Econometric Theory, **22**, 2006, 513–527. Printed in the United States of America.
DOI: 10.1017/S0266466606060257

GENERALIZED EMPIRICAL LIKELIHOOD INFERENCE FOR NONLINEAR AND TIME SERIES MODELS UNDER WEAK IDENTIFICATION

TAISUKE OTSU
Cowles Foundation, Yale University

This paper studies robust inference methods for nonlinear moment restriction models with weakly identified parameters in time series contexts. Our methods are based on generalized empirical likelihood with kernel smoothing. The proposed test statistics, which follow the standard χ^2 limiting distributions, are robust to weak identification and dependent data.

1. INTRODUCTION

This paper studies robust inference methods for nonlinear moment restriction models with weakly identified parameters in time series contexts. We propose two robust test statistics based on generalized empirical likelihood (GEL) by Newey and Smith (2004). The first test statistic, the S_{GEL}-statistic, is the kernel-smoothed GEL criterion function of Smith (2004) itself.[1] Kernel smoothing to moment restrictions induces a similar effect as heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimation. The second test statistic, the K_{GEL}-statistic, is derived from a GEL criterion function that uses transformed moment restrictions whose dimension is identical to the number of parameters. We derive the asymptotic properties of the S_{GEL}- and K_{GEL}-statistics under the null and fixed alternative hypotheses and show that (i) the S_{GEL}-statistic is robust to weak identification and dependent data and (ii) the K_{GEL}-statistic is also robust to weak identification and dependent data and its limiting distribution does not depend on the number of moment restrictions (or instrumental variables). A small simulation study based on a linear instrumental variables time series model indicates that under dependent data, kernel smoothing should be used to avoid serious size distortions.

The author is deeply grateful to Bruce Hansen, John Kennan, and Gautam Tripathi for their guidance and time. Comments from a coeditor and two anonymous referees substantially helped this revision. The author also thanks Allan Gregory, Patrik Guggenberger, Philip Haile, Hiroyuki Kasahara, Matthew Kim, Yuichi Kitamura, and seminar participants at Queen's University, University of Wisconsin, and the 2003 North America Summer Meeting of the Econometric Society for helpful discussions and suggestions. Financial support from the Alice Gengler Wisconsin Distinguished Graduate Fellowship and Wisconsin Alumni Research Foundation Dissertation Fellowship is gratefully acknowledged. Address correspondence to Taisuke Otsu, e-mail: taisuke.otsu@yale.edu.

© 2006 Cambridge University Press 0266-4666/06 $12.00

https://doi.org/10.1017/S0266466606060257 Published online by Cambridge University Press

Recently, several papers closely related to ours have appeared. First, Guggenberger and Smith (2005a) proposed GEL-based robust test statistics to weak identification. Indeed, the outlines of our mathematical proofs follow those of Guggenberger and Smith. If the data are independent or the moment restrictions are martingale difference sequences, the test statistics of Guggenberger and Smith are asymptotically equivalent to ours. However, if the data are dependent and a HAC-type correction is required, the results of Guggenberger and Smith no longer hold. Therefore, this paper can be considered an extension of Guggenberger and Smith (2005a) to time series contexts.[2] Second, Kleibergen (2005) proposed a generalized method of moments–based (GMM-based) robust statistic, called the K-statistic, which is an extension of Kleibergen (2002) to dynamic and nonlinear setups. By incorporating HAC covariance matrix estimation, the K-statistic becomes asymptotically equivalent to ours. However, Kleibergen (2005) did not derive the power property of the K-statistic. On the other hand, we derive the asymptotic properties of our GEL-based statistics under the null and alternative hypotheses. Third, Moreira (2003) proposed a similar conditional likelihood ratio test with good power properties for independent and identically distributed (i.i.d.) linear instrumental variables models.[3] However, it is unclear if the Moreira (2003) test can be extended to non-i.i.d. or nonlinear models.[4] In contrast, our methods are robust to those situations. Finally, Guggenberger and Smith (2005b) recently extended Guggenberger and Smith (2005a) to time series contexts. Their test statistics are asymptotically equivalent to ours and show attractive finite-sample properties. This paper extends the work of Guggenberger and Smith (2005b) by consideration of (i) testing for subvectors of parameters, (ii) general kernel functions for smoothing moment restrictions, (iii) asymptotic properties under the null and fixed alternative hypotheses, and (iv) more primitive assumptions to derive the asymptotic results. An extensive simulation study for these methods is an important topic for future research. Another interesting difference from the preceding existing test statistics is that our K_{GEL}-statistic is not necessarily quadratic. Therefore, although it is beyond the scope of this paper, we can expect that the higher order properties of these statistics are different.

We use the following notation: $\|\cdot\|$ is the Frobenius norm; $\mathrm{vec}(\cdot)$ is the column vectorization operator; $\mathrm{int}(S)$ is the interior of a set S; $\lambda_{\min}(A)$ is the smallest eigenvalue of a matrix A in absolute value. The symbols $\xrightarrow{d}$ and $\xrightarrow{p}$ mean convergence in distribution and probability, respectively. The abbreviations a.s. and w.p.a.1 mean "almost surely" and "with probability approaching one," respectively.

2. SETUP

Our notation closely follows that of Guggenberger and Smith (2005a) and Smith (2004). Let $\{z_t\}_{t=1}^T$ be a sequence of $l \times 1$ random vectors, $\theta \in \Theta \subset R^p$ be a

https://doi.org/10.1017/S0266466606060257 Published online by Cambridge University Press

$p \times 1$ vector of parameters, θ_0 be the true value of θ, and $g : R^l \times \Theta \to R^m$ be a vector of known functions, where $m \geq p$. Dynamic and nonlinear unconditional moment restriction models are written as

$$E[g_t(\theta_0)] \equiv E[g(z_t, \theta_0)] = 0, \tag{1}$$

for $t = 1, \ldots, T$, where $g_t(\theta)$ may depend on T (see Assumption 1, which follows). Let $\theta = (\alpha_1', \alpha_2', \beta')' \in \mathcal{A}_1 \times \mathcal{A}_2 \times \mathcal{B}$, where α_1, α_2, and β are $p_{\alpha_1} \times 1$, $p_{\alpha_2} \times 1$, and $p_\beta \times 1$ subvectors, respectively. Suppose that $\alpha \equiv (\alpha_1', \alpha_2')'$ are parameters of interest and β are nuisance parameters. Let $\theta_0 = (\alpha_{01}', \alpha_{02}', \beta_0')'$. We consider the following general identification setup of Stock and Wright (2000) and Guggenberger and Smith (2005a).

Assumption 1. Assume that

(i) Θ is compact and $\theta_0 \in int(\Theta)$;
(ii) $E[T^{-1} \sum_{t=1}^{T} g_t(\theta)] = T^{-1/2} m_{1T}(\theta) + m_2(\alpha_2, \beta)$, where m_{1T}, $m_1 : \Theta \to R^m$ and $m_2 : \mathcal{A}_2 \times \mathcal{B} \to R^m$ are continuous functions such that $m_{1T}(\theta) \to m_1(\theta)$ uniformly on Θ, $m_1(\theta_0) = 0$, and $m_2(\alpha_2, \beta) = 0$ if and only if $(\alpha_2, \beta) = (\alpha_{02}, \beta_0)$;
(iii) $m_2(\alpha_2, \beta)$ is continuously differentiable in a neighborhood of (α_{02}, β_0) and $\partial m_2(\alpha_{02}, \beta_0)/\partial(\alpha_2', \beta')$ has full column rank $p_{\alpha_2} + p_\beta$.

This assumption implies that (α_{02}, β_0) is strongly identified, whereas α_{01} is weakly identified, i.e., $E[g_t(\alpha_1, \alpha_{02}, \beta_0)]$ can be close to zero even for $\alpha_1 \neq \alpha_{01}$. Our setup includes the no nuisance parameter case (i.e., $p_\beta = 0$) and the no strongly identified parameter case (i.e., $p_{\alpha_2} = 0$) as special cases.[5] This paper considers GEL-based robust testing for $H_0 : \alpha = \alpha_0$ under Assumption 1 in time series contexts.

We now introduce GEL. First, to deal with time series data, we consider the kernel smoothed moment restrictions by Kitamura and Stutzer (1997) and Smith (2004):

$$g_{tT}(\theta) \equiv S_T^{-1} \sum_{s=t-T}^{t-1} k\left(\frac{s}{S_T}\right) g_{t-s}(\theta), \tag{2}$$

for $t = 1, \ldots, T$, where S_T is a bandwidth parameter and $k(\cdot)$ is a real-valued kernel function. As we will see subsequently, smoothing of the moment restrictions in GEL induces a HAC-type modification to estimating the long-run covariance matrix for the process of $g_t(\theta)$.

