\frac{h_3}{2}\frac{\rho_3}{\rho_2^2}\sum_{i=1}^{\ell}\sum_{u=-\infty}^{+\infty}\sum_{v=-\infty}^{+\infty} E[m_{it}m_{t-u}m_{t-v}']\Omega\zeta_T\zeta_T'\Omega e_i \\
& + \frac{1}{\sqrt{T}}O_p\left(\frac{1}{\delta_T}+\frac{\delta_T^3}{T}\right), \\
0 = {} & -Q'\rho_2\delta_T^{-1}\sqrt{T}\hat\lambda + \frac{1}{\sqrt{T}}\Delta_{\partial m}'\Omega\zeta_T \\
& - \frac{1}{\sqrt{T}}\sum_{j=1}^{k} E\left[\frac{\partial m_{\theta t}'}{\partial\theta_j}\right]\sqrt{T}\hat\lambda\sqrt{T}(\hat\theta_j-\theta_j) \\
& - \frac{1}{2\sqrt{T}}\sum_{i=1}^{\ell}\sum_{u=-\infty}^{+\infty} E[m_{\theta t}'(e_i'm_{t-u}+e_im_{t-u}')]\Omega\zeta_T\zeta_T'\Omega e_i \\
& + \frac{1}{\sqrt{T}}O_p\left(\frac{1}{\delta_T}+\frac{\delta_T^2}{T}\right).
\end{aligned}
$$

Note that the remainders in both equations are $o_p(1/\sqrt{T})$ when Assumption 4 holds. Premultiplying the first equation by $Q'V^{-1}$, adding the second equation, expressing out $\sqrt{T}(\hat\theta-\theta)$, and taking expectations getting rid of terms of higher order, we obtain the last result using Lemmas 2 and 3, and the fact that the bias component

$$
\frac{\Sigma}{2}\sum_{i=1}^{\ell}\sum_{u=-\infty}^{+\infty} E[m_{\theta t}'m_{i,t-u}+m_{\theta t}'e_im_{t-u}']E[\Omega\zeta_T\zeta_T'\Omega]e_i
$$

offsets $-\Sigma E[\Delta_{\partial m}'\Omega\zeta_T]$ up to $o(1)$. The bias expression for CU is a special case when $h_3=0$. *Q.E.D.*

PROOF OF THEOREM 2: From the first part of expansions in Lemma A4 of Newey and Smith (2004), the second order expansion for GMM and GEL are, respectively,

$$\begin{aligned}
&\sqrt{T}(\hat{\theta}_{\mathrm{GMM}} - \theta) \\
&\quad = -\Xi\zeta_T + \frac{1}{\sqrt{T}}\Xi\Delta_{mm}\Omega\zeta_T + \frac{1}{\sqrt{T}}\Xi\Delta_{\partial m}\Xi\zeta_T - \frac{1}{\sqrt{T}}\Sigma\Delta'_{\partial m}\Omega\zeta_T \\
&\qquad - \frac{1}{\sqrt{T}}\frac{\Xi}{2}\sum_{j=1}^{k} E\left[\frac{\partial m_{\theta t}}{\partial\theta_j}\right]\Xi\zeta_T\zeta'_T\Xi' e_j \\
&\qquad + \frac{1}{\sqrt{T}}\Sigma\sum_{j=1}^{k} E\left[\frac{\partial m_{\theta t}}{\partial\theta_j}\right]\Omega\zeta_T\zeta'_T\Xi' e_j \\
&\qquad + \frac{1}{\sqrt{T}}\Xi\sum_{j=1}^{k} E\left[\frac{\partial m_t m'_t}{\partial\theta_j}\right]\Omega\zeta_T\zeta'_T\Xi'_W e_j + O_p\left(\frac{1}{T}\right), \\
&\sqrt{T}(\hat{\theta}_{\mathrm{GEL}} - \theta) \\
&\quad = -\Xi\zeta_T + \frac{1}{\sqrt{T}}\Xi\Delta_{mm}\Omega\zeta_T + \frac{1}{\sqrt{T}}\Xi\Delta_{\partial m}\Xi\zeta_T - \frac{1}{\sqrt{T}}\Sigma\Delta'_{\partial m}\Omega\zeta_T \\
&\qquad - \frac{1}{\sqrt{T}}\frac{\Xi}{2}\sum_{j=1}^{k} E\left[\frac{\partial m_{\theta t}}{\partial\theta_j}\right]\Xi\zeta_T\zeta'_T\Xi' e_j \\
&\qquad + \frac{1}{\sqrt{T}}\frac{h_3\Xi}{2}\sum_{i=1}^{\ell} E[m_{it}m_t m'_t]\Omega\zeta_T\zeta'_T\Omega e_i \\
&\qquad + \frac{1}{\sqrt{T}}\frac{\Sigma}{2}\sum_{i=1}^{\ell} E[m'_{\theta t}(e'_i m_{it} + e_i m'_t)]\Omega\zeta_T\zeta'_T\Omega e_i + O_p\left(\frac{1}{T}\right).
\end{aligned}$$