Next, let $\rho(\cdot)$ be a concave function on its domain Q. Denote λ as an $m \times 1$ parameter vector with the parameter space $\hat{\Lambda}_T(\theta) \equiv \{\lambda \in R^m : \kappa\lambda' g_{tT}(\theta) \in Q$

https://doi.org/10.1017/S0266466606060257 Published online by Cambridge University Press

for $t = 1, \ldots, T\}$, where $\kappa \equiv \kappa_1/\kappa_2$ is a constant for normalization, $\kappa_1 \equiv \int k(x)\,dx$, and $\kappa_2 \equiv \int k(x)^2\,dx$. A dependent version of GEL by Smith (2004) is defined as

$$S_{GEL}(\theta) \equiv \sup_{\lambda \in \hat{\Lambda}_T(\theta)} \hat{P}(\theta, \lambda) \equiv \sup_{\lambda \in \hat{\Lambda}_T(\theta)} \frac{1}{T} \sum_{t=1}^{T} \rho(\kappa \lambda' g_{tT}(\theta)). \tag{3}$$

As special cases, $S_{GEL}(\theta)$ includes (i) continuous updating: $\rho(v) = -(1 + v)^2/2$, (ii) empirical likelihood: $\rho(v) = \ln(1 - v)$, and (iii) exponential tilting: $\rho(v) = -\exp v$.

To compute $S_{GEL}(\theta)$, we must choose the kernel function $k(\cdot)$ and the bandwidth parameter S_T. Slight modifications of Smith (2004, Lem. A.3) and Smith (2005, Thm. 2.1) imply that $\kappa_2^{-1}(S_T/T)\sum_{t=1}^{T} g_{tT}(\theta_0) g_{tT}(\theta_0)'$ is asymptotically equivalent to the HAC covariance estimator with the kernel function $k^*(x) \equiv \int k(x - y)k(y)\,dy$ for the long-run covariance matrix $\Omega(\theta_0) \equiv \lim_{T\to\infty} \mathrm{Var}\{T^{-1/2} \sum_{t=1}^{T} g_t(\theta_0)\}$. Therefore, the kernel $k(\cdot)$ should be chosen so that the induced kernel $k^*(\cdot)$ will be a reasonable kernel function for estimating $\Omega(\theta_0)$.[6] In our simulation study, we adopt the truncated kernel, $k(x) = I(|x| \leq 1)$, where the induced kernel is the Bartlett kernel, $k^*(x) = I(|x| \leq 2)(1 - |x/2|)$. Setting $S_T = (2\ell_T + 1)/2$, the smoothed moment restrictions are written as $g_{tT}(\theta) = 2(2\ell_T + 1)^{-1} \sum_{s=\max\{t-T, -\ell_T\}}^{\min\{t-1, \ell_T\}} g_{t-s}(\theta)$.

3. MAIN RESULTS

3.1. S_{GEL}-Statistic

Let $\hat{\theta}_\alpha \equiv (\alpha, \hat{\beta}(\alpha))$, where $\hat{\beta}(\alpha)$ is the GEL estimator for β_0 given α. The S_{GEL}-statistic for testing $H_0: \alpha = \alpha_0$ is defined as

$$\hat{S}_{GEL}(\alpha) \equiv S_{GEL}(\hat{\theta}_\alpha) \equiv \min_{\beta \in \mathcal{B}} S_{GEL}(\alpha, \beta). \tag{4}$$

Thus, the S_{GEL}-statistic is the GEL criterion function itself, in which the (strongly identified) nuisance parameters β are profiled out. Denote $\rho_1(v) \equiv d\rho(v)/dv$, $\rho_2(v) \equiv d^2\rho(v)/dv^2$, $\theta_{a\beta} \equiv (a', \beta')'$, $a \equiv (a_1', \alpha_{02}')'$, and $\mathcal{N}$ as some closed neighborhood of β_0. To derive the asymptotic distribution of $\hat{S}_{GEL}(a)$, we impose the following assumptions.

Assumption 2. $\rho: Q \to R$ is second-order continuously differentiable in a neighborhood of 0 and is strictly concave on Q, where Q is an open interval containing 0 and $\rho_1(0) = \rho_2(0) = -1$.

Assumption 3. $\{z_t\}_{t=1}^{T}$ is stationary and strong mixing with mixing coefficients $\sum_{j=1}^{\infty} j^2 \alpha(j)^{(\nu-1)/\nu} < \infty$ for some $\nu > 1$.

https://doi.org/10.1017/S0266466606060257 Published online by Cambridge University Press

Assumption 4. Assume that

(i) $S_T = O(T^{1/2-\delta})$ for some $\frac{1}{6} < \delta < \frac{1}{2}$ and $S_T \to \infty$;
(ii) $k: R \to [-\bar{k}, \bar{k}]$ for some $\bar{k} < \infty$, $k(0) \neq 0$, $\kappa_1 \neq 0$, $k(x)$ is continuous at 0 and almost everywhere, $|(2\pi)^{-1}\int_{-\infty}^{\infty} \exp(-ix\omega)k(x)\,dx| \geq 0$ for each $\omega \in R$, and $\int_{-\infty}^{0} \sup_{y \leq x}|k(y)|\,dx + \int_{0}^{\infty} \sup_{y \geq x}|k(y)|\,dx < \infty$.

Assumption 5. Assume that

(i) $g_t(\theta_{a\beta})$ is continuous at $\theta_{a\beta}$ a.s. for each $\beta \in \mathcal{B}$ and $E[\sup_{\beta \in \mathcal{B}}\|g_t(\theta_{a\beta})\|^{\eta}] < \infty$ for some $\eta > \max\{4\nu, 1/\delta\}$;
(ii) $\inf_{\beta \in \mathcal{B}} \lambda_{\min}(\Omega(\theta_{a\beta})) \geq \epsilon$ for some $\epsilon > 0$ and $\Omega(\theta_{a\beta})$ is uniformly bounded on $\beta \in \mathcal{B}$ and is continuous at $\theta_{a\beta_0}$;
(iii) $\partial g_t(\theta_{a\beta})/\partial\beta'$ is continuous at $\theta_{a\beta}$ a.s. for each $\beta \in \mathcal{N}$, $E[\sup_{\beta \in \mathcal{N}}\|\partial g_t(\theta_{a\beta})/\partial\beta'\|^2] < \infty$, $\partial m_{1T}(\theta_{a\beta})/\partial\beta'$ is continuous at $\theta_{a\beta}$ for each $\beta \in \mathcal{N}$, and $\partial m_{1T}(\theta_{a\beta})/\partial\beta'$ converges to some function uniformly on $\beta \in \mathcal{N}$.

Similar comments from Smith (2004) apply. In contrast to Guggenberger and Smith (2005a) and Kleibergen (2005), we provide more primitive conditions. Assumption 3 allows for weakly dependent data. Assumption 4 contains standard conditions for S_T and $k(\cdot)$. The condition $\frac{1}{6} < \delta$ is required in the proof of Lemma 2(iii) in the Appendix. The last part of Assumption 4(ii) is required to avoid pathological cases (see Jansson, 2002). Let $\kappa^* \equiv \kappa_2/\kappa_1^2$, $M_{2\beta} \equiv \partial m_2(\alpha_{02}, \beta_0)/\partial\beta'$, and $\Delta(\theta_{a\beta_0}) \equiv I_m - M_{2\beta}[M_{2\beta}'\Omega(\theta_{a\beta_0})^{-1}M_{2\beta}]^{-1}$ $M_{2\beta}'\Omega(\theta_{a\beta_0})^{-1}$. The limiting distribution of the S_{GEL}-statistic is obtained as follows.

THEOREM 1. *Suppose that Assumptions 1–5 hold. Then*

$$2\kappa^* T S_T^{-1}(\hat{S}_{GEL}(a) - \rho(0)) \xrightarrow{d} \chi^2(m - p_\beta, \Upsilon) \quad \textit{and}$$

$$2\kappa^* T S_T^{-1}(\hat{S}_{GEL}(\alpha_0) - \rho(0)) \xrightarrow{d} \chi^2(m - p_\beta),$$

where $\Upsilon \equiv m_1(\theta_{a\beta_0})'\Omega(\theta_{a\beta_0})^{-1}\Delta(\theta_{a\beta_0})m_1(\theta_{a\beta_0})$ *is the noncentrality parameter.*

Remark 1. The first part of Theorem 1 implies that the test for $H_0: \alpha = \alpha_0$ based on the S_{GEL}-statistic is inconsistent. The inconsistency results of hypothesis testing are common in the weak identification setup. However, we can construct a robust confidence set for α_0, i.e., $\{\alpha \in \mathcal{A}: 2\kappa^* T S_T^{-1}(\hat{S}_{GEL}(\alpha) - \rho(0)) \leq \chi_r^2(m - p_\beta)\}$, where $\chi_r^2(m - p_\beta)$ is the $(1 - r)$-critical value of $\chi^2(m - p_\beta)$. The dimensions of α_1 and α_2 have no effect on the limiting distribution.

Remark 2. Theorem 1 is a generalization of Guggenberger and Smith (2005a, Thm. 5) to our time series setup. Guggenberger and Smith (2005a, Assump. M_α(ii)) assumed that $T^{-1}\sum_{t=1}^{T} g_t(\theta_{a\beta})g_t(\theta_{a\beta})' \xrightarrow{p} \Omega(\theta_{a\beta})$ uniformly on $\beta \in \mathcal{B}$,

https://doi.org/10.1017/S0266466606060257 Published online by Cambridge University Press

which is not typically satisfied in dependent data. If this assumption is violated, Guggenberger and Smith's asymptotic result for their $GELR_\rho$-statistic no longer holds. Besides, they employed the GEL estimator without kernel smoothing, which is generally inefficient in our time series setup. Because of kernel smoothing, the S_{GEL}-statistic is robust to dependent data.

3.2. K_{GEL}-Statistic

Theorem 1 indicates that the limiting distribution of the S_{GEL}-statistic depends on the number of moment restrictions m (i.e., the dimension of $g_t(\theta)$). If m is large, the power of the S_{GEL}-statistic may be low.[7] To avoid such dependence on m, we transform the moment restrictions from $g_{tT}(\hat{\theta}_\alpha)$ to $B(\hat{\theta}_\alpha)g_{tT}(\hat{\theta}_\alpha)$, where $B(\hat{\theta}_\alpha)$ is a $p \times m$ matrix and p is the number of parameters. Let $\Lambda \equiv diag(T^{1/2},\ldots,T^{1/2},1,\ldots,1)$, where the first p_{α_1} diagonal elements are $T^{1/2}$ and the remaining diagonal elements are ones. Based on Guggenberger and Smith (2005a), we specify the transformation matrix as

$$B(\hat{\theta}_\alpha) \equiv D^*_{\rho T}(\hat{\theta}_\alpha)'\hat{\Omega}_T(\hat{\theta}_\alpha)^{-1}$$
$$\equiv \Lambda\left(\frac{1}{T}\sum_{t=1}^{T}\rho_1(\kappa\lambda(\hat{\theta}_\alpha)'g_{tT}(\hat{\theta}_\alpha))G_{tT}(\hat{\theta}_\alpha)'\right)\hat{\Omega}_T(\hat{\theta}_\alpha)^{-1},$$

where $\hat{\Omega}_T(\hat{\theta}_\alpha)$ is some consistent estimator for $\Omega(\theta_{\alpha\beta_0})$ such as some HAC covariance matrix estimator and the Smith (2005) automatic HAC estimator (see also Lemma 1(ii)) in the Appendix, $G_{tT}(\hat{\theta}_\alpha) \equiv \partial g_{tT}(\hat{\theta}_\alpha)/\partial\theta'$, and $\lambda(\hat{\theta}_\alpha) \equiv \arg\sup_{\lambda\in\hat{\Lambda}_T(\hat{\theta}_\alpha)}\hat{P}(\hat{\theta}_\alpha,\lambda)$, which exists w.p.a.1 from Lemma 2(ii) in the Appendix. This transformation is based on the first-order condition (FOC) for θ in minimizing $S_{GEL}(\theta)$. A crucial requirement is that $D^*_{\rho T}(\hat{\theta}_\alpha)$ is asymptotically independent of $\lambda(\hat{\theta}_\alpha)$.[8] The K_{GEL}-statistic for testing $H_0\colon \alpha = \alpha_0$ is defined as

$$\hat{K}_{GEL}(\alpha) \equiv K_{GEL}(\hat{\theta}_\alpha) \equiv \sup_{\gamma\in\hat{\Gamma}_T(\hat{\theta}_\alpha)}\hat{P}_K(\hat{\theta}_\alpha,\gamma)$$
$$\equiv \sup_{\gamma\in\hat{\Gamma}_T(\hat{\theta}_\alpha)}\frac{1}{T}\sum_{t=1}^{T}\rho(\kappa\gamma' D^*_{\rho T}(\hat{\theta}_\alpha)'\hat{\Omega}_T(\hat{\theta}_\alpha)^{-1}g_{tT}(\hat{\theta}_\alpha)), \tag{5}$$

where $\hat{\Gamma}_T(\theta) \equiv \{\gamma \in R^p\colon \kappa\gamma' D^*_{\rho T}(\theta)'\hat{\Omega}_T(\theta)^{-1}g_{tT}(\theta) \in Q \text{ for } t = 1,\ldots,T\}$. Let $G_{\alpha_1 t}(\theta) \equiv \partial g_t(\theta)/\partial\alpha_1'$, $G_{\alpha_2 t}(\theta) \equiv \partial g_t(\theta)/\partial\alpha_2'$, $G_{\beta t}(\theta) \equiv \partial g_t(\theta)/\partial\beta'$, and $V(\theta) \equiv \begin{pmatrix}\Omega_{\alpha_1}(\theta) & \Omega_{\alpha_1 g}(\theta)\\ \Omega_{\alpha_1 g}(\theta)' & \Omega(\theta)\end{pmatrix} \equiv \lim_{T\to\infty}\mathrm{Var}\begin{pmatrix}T^{-1/2}\sum_{t=1}^{T}\mathrm{vec}G_{\alpha_1 t}(\theta)\\ T^{-1/2}\sum_{t=1}^{T}g_t(\theta)\end{pmatrix}$. We add the following assumption.

Assumption 6. Assume that

(i) $G_{\alpha_2 t}(\theta_{a\beta})$ is continuous at $\theta_{a\beta}$ a.s. for each $\beta \in \mathcal{N}$, $E[\sup_{\beta\in\mathcal{N}}\|G_{\alpha_2 t}(\theta_{a\beta})\|] < \infty$, $\partial m_{1T}(\theta_{a\beta})/\partial\alpha_2'$ is continuous at $\theta_{a\beta}$ for

https://doi.org/10.1017/S0266466606060257 Published online by Cambridge University Press

each $\beta \in \mathcal{N}$, and $\partial m_{1T}(\theta_{a\beta})/\partial \alpha_2'$ converges to some function uniformly on $\beta \in \mathcal{N}$;

(ii) $G_{\alpha_1 t}(\theta_{a\beta})$ is continuous at $\theta_{a\beta}$ a.s. for each $\beta \in \mathcal{N}$, $E[\sup_{\beta\in\mathcal{N}}\|G_{\alpha_1 t}(\theta_{a\beta})\|^{\eta}] < \infty$ for some $\eta > \max\{4\nu, 1/\delta\}$, $\partial \operatorname{vec} G_{\alpha_1 t}(\theta_{a\beta})/\partial\beta'$ is continuous at $\theta_{a\beta}$ a.s. for each $\beta \in \mathcal{N}$, $E[\sup_{\beta\in\mathcal{N}}\|\partial \operatorname{vec} G_{\alpha_1 t}(\theta_{a\beta})/\partial\beta'\|^2] < \infty$, $\partial \operatorname{vec}(\partial m_{1T}(\theta_{a\beta})/\partial\alpha_1')/\partial\beta' \to \partial \operatorname{vec}(\partial m_1(\theta_{a\beta})/\partial\alpha_1')/\partial\beta'$ uniformly on $\beta \in \mathcal{N}$, and $V(\theta_{a\beta})$ is positive definite and continuous at $\theta_{a\beta_0}$;

(iii) $\hat{\Omega}_T(\hat{\theta}_a) \xrightarrow{p} \Omega(\theta_{a\beta_0})$.

A crucial assumption is Assumption 6(ii), which ensures a central limit theorem for $T^{-1/2}\sum_{t=1}^{T} \operatorname{vec} G_{\alpha_1 t}(\theta)$ (see Lemma 3(iii) in the Appendix). This assumption is not required in the standard GMM setup. The limiting distribution of the K_{GEL}-statistic is obtained as follows.

THEOREM 2. *Suppose that Assumptions 1–6 hold. Then*

$$2\kappa^* T S_T^{-1}(\hat{K}_{GEL}(a) - \rho(0)) \xrightarrow{d} (\Upsilon_K(a) + \xi_\alpha)'(\Upsilon_K(a) + \xi_\alpha), \quad \textit{and}$$

$$2\kappa^* T S_T^{-1}(\hat{K}_{GEL}(\alpha_0) - \rho(0)) \xrightarrow{d} \chi^2(p_\alpha),$$

where $\Upsilon_K(a)$ *is defined in (A.9) in the Appendix,* $\xi_\alpha \sim N(0, I_{p_\alpha})$, *and* $\Upsilon_K(a)$ *and* ξ_α *are independent.*

Remark 3. The limiting distribution of $\hat{K}_{GEL}(a)$ does not depend on m. As is the case with the S_{GEL}-statistic, the test for $H_0: \alpha = \alpha_0$ based on the K_{GEL}-statistic is inconsistent. However, we can construct a robust confidence set for α_0 as $\{\alpha \in \mathcal{A} : 2\kappa^* T S_T^{-1}(\hat{K}_{GEL}(\alpha) - \rho(0)) \leq \chi_r^2(p_\alpha)\}$. If p_α is fixed, the dimensions of α_1 and α_2 have no effect on the limiting distribution of $\hat{K}_{GEL}(a)$.