Taking expectations of the leading seven terms, we obtain the first and third conclusions using Lemma 2. The bias expression for CU is a special case when $h_3 = 0$. Finally, the last conclusion follows from the observation that under the additional condition, $E[m_t m'_{t+u}\Omega m_t] = 0$ and $E[m'_{\theta t}\Omega m_{t+u}] = 0$ when $u > 0$. *Q.E.D.*

REFERENCES

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

BAO, Y., AND A. ULLAH (2003): "The Second-Order Bias and Mean Squared Error of Estimators in Time Series Models," Unpublished Manuscript, University of California–Riverside.

DONALD, S., AND W. K. NEWEY (2000): "A Jacknife Interpretation of the Continuous Updating Estimator," *Economics Letters*, 67, 239–243.

GÖTZE, F., AND H. R. KÜNSCH (1996): "Second-Order Correctness of the Blockwise Bootstrap for Stationary Observations," *The Annals of Statistics*, 24, 1914–1933.

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

HANSEN, L. P., J. HEATON, AND A. YARON (1996): "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14, 262–280.

IMBENS, G. W. (1997): "One-Step Estimators for Over-Identified Generalized Method of Moments Models," *Review of Economic Studies*, 64, 359–383.

IMBENS, G. W., R. H. SPADY, AND P. JOHNSON (1998): "Information Theoretic Approaches to Inference in Moment Condition Models," *Econometrica*, 66, 333–357.

INOUE, A., AND M. SHINTANI (2003): "Bootstrapping GMM Estimators for Time Series," *Journal of Econometrics*, forthcoming.

KITAMURA, Y., AND M. STUTZER (1997): "An Information-Theoretic Alternative to Generalized Method of Moments Estimation," *Econometrica*, 65, 861–874.

NAGAR, A. L. (1959): "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations," *Econometrica*, 27, 573–595.

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

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

PHILLIPS, P. C. B. (1977): "Approximations to Some Finite Sample Distributions Associated with a First-Order Stochastic Difference Equation," *Econometrica*, 45, 463–486.

QIN, J., AND J. LAWLESS (1994): "Empirical Likelihood and General Estimating Equations," *The Annals of Statistics*, 22, 300–325.

RILSTONE, P., V. K. SRIVASTAVA, AND A. ULLAH (1996): "The Second-Order Bias and Mean-Squared Error of Nonlinear Estimators," *Journal of Econometrics*, 75, 369–395.

ROSENBLATT, M., AND J. W. VAN NESS (1965): "Estimation of the Bispectrum," *The Annals of Mathematical Statistics*, 36, 1120–1136.

SMITH, R. J. (1997): "Alternative Semi-Parametric Likelihood Approaches to Generalized Method of Moments Estimation," *Economic Journal*, 107, 503–519.

——— (2001): "GEL Criteria for Moment Condition Models," Unpublished Manuscript, University of Bristol.