Remark 4. Theorem 2 is a generalization of Guggenberger and Smith (2005a, Thm. 6) to our time series setup. In addition to Assumption M_α(ii) of Guggenberger and Smith (2005a), Assumption M_α(viii) assumed that $T^{-1}\sum_{t=1}^{T} \operatorname{vec} G_{\alpha_1 t}(\theta_{a\beta}) g_t(\theta_{a\beta})' \xrightarrow{p} \Omega_{\alpha_1 g}(\theta_{a\beta})$ uniformly on $\beta \in \mathcal{N}$. Neither of these assumptions is typically satisfied in dependent data. If these assumptions are violated, the Guggenberger and Smith asymptotic results for their S_ρ- and LM_ρ-statistics no longer hold. Because of kernel smoothing, the K_{GEL}-statistic is robust to dependent data.

Remark 5. For i.i.d. data, we can use the unsmoothed moment restrictions $g_t(\theta)$ instead of $g_{tT}(\theta)$ for defining the S_{GEL}- and K_{GEL}-statistics. Under similar assumptions, results analogous to Theorems 1 and 2 hold.

https://doi.org/10.1017/S0266466606060257 Published online by Cambridge University Press

3.3. Simulation Results

To highlight the importance of kernel smoothing, we focus on the linear instrumental variables time series model with a nuisance parameter:

$$y_t = \alpha x_{1t} + \beta x_{2t} + u_t, \qquad u_t = \rho_u u_{t-1} + \epsilon_{ut},$$

$$(x_{1t}, x_{2t})' = \Pi z_t + \epsilon_{xt}, \qquad z_t = \rho_z I_m z_{t-1} + \epsilon_{zt},$$

for $t = 1, \ldots T$, where $\Pi = \left[\begin{smallmatrix} 0.1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 1 \end{smallmatrix}\right]$, $\left[\begin{smallmatrix} \epsilon_{ut} \\ \epsilon_{xt} \end{smallmatrix}\right] \sim N\left(0, \left[\begin{smallmatrix} 1 & 0.8 & 0.8 \\ 0.8 & 1 & 0.3 \\ 0.8 & 0.3 & 1 \end{smallmatrix}\right]\right)$, and $\epsilon_{zt} \sim N(0, I_m)$. The true parameters are $\alpha_0 = 0$ (weakly identified parameter of interest) and $\beta_0 = 1$ (strongly identified nuisance parameter), and β is estimated by the corresponding GEL or GMM estimator. For illustration purposes, we only report the case of $T = 100$ and $m = 10$.[9] The number of replication is 10,000.[10] We report the p-value plot by Davidson and MacKinnon (1998), where the x-axis shows probability values and the y-axis shows actual sizes of the test

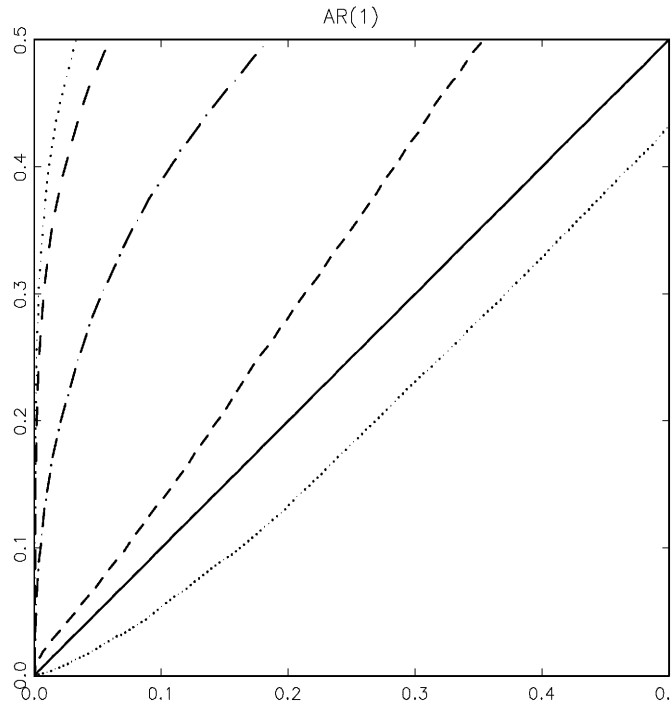


FIGURE 1. P-value plots for the LM_{EL} (dashed), K_{EL} (dotted), K_{EL} with $\ell_T = 2$ (short dashes), K (dots and dashes), and K with $\ell_T = 2$ (closely spaced dots) tests.

https://doi.org/10.1017/S0266466606060257 Published online by Cambridge University Press

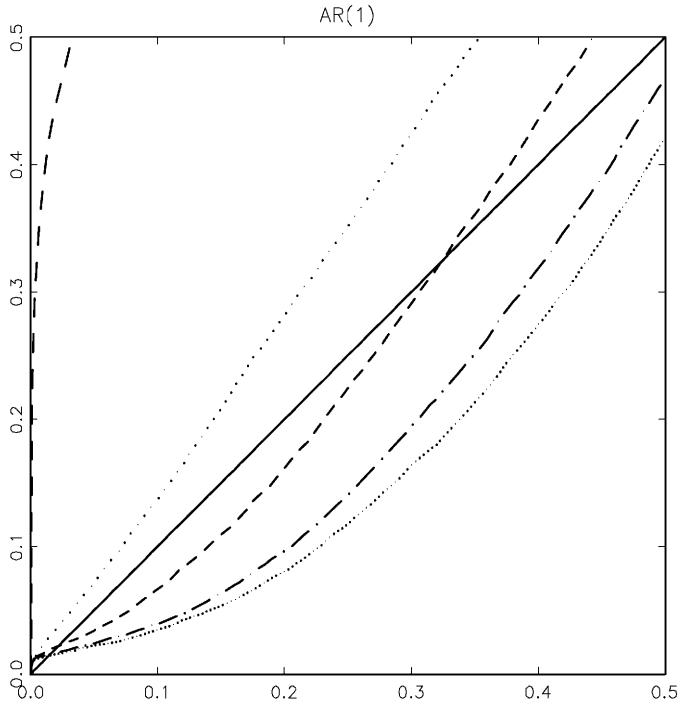


FIGURE 2. *P*-value plots for the K_{EL} tests with $\ell_T = 0$ (dashed), $\ell_T = 2$ (dotted), $\ell_T = 4$ with $\ell_T = 6$ (short dashes), and automatic ℓ_T (dots and dashes).

statistics. If the *p*-value plot is close to the 45-degree line, the test statistic is well approximated by the limiting distribution.

Figures 1 and 2 report the *p*-value plots for the case of $\rho_u = \rho_z = 0.9$. Figure 1 illustrates the *p*-value plots for the LM_{EL}, unsmoothed K_{EL}, smoothed K_{EL} with $\ell_T = 2$, unsmoothed K, and smoothed K with $\ell_T = 2$ tests.[11] We find that (i) the LM_{EL}, unsmoothed K_{EL}, and unsmoothed K tests show severe overrejection and (ii) the smoothed K_{EL} and K tests with $\ell_T = 2$ show better size properties, particularly in smaller probability values (less than 0.10).[12] Thus, under time series data, we suggest the use of the smoothed K_{EL} or K test. This robustness to dependent data is compatible with our theoretical result. Figure 2 illustrates the *p*-value plots for the K_{EL} test with different bandwidth parameters ($\ell_T = 0,2,4,6$ and Newey and West, 1994, automatic choice). We find that (i) as ℓ_T increases, the actual size decreases; and (ii) in smaller probability values (less than 0.05), the size properties are less sensitive to the choice of ℓ_T. Although it is outside the scope of this paper, it would be interesting to make a comparison with adapted versions of the Guggenberger and Smith (2005b) full vector tests or the Guggenberger and Wolf (2004) subsampling tests in this setup.

https://doi.org/10.1017/S0266466606060257 Published online by Cambridge University Press

NOTES

1. Kitamura (1997) and Kitamura and Stutzer (1997) employed kernel smoothed moment restrictions for empirical likelihood and exponential tilting, respectively. Smith (2004) generalized this idea to the GEL context.

2. Caner (2003) also proposed robust test statistics based on exponential tilting. Caner (2003) assumed independent data, and his results are similar to those of Guggenberger and Smith (2005a).

3. Andrews, Moreira, and Stock (2004) proposed optimal power properties of their similar tests.

4. In a simulation study, Guggenberger and Smith (2005a) reported that the Moreira (2003) test tends to overreject the null hypothesis under conditional heteroskedasticity.

5. See Stock and Wright (2000) for a detailed discussion of this assumption.

6. Although it is beyond the scope of this paper, the optimal choice of $k(\cdot)$ and S_T is an interesting challenge. Smith (2004, Sect. 2.6) considered an optimal bandwidth, which minimized the mean squared error in estimating the long-run covariance matrix. In our simulation study, we employ the Newey and West (1994) automatic bandwidth choice rule. However, because our main interest is to test $H_0: \alpha_0 = \alpha$ (or construct a valid confidence set for α_0), it is important to consider other optimality criteria, such as the error in coverage probability of the GEL-based confidence sets.

7. Moreover, our preliminary simulation results illustrate that the size property of the S_{GEL}-statistic is sensitive to m.

8. Instead of $D^*_{\rho T}(\hat{\theta}_\alpha)$, we may use the robust Jacobian matrix in Kleibergen (2005). Differences in asymptotically equivalent choices of $B(\hat{\theta}_\alpha)$ have no effect on the first-order asymptotic properties of the K_{GEL}-statistic. The optimal choice of $B(\hat{\theta}_\alpha)$ based on the higher order properties is a challenging topic. In our preliminary simulation (not reported), the size and power properties of the GEL-based tests are not sensitive to the choice of $B(\hat{\theta}_\alpha)$.

9. More extensive results are available from the author upon request. We also tried the cases of $T = 25$, 50, and 250, $m = 5$, and independent MA(1), $t(2)$, and $\chi^2(1)$ error terms. In the working paper version, we also considered the Wald, GMM-LM, Anderson–Rubin, homoskedastic K, S, and S_{EL} test statistics. Although we analyzed the exponential tilting–based tests, the results were similar to those of the empirical likelihood–based tests.

10. In terms of computational time, the S, K, and LM_ρ tests are faster than the S_{EL} and K_{EL} tests. However, even when $n = 250$, each Monte Carlo replication takes less than one minute on a Pentium 4 CPU 1.38GHz.

11. We employ the truncated kernel for the kernel function $k(\cdot)$ and the Bartlett kernel for the HAC covariance estimator.

12. For the MA(1) case, we obtained similar results, but the magnitudes of overrejection for the unsmoothed tests were smaller than in the AR(1) case.

REFERENCES

Andrews, D.W.K., M. Moreira, & J.H. Stock (2004) Optimal Invariant Similar Tests for Instrumental Variables Regression. Working paper, Yale University.

Caner, M. (2003) Exponential Tilting with Weak Instruments: Estimation and Testing. Working paper, University of Pittsburgh.

Davidson, R. & J.G. MacKinnon (1998) Graphical methods for investigating the size and power of hypothesis tests. *Manchester School* 66, 1–26.

Guggenberger, P. & R.J. Smith (2005a) Generalized empirical likelihood estimators and tests under partial, weak and strong identification. *Econometric Theory* 21, 667–709.

Guggenberger, P. & R.J. Smith (2005b) Generalized Empirical Likelihood Tests in Time Series Models with Potential Identification Failure. Working paper, University College London.

https://doi.org/10.1017/S0266466606060257 Published online by Cambridge University Press

Guggenberger, P. & M. Wolf (2004) Subsampling Tests of Parameter Hypotheses and Overidentifying Restrictions with Possible Failure of Identification. Working paper, UCLA.
Jansson, M. (2002) Consistent covariance matrix estimation for linear processes. *Econometric Theory* 18, 1449–1459.
Kitamura, Y. (1997) Empirical likelihood methods with weakly dependent processes. *Annals of Statistics* 25, 2084–2102.
Kitamura, Y. & M. Stutzer (1997) An information-theoretic alternative to generalized method of moments estimation. *Econometrica* 65, 861–874.
Kleibergen, F. (2002) Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica* 70, 1781–1805.
Kleibergen, F. (2005) Testing parameters in GMM without assuming that they are identified. *Econometrica* 73, 1103–1123.
Moreira, M.J. (2003) A conditional likelihood ratio test for structural models. *Econometrica* 71, 1027–1048.
Newey, W.K. & R.J. Smith (2004) Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* 72, 219–256.
Newey, W.K. & K.D. West (1994) Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* 61, 631–654.
Smith, R.J. (2004) GEL Criteria for Moment Condition Models. Working paper, University College London.
Smith, R.J. (2005) Automatic positive semi-definite HAC covariance matrix and GMM estimation. *Econometric Theory* 21, 158–170.
Stock, J.H. & J.H. Wright (2000) GMM with weak identification. *Econometrica* 68, 1055–1096.

MATHEMATICAL APPENDIX

Notation. Let C be a generic positive constant, $\hat{g}(\theta) \equiv T^{-1}\sum_{t=1}^{T} g_t(\theta)$, $\hat{g}_T(\theta) \equiv T^{-1}\sum_{t=1}^{T} g_{tT}(\theta)$, $G_{\beta tT}(\theta) \equiv \partial g_{tT}(\theta)/\partial\beta'$, $\hat{G}_{\beta T}(\theta) \equiv T^{-1}\sum_{t=1}^{T} G_{\beta tT}(\theta)$, $G_{\alpha_1 tT}(\theta) \equiv \partial g_{tT}(\theta)/\partial\alpha_1'$, $\hat{G}_{\alpha_1 T}(\theta) \equiv T^{-1}\sum_{t=1}^{T} G_{\alpha_1 tT}(\theta)$, $G_{\alpha_2 tT}(\theta) \equiv \partial g_{tT}(\theta)/\partial\alpha_2'$, $\hat{G}_{\alpha_2 T}(\theta) \equiv T^{-1}\sum_{t=1}^{T} G_{\alpha_2 tT}(\theta)$, and $\Omega \equiv \Omega(\theta_{a\beta_0})$.

Proofs of the following lemmas are available from the author upon request.

LEMMA 1. *Suppose that Assumptions 1 and 3–5 hold. Then*

(i) $\sup_{\beta\in\mathcal{B}}\|\hat{g}_T(\theta_{a\beta}) - \kappa_1 E[\hat{g}(\theta_{a\beta})]\| \xrightarrow{p} 0$;
(ii) $\kappa_2^{-1}(S_T/T)\sum_{t=1}^{T} g_{tT}(\theta_{a\beta})g_{tT}(\theta_{a\beta})' \xrightarrow{p} \Omega$ *for* $\beta = \beta_0$ *and* $\beta = \hat{\beta}(a)$;
(iii) $T^{1/2}[\hat{g}_T(\theta_{a\beta_0}) - E[\hat{g}_T(\theta_{a\beta_0})]] \xrightarrow{d} N(0, \kappa_1^2\Omega)$.

LEMMA 2. *Suppose that Assumptions 1–5 hold. Then*

(i) for some $1/2\eta\delta < \zeta < \frac{1}{2}$ *and* $c > 0$, $\Lambda_T \equiv \{\lambda \in R^m : \|\lambda\| \le c(T^{-1}S_T^2)^{\zeta}\} \subset \hat{\Lambda}_T(\theta_{a\beta})$ *uniformly on* $\beta \in \mathcal{B}$ *w.p.a.1 and* $\sup_{\beta\in\mathcal{B},\lambda\in\Lambda_T,1\le t\le T}|\lambda' g_{tT}(\theta_{a\beta})| \xrightarrow{p} 0$;
(ii) for $\beta = \beta_0$ *and* $\hat{\beta}(a)$, *there exists* $\lambda(\theta_{a\beta}) \in \hat{\Lambda}_T(\theta_{a\beta})$ *w.p.a.1 such that* $\hat{P}(\theta_{a\beta}, \lambda(\theta_{a\beta})) = \sup_{\lambda\in\hat{\Lambda}_T(\theta_{a\beta})}\hat{P}(\theta_{a\beta}, \lambda)$, $\lambda(\theta_{a\beta}) = O_p(T^{-1/2}S_T)$, *and* $\hat{P}(\theta_{a\beta}, \lambda(\theta_{a\beta})) - \rho(0) = O_p(T^{-1}S_T)$;
(iii) $\sup_{\beta\in\mathcal{N}}\|\hat{G}_{\beta T}(\theta_{a\beta}) - \kappa_1(T^{-1/2}\partial m_{1T}(\theta_{a\beta})/\beta' + \partial m_2(\alpha_{02}, \beta)/\partial\beta')\| \xrightarrow{p} 0$, $\hat{\beta}(a) \xrightarrow{p} \beta_0$, $T^{1/2}(\hat{\beta}(a) - \beta_0) = O_p(1)$, *and* $\hat{g}_T(\hat{\theta}_a) = O_p(T^{-1/2})$.

https://doi.org/10.1017/S0266466606060257 Published online by Cambridge University Press

LEMMA 3. *Suppose that Assumptions 1 and 3–6(i)–(ii) hold. Then*

(i) $\sup_{\beta\in\mathcal{N}}\|\hat{G}_{\alpha_1 T}(\theta_{a\beta}) - \kappa_1 T^{-1/2}\partial m_{1T}(\theta_{a\beta})/\alpha_1'\| \xrightarrow{p} 0$ *and* $\sup_{\beta\in\mathcal{N}}\|\hat{G}_{\alpha_2 T}(\theta_{a\beta}) - \kappa_1(T^{-1/2}\partial m_{1T}(\theta_{a\beta})/\alpha_2' + \partial m_2(\alpha_{02},\beta)/\partial\alpha_2')\| \xrightarrow{p} 0$*;*

(ii) $\kappa_2^{-1}(S_T/T)\sum_{t=1}^{T}((\operatorname{vec} G_{\alpha_1 tT}(\theta_{a\beta}))', g_{tT}(\theta_{a\beta})')'((\operatorname{vec} G_{\alpha_1 tT}(\theta_{a\beta}))', g_{tT}(\theta_{a\beta})') \xrightarrow{p} V(\theta_{a\beta_0})$ *for* $\beta = \beta_0$ *and* $\beta = \hat{\beta}(a)$*;*

(iii) $T^{1/2}[((\operatorname{vec}\hat{G}_{\alpha_1 T}(\theta_{a\beta_0}))', \hat{g}_T(\theta_{a\beta_0})')' - E[((\operatorname{vec}\hat{G}_{\alpha_1 T}(\theta_{a\beta_0}))', \hat{g}_T(\theta_{a\beta_0})')']] \xrightarrow{d} N(0, \kappa_1^2 V(\theta_{a\beta_0}))$.

LEMMA 4. *Suppose that Assumptions 1–6 hold. Then*

(i) for some $1/2\eta\delta < \zeta < \frac{1}{2}$ *and* $c > 0$, $\Gamma_T \equiv \{\gamma \in R^p : \|\gamma\| \le c(T^{-1}S_T^2)^{\zeta}\} \subset \hat{\Gamma}_T(\hat{\theta}_a)$ *w.p.a.1 and* $\sup_{\gamma\in\Gamma_T, 1\le t\le T}|\gamma' D_{\rho T}^*(\hat{\theta}_a)'\hat{\Omega}_T(\hat{\theta}_a)^{-1} g_{tT}(\hat{\theta}_a)| \xrightarrow{p} 0$*;*

(ii) there exists $\gamma(\hat{\theta}_a) \in \operatorname{int}(\Gamma_T) \subset \hat{\Gamma}_T(\hat{\theta}_a)$ *w.p.a.1 such that* $\hat{P}_K(\hat{\theta}_a, \gamma(\hat{\theta}_a)) = \sup_{\gamma\in\hat{\Gamma}_T(\hat{\theta}_a)}\hat{P}_K(\hat{\theta}_a,\gamma)$.

Proof of Theorem 1. The outline of the proof is similar to that of Theorem 5 of Guggenberger and Smith (2005a). Lemma 2(ii) implies that the FOC for $\lambda(\hat{\theta}_a)$,

$$S_T T^{-1}\sum_{t=1}^{T}\rho_1(\kappa\lambda(\hat{\theta}_a)' g_{tT}(\hat{\theta}_a)) g_{tT}(\hat{\theta}_a) = 0, \tag{A.1}$$

holds w.p.a.1. An expansion of (A.1) around $\lambda(\hat{\theta}_a) = 0$ yields $0 = -S_T\hat{g}_T(\hat{\theta}_a) - \kappa_1\dot{\Omega}_{\lambda\theta}\lambda(\hat{\theta}_a)$, where $\dot{\Omega}_{\lambda\theta} \equiv -\kappa_2^{-1}S_T T^{-1}\sum_{t=1}^{T}\rho_2(\kappa\dot{\lambda}' g_{tT}(\hat{\theta}_a)) g_{tT}(\hat{\theta}_a) g_{tT}(\hat{\theta}_a)'$ and $\dot{\lambda}$ is a point on the line joining 0 and $\lambda(\hat{\theta}_a)$. Lemmas 1(ii) and 2(i) and $\rho_2(0) = -1$ imply that $\dot{\Omega}_{\lambda\theta} \xrightarrow{p} \Omega$. From Assumption 5(ii), $\dot{\Omega}_{\lambda\theta}$ is invertible w.p.a.1, and $\lambda(\hat{\theta}_a)$ is written as

$$\lambda(\hat{\theta}_a) = -\kappa_1^{-1} S_T\dot{\Omega}_{\lambda\theta}^{-1}\hat{g}_T(\hat{\theta}_a), \tag{A.2}$$

w.p.a.1. A second-order Taylor expansion of $\hat{P}(\hat{\theta}_a, \lambda(\hat{\theta}_a))$ around $\lambda(\hat{\theta}_a) = 0$ yields

$$\begin{aligned} 2\kappa_2 T S_T^{-1}(\hat{S}_{GEL}(a) - \rho(0)) &= 2T\hat{g}_T(\hat{\theta}_a)'\dot{\Omega}_{\lambda\theta}^{-1}\hat{g}_T(\hat{\theta}_a) - T\hat{g}_T(\hat{\theta}_a)'\dot{\Omega}_{\lambda\theta}^{-1}\ddot{\Omega}_{\lambda\theta}\dot{\Omega}_{\lambda\theta}^{-1}\hat{g}_T(\hat{\theta}_a) \\ &= T\hat{g}_T(\hat{\theta}_a)'\Omega^{-1}\hat{g}_T(\hat{\theta}_a) + o_p(1), \end{aligned} \tag{A.3}$$

where $\ddot{\Omega}_{\lambda\theta}$ is defined similarly to $\dot{\Omega}_{\lambda\theta}$ and the second equality follows from $\dot{\Omega}_{\lambda\theta}$, $\ddot{\Omega}_{\lambda\theta} \xrightarrow{p} \Omega$ and Lemma 2(iii). We next derive the asymptotic distribution of $T^{1/2}\hat{g}_T(\hat{\theta}_a)$. A similar argument for obtaining $\dot{\Omega}_{\lambda\theta} \xrightarrow{p} \Omega$ implies that the derivative of (A.1) with respect to $\lambda(\hat{\theta}_a)$ converges in probability to a nonsingular matrix, $-\kappa_1\Omega$. Thus, applications of the implicit function theorem and the envelope theorem to (A.1) imply $T^{-1}\sum_{t=1}^{T}\rho_1(\kappa\lambda(\hat{\theta}_a)' g_{tT}(\hat{\theta}_a)) G_{\beta tT}(\hat{\theta}_a)'\lambda(\hat{\theta}_a) = 0$ w.p.a.1. An expansion of (A.1) around $(\hat{\beta}(a), \lambda(\hat{\theta}_a)) = (\beta_0, 0)$ yields $0 = -\hat{g}_T(\theta_{a\beta_0}) + T^{-1}\sum_{t=1}^{T}[\rho_1(\kappa\bar{\lambda}' g_{tT}(\hat{\theta}_a)) G_{\beta tT}(\theta_{a\bar{\beta}})(\hat{\beta}(a) - \beta_0) + \kappa\rho_2(\kappa\bar{\lambda}' g_{tT}(\hat{\theta}_a)) g_{tT}(\theta_{a\bar{\beta}}) g_{tT}(\hat{\theta}_a)'\lambda(\hat{\theta}_a)]$ w.p.a.1, where $(\bar{\beta}, \bar{\lambda})$ is a point on the line joining $(\beta_0, 0)$ and $(\hat{\beta}(a), \lambda(\hat{\theta}_a))$. By stacking these equations and using (A.2),

$$\begin{aligned} 0 &= \begin{pmatrix} 0 \\ -\hat{g}_T(\theta_{a\beta_0}) \end{pmatrix} + \begin{pmatrix} 0 & A_{12} \\ A_{21} & A_{22} \end{pmatrix}\begin{pmatrix} \hat{\beta}(a) - \beta_0 \\ -\kappa_1^{-1}\dot{\Omega}_{\lambda\theta}^{-1}\hat{g}_T(\hat{\theta}_a) \end{pmatrix} \\ &\equiv \begin{pmatrix} 0 \\ -\hat{g}_T(\theta_{a\beta_0}) \end{pmatrix} + A\begin{pmatrix} \hat{\beta}(a) - \beta_0 \\ -\kappa_1^{-1}\dot{\Omega}_{\lambda\theta}^{-1}\hat{g}_T(\hat{\theta}_a) \end{pmatrix} \end{aligned} \tag{A.4}$$

https://doi.org/10.1017/S0266466606060257 Published online by Cambridge University Press

holds w.p.a.1, where $A_{12} \equiv T^{-1}\sum_{t=1}^{T}\rho_1(\kappa\lambda(\hat{\theta}_a)'g_{tT}(\hat{\theta}_a))G_{\beta tT}(\hat{\theta}_a)'$, $A_{21} \equiv T^{-1}\sum_{t=1}^{T}\rho_1(\kappa\bar{\lambda}'g_{tT}(\hat{\theta}_a))G_{\beta tT}(\theta_{a\bar{\beta}})$, and $A_{22} \equiv \kappa(S_T/T)\sum_{t=1}^{T}\rho_2(\kappa\bar{\lambda}'g_{tT}(\hat{\theta}_a))g_{tT}(\theta_{a\bar{\beta}})g_{tT}(\hat{\theta}_a)'$. Lemmas 1(ii) and 2(i) and (iii) and Assumption 1 imply that $A_{12} \xrightarrow{p} -\kappa_1 M_{2\beta}'$, $A_{21} \xrightarrow{p} -\kappa_1 M_{2\beta}$, and $A_{22} \xrightarrow{p} -\kappa_1\Omega$. Assumptions 1(iii) and 5(ii) imply that A is invertible w.p.a.1. Thus, solving (A.4) for $(\hat{\beta}(a) - \beta_0)$ yields $T^{1/2}(\hat{\beta}(a) - \beta_0) = -\kappa_1^{-1}HT^{1/2}\hat{g}_T(\theta_{a\beta_0}) + o_p(1)$, where $H \equiv (M_{2\beta}'\Omega^{-1}M_{2\beta})^{-1}M_{2\beta}'\Omega^{-1}$. An expansion of $\hat{g}_T(\hat{\theta}_a)$ around $\hat{\beta}(a) = \beta_0$ and Lemma 2(iii) yield

$$
\begin{aligned}
T^{1/2}\hat{g}_T(\hat{\theta}_a) &= T^{1/2}\hat{g}_T(\theta_{a\beta_0}) + \hat{G}_{\beta T}(\theta_{a\dot{\beta}})T^{1/2}(\hat{\beta}(a) - \beta_0) \\
&= (I_m - M_{2\beta}H)T^{1/2}\hat{g}_T(\theta_{a\beta_0}) + o_p(1), \qquad \textbf{(A.5)}
\end{aligned}
$$

where $\dot{\beta}$ is a point on the line joining $\hat{\beta}(a)$ and β_0. From Lemma 1(iii) and Assumption 1, $T^{1/2}\hat{g}_T(\theta_{a\beta_0}) \xrightarrow{d} \kappa_1\xi \sim N(\kappa_1 m_1(\theta_{a\beta_0}), \kappa_1^2\Omega)$. Therefore, from (A.3) and (A.5), we have $2\kappa_1^{-2}\kappa_2 TS_T^{-1}(\hat{S}_{GEL}(a) - \rho(0)) \xrightarrow{d} \xi'(I_m - M_{2\beta}H)'\Omega^{-1}(I_m - M_{2\beta}H)\xi = \xi'\Omega^{-1}\Delta(\theta_{a\beta_0})\xi$. ■

Proof of Theorem 2. We follow the argument in the proof of Theorem 6 in Guggenberger and Smith (2005a). Let $D_{\rho T}^*(\hat{\theta}_a) = (D_{\rho\alpha T}^*(\hat{\theta}_a), D_{\rho\beta T}^*(\hat{\theta}_a))$. First, we show that $T^{1/2}\hat{g}_T(\hat{\theta}_a)$ and $D_{\rho\alpha T}^*(\hat{\theta}_a)$ are asymptotically independent. Let $M_{2\alpha_2} \equiv \partial m_2(\alpha_{02}, \beta_0)/\partial\alpha_2'$. An expansion of $D_{\rho\alpha T}^*(\hat{\theta}_a) = (D_{\rho\alpha_1 T}^*(\hat{\theta}_a), D_{\rho\alpha_2 T}^*(\hat{\theta}_a))$ around $\lambda(\hat{\theta}_a)'g_{tT}(\hat{\theta}_a) = 0$, Lemmas 1(ii) and 2(i), and (A.2) yield

$$
\begin{aligned}
D_{\rho\alpha_1 T}^*(\hat{\theta}_a) &= -T^{-1/2}\sum_{t=1}^{T} G_{\alpha_1 tT}(\hat{\theta}_a) \\
&\quad + \kappa_2^{-1}\frac{S_T}{T}\sum_{t=1}^{T} G_{\alpha_1 tT}(\hat{\theta}_a)g_{tT}(\hat{\theta}_a)'\Omega^{-1}T^{1/2}\hat{g}_T(\hat{\theta}_a) + o_p(1), \\
D_{\rho\alpha_2 T}^*(\hat{\theta}_a) &= -T^{-1}\sum_{t=1}^{T} G_{\alpha_2 tT}(\hat{\theta}_a) \\
&\quad + \kappa_2^{-1}\frac{S_T}{T}\sum_{t=1}^{T} G_{\alpha_2 tT}(\hat{\theta}_a)g_{tT}(\hat{\theta}_a)'\Omega^{-1}\hat{g}_T(\hat{\theta}_a) + o_p(1) = -\kappa_1 M_{2\alpha_2} + o_p(1),
\end{aligned}
$$

where the second equality in the second equation follows from Lemmas 2(iii) and 3(i) and $\sup_{\beta\in\mathcal{B},1\le t\le T}\|g_{tT}(\theta_{a\beta})\| = O_p(T^{1/\eta})$. An expansion of $T^{1/2}\operatorname{vec}\hat{G}_{\alpha_1 T}(\hat{\theta}_a)$ around $\hat{\beta}(a) = \beta_0$ and $T^{1/2}(\hat{\beta}(a) - \beta_0) = -\kappa_1^{-1}HT^{1/2}\hat{g}_T(\theta_{a\beta_0}) + o_p(1)$ (see the proof of Theorem 1) yield

$$
\begin{aligned}
T^{1/2}\operatorname{vec}\hat{G}_{\alpha_1 T}(\hat{\theta}_a) &= T^{1/2}\operatorname{vec}\hat{G}_{\alpha_1 T}(\theta_{a\beta_0}) \\
&\quad - \kappa_1^{-1}(\partial\operatorname{vec}\hat{G}_{\alpha_1 T}(\dot{\theta})/\partial\beta')(HT^{1/2}\hat{g}_T(\theta_{a\beta_0}) + o_p(1)) \\
&= T^{1/2}\operatorname{vec}\hat{G}_{\alpha_1 T}(\theta_{a\beta_0}) + o_p(1), \qquad \textbf{(A.6)}
\end{aligned}
$$

https://doi.org/10.1017/S0266466606060257 Published online by Cambridge University Press

where $\dot{\theta}$ is a point on the line joining $\theta_{a\beta_0}$ and $\hat{\theta}_a$ and the second equality follows from Lemma 2(iii) and $\sup_{\beta\in\mathcal{N}}\|\partial \operatorname{vec} \hat{G}_{\alpha_1 T}(\theta_{a\beta})/\partial\beta' - T^{-1/2}\partial(\operatorname{vec}(\partial m_{1T}(\theta_{a\beta})/\partial\alpha_1'))/\partial\beta'\| \xrightarrow{p} 0$ (by a similar argument as Lemma 3(i) using Assumption 6(ii)). From Lemma 3(ii), (A.5), and (A.6), stacking $\operatorname{vec} D^*_{\rho\alpha T}(\hat{\theta}_a)$ and $T^{1/2}\hat{g}_T(\hat{\theta}_a)$ yields

$$\operatorname{vec}(D^*_{\rho\alpha T}(\hat{\theta}_a), T^{1/2}\hat{g}_T(\hat{\theta}_a)) = \psi + \Psi\mathcal{G} + o_p(1), \tag{A.7}$$

where $\Psi \equiv \begin{pmatrix} -I_{mp_{\alpha_1}} & \Omega_{\alpha_1 g}(\theta_{a\beta_0})\Omega^{-1}0 \\ 0 & 0 \\ 0 & I_m \end{pmatrix}\begin{pmatrix} I_{mp_{\alpha_1}} & 0 \\ 0 & I_m - M_{2\beta}H \end{pmatrix}$, $\mathcal{G} \equiv \begin{pmatrix} T^{1/2}\operatorname{vec}\hat{G}_{\alpha_1 T}(\theta_{a\beta_0}) \\ T^{1/2}\hat{g}_T(\theta_{a\beta_0}) \end{pmatrix}$, and $\psi \equiv \operatorname{vec}(0, -\kappa_1 M_{2\alpha_2}, 0)$. Thus, from Lemma 3(iii), $\operatorname{vec}(D^*_{\rho\alpha T}(\hat{\theta}_a), T^{1/2}\hat{g}_T(\hat{\theta}_a))$ is asymptotically normal with the asymptotic covariance matrix $\kappa_1^2\Psi V(\theta_{a\beta_0})\Psi'$. Because the asymptotic covariance of $T^{1/2}\hat{g}_T(\hat{\theta}_a)$ and $\operatorname{vec}(D^*_{\rho\alpha T}(\hat{\theta}_a))$ (i.e., the off-diagonal component of $\kappa_1^2\Psi V(\theta_{a\beta_0})\Psi'$) is 0, $D^*_{\rho\alpha T}(\hat{\theta}_a)$ and $T^{1/2}\hat{g}_T(\hat{\theta}_a)$ are asymptotically independent.

Next, we derive the asymptotic distribution of $\hat{K}_{GEL}(a)$. Lemma 4(ii) implies that the FOC for $\gamma(\hat{\theta}_a)$,

$$\frac{S_T}{T}\sum_{t=1}^{T}\rho_1(\kappa'\gamma(\hat{\theta}_a)'D^*_{\rho T}(\hat{\theta}_a)'\hat{\Omega}_T(\hat{\theta}_a)^{-1}g_{tT}(\hat{\theta}_a))D^*_{\rho T}(\hat{\theta}_a)'\hat{\Omega}_T(\hat{\theta}_a)^{-1}g_{tT}(\hat{\theta}_a) = 0, \tag{A.8}$$

holds w.p.a.1. An expansion of (A.8) around $\gamma(\hat{\theta}_a) = 0$ yields $0 = -S_T D^*_{\rho T}(\hat{\theta}_a)'\hat{\Omega}_T(\hat{\theta}_a)^{-1}\hat{g}_T(\hat{\theta}_a) - \kappa_1\dot{\Phi}(\hat{\theta}_a)\gamma(\hat{\theta}_a)$ w.p.a.1, where $\dot{\Phi}(\hat{\theta}_a) \equiv D^*_{\rho T}(\hat{\theta}_a)'\hat{\Omega}_T(\hat{\theta}_a)^{-1}\dot{\Omega}_T(\hat{\theta}_a)\hat{\Omega}_T(\hat{\theta}_a)^{-1}D^*_{\rho T}(\hat{\theta}_a)$, $\dot{\Omega}_T(\hat{\theta}_a) \equiv -\kappa_2^{-1}(S_T/T)\sum_{t=1}^{T}\rho_2(\kappa\dot{\gamma}'D^*_{\rho T}(\hat{\theta}_a)'\hat{\Omega}_T(\hat{\theta}_a)^{-1}g_{tT}(\hat{\theta}_a))g_{tT}(\hat{\theta}_a)g_{tT}(\hat{\theta}_a)'$ and $\dot{\gamma}$ is a point on the line joining 0 and $\gamma(\hat{\theta}_a)$. By combining an argument similar to the proof of Theorem 6 in Guggenberger and Smith (2005a), Assumptions 1 and 6(i) and (ii), Lemma 3, and (A.7), we can show that $D^*_{\rho T}(\hat{\theta}_a)$ converges to a full column rank random matrix a.s. Thus, from $\hat{\Omega}_T(\hat{\theta}_a), \dot{\Omega}_T(\hat{\theta}_a) \xrightarrow{p} \Omega$ (by Assumption 6(iii) and Lemmas 1(ii) and 4(i)), $\dot{\Phi}(\hat{\theta}_a)$ is invertible w.p.a.1 and $\gamma(\hat{\theta}_a) = -\kappa_1^{-1}S_T\dot{\Phi}(\hat{\theta}_a)^{-1}D^*_{\rho T}(\hat{\theta}_a)'\hat{\Omega}_T(\hat{\theta}_a)^{-1}\hat{g}_T(\hat{\theta}_a)$ holds w.p.a.1. A second-order Taylor expansion of $\hat{P}_K(\hat{\theta}_a, \gamma(\hat{\theta}_a))$ around $\gamma(\hat{\theta}_a) = 0$ yields

$$\begin{aligned}
&2\kappa_2 TS_T^{-1}(\hat{K}_{GEL}(a) - \rho(0)) \\
&\quad = 2T\hat{g}_T(\hat{\theta}_a)'\hat{\Omega}_T(\hat{\theta}_a)^{-1}D^*_{\rho T}(\hat{\theta}_a)\dot{\Phi}(\hat{\theta}_a)^{-1}D^*_{\rho T}(\hat{\theta}_a)'\hat{\Omega}_T(\hat{\theta}_a)^{-1}\hat{g}_T(\hat{\theta}_a) \\
&\qquad - T\hat{g}_T(\hat{\theta}_a)'\hat{\Omega}_T(\hat{\theta}_a)^{-1}D^*_{\rho T}(\hat{\theta}_a)\dot{\Phi}(\hat{\theta}_a)^{-1}\ddot{\Phi}(\hat{\theta}_a)\dot{\Phi}(\hat{\theta}_a)^{-1}D^*_{\rho T}(\hat{\theta}_a)'\hat{\Omega}_T(\hat{\theta}_a)^{-1}\hat{g}_T(\hat{\theta}_a) \\
&\quad = T\hat{g}_T(\hat{\theta}_a)'\Omega^{-1}D^*_{\rho T}(\hat{\theta}_a)[D^*_{\rho T}(\hat{\theta}_a)'\Omega^{-1}D^*_{\rho T}(\hat{\theta}_a)]^{-1}D^*_{\rho T}(\hat{\theta}_a)'\Omega^{-1}\hat{g}_T(\hat{\theta}_a) + o_p(1) \\
&\quad = T(\hat{g}_T(\hat{\theta}_a)'\Omega^{-1}D^*_{\rho\alpha T}(\hat{\theta}_a), o_p(1))[D^*_{\rho T}(\hat{\theta}_a)'\Omega^{-1}D^*_{\rho T}(\hat{\theta}_a)]^{-1} \\
&\qquad \times \begin{pmatrix} D^*_{\rho\alpha T}(\hat{\theta}_a)'\Omega^{-1}\hat{g}_T(\hat{\theta}_a) \\ o_p(1) \end{pmatrix} + o_p(1),
\end{aligned}$$

where $\ddot{\Phi}(\hat{\theta}_a)$ is defined similarly to $\dot{\Phi}(\hat{\theta}_a)$, the second equality follows from Assumption 6(iii) and Lemmas 1(ii) and 4(i), and the third equality follows from the FOC for $\hat{\beta}(a)$ (i.e., $0 = D^*_{\rho\beta T}(\hat{\theta}_a)'\lambda(\hat{\theta}_a)$) and (A.2). From (A.7), we can denote $D^*_{\rho\alpha T}(\hat{\theta}_a) \xrightarrow{d} \kappa_1\bar{D}_\alpha$ and $T^{1/2}\hat{g}_T(\hat{\theta}_a) \xrightarrow{d} \kappa_1\bar{g}$. By the inversion formula of partitioned matrices, the $p_\alpha \times p_\alpha$ upper left submatrix of the limit of $[D^*_{\rho T}(\hat{\theta}_a)'\Omega^{-1}D^*_{\rho T}(\hat{\theta}_a)]^{-1}$ is

https://doi.org/10.1017/S0266466606060257 Published online by Cambridge University Press

$[\bar{D}_\alpha'^{-1}\Omega^{-1}\Delta(\theta_{a\beta_0})\bar{D}_\alpha]^{-1}$. Therefore, the continuous mapping theorem and asymptotic independence of $D^*_{\rho\alpha T}(\hat{\theta}_a)$ and $T^{1/2}\hat{g}_T(\hat{\theta}_a)$ imply that

$$2\left(\frac{\kappa_2}{\kappa_1^2}\right)TS_T^{-1}(\hat{K}_{GEL}(a)-\rho(0)) \xrightarrow{d} \bar{g}'\Omega^{-1}\bar{D}_\alpha[\bar{D}_\alpha'^{-1}\Omega^{-1}\Delta(\theta_{a\beta_0})\bar{D}_\alpha]^{-1}\bar{D}_\alpha'\Omega^{-1}\bar{g}$$

$$= (\Upsilon_K(a)+\xi_\alpha)'(\Upsilon_K(a)+\xi_\alpha), \tag{A.9}$$

where $\Upsilon_K(a) \equiv [\bar{D}_\alpha'^{-1}\Omega^{-1}\Delta(\theta_{a\beta_0})\bar{D}_\alpha]^{-1/2}\bar{D}_\alpha\Omega^{-1}\Delta(\theta_{a\beta_0})m_1(\theta_{a\beta_0})$ and $\xi_\alpha \sim N(0, I_{p_\alpha})$. Also, from Assumption 1(ii), we have $\Upsilon_K(\alpha_0) = 0$. ■

https://doi.org/10.1017/S0266466606060257 Published online by Cambridge University Press