Using implied probabilities to improve the estimation of unconditional moment restrictions for weakly dependent data: Supplementary material[1]

Alain Guay

University of Quebec at Montreal (UQAM)

CIRPÉE, CIREQ and L.E.A.D.

Florian Pelgrin

EDHEC Business School, Lille

and

Faculty of Business and Economics (HEC)

University of Lausanne

This version: August 2013

[1]Florian Pelgrin gratefully acknowledges financial support from the National Center of Competence in Research "Financial Valuation and Risk Management". The National Centers of Competence in Research (NCCR) are a research instrument of the Swiss National Science Foundation. Alain Guay gratefully acknowledges financial support from le Fonds québécois de la recherche sur la société et la culture (FQRSC). All proofs and extensions are available on the following link, http://www.er.uqam.ca/nobel/r27460/.

Email addresses : Florian Pelgrin, florian.pelgrin@edhec.edu and Alain Guay: guay.alain@uqam.ca

Appendix 1: Assumptions and proofs of the main results

Assumptions

Let $g_t = g(z_t, \theta^0)$ and $g_t^* = g(z_t, \theta^*)$ denote, respectively, the moments conditions evaluated at θ^0 and θ^*. Let $g_{\theta,t}$ and $g_{\theta,t}^*$ denote the corresponding first-order partial derivative of g with respect to the parameters of interest. To prove the higher-order properties, we adopt the same assumptions as in Anatolyev (2005) with the exception that the uniform kernel proposed by Kitamura and Stutzer (1997) is used for the reasons aforementioned in the text.

Assumptions A

A1 The sequence z_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 $\nu > 1$.

A2 The moment conditions (1) holds for unique $\theta^0 \in int(\Theta)$, where $\Theta \subseteq R^p$ is compact.

A3 The function $g(z_t, \theta^*)$ is Borel measurable for all $\theta^* \in \Theta$ and is twice continuously differentiable in θ^* for all $\theta^* \in \Theta$ and for z_t in its support.

A4 For some stationary series d_t with finite $E(d_t^8)$, $\sup_{\theta^* \in \Theta} \max\{\| g_t^* \|, \| g_{\theta,t}^* \|, \| \partial g_{\theta,t}^*/\partial\theta_j \|, \| \partial^2 g_{\theta,t}^*/\partial\theta_j\partial\theta' \| \; \forall j = 1, \cdots, p\} \leq d_t$ and $\max\{\| g_t^* - g_t \|, \| g_{\theta,t}^* - g_{\theta,t} \|, \| \partial g_{\theta,t}^*/\partial\theta_j - \partial g_{\theta,t}/\partial\theta_j \| \; \forall j = 1, \cdots, p\} \leq d_t \| \theta^* - \theta \|$ for all $\theta^* \in \Theta$.

A5 The matrices $G = E(g_{\theta,t})$ and $\Omega = \sum_{s=-\infty}^{\infty} E(g_t g_{t-s})$ are of full rank.

A6 $K_T \to \infty$ as $T \to \infty$ and $K_T = o(T^{1/3})$.

A7 $\rho(\cdot)$ is concave and three times continuously differentiable on its domain; an open interval Φ containing zero, has bounded Lipschitz third derivative in a neighborhood of zero and $\rho_1 = \rho_2 = -1$.

Proof of Proposition 1

To show this result, we first follow Smith (2011) and start from the definition of the smoothed generalized empirical likelihood estimator and the definition of the corresponding smoothed implied probabilities. The smoothed generalized empirical likelihood (SGEL) estimator is a solution to the following saddle point problem (Smith, 2011):

$$\hat{\theta}_T^{SGEL} = \arg\min_{\theta \in \Theta} \sup_{\lambda \in \hat{\Lambda}_T(\theta)} \sum_{t=1}^{T} \frac{[\rho(k\lambda' g_{tT}(\theta)) - \rho_0]}{T}$$

where $k = \frac{k_1}{k_2}$, $\rho_j() = \partial^j \rho()/\partial \phi^j$ and $\rho_j = \rho_j(0)$ for $j = 0, 1, 2, \ldots$, and $\rho_1 = \rho_2 = -1$ (normalization assumption). Following Smith (2011), the associated empirical (implied) probabilities are given by:

$$\pi_t^{SGEL}(\hat{\theta}_T^{SGEL}) = \frac{\rho_1(k\hat{\lambda}_T' g_{tT}(\hat{\theta}_T^{SGEL}))}{\sum_{t=1}^T \rho_1(k\hat{\lambda}_T' g_{tT}(\hat{\theta}_T^{SGEL}))} \tag{A.1}$$

for $t = 1, \ldots, T$ where $\hat{\lambda}_T = \arg\max_{\lambda \in \hat{\Lambda}(\theta)} \sum_{t=1}^T \frac{\left[\rho(k\lambda' g_{tT}(\hat{\theta}_T^{SGEL})) - \rho_0\right]}{T}$. Eq. (A.1) also holds true at any efficient first-order estimator $\hat{\theta}_T$ such that $\sqrt{T}\left(\hat{\theta}_T - \theta^0\right) = \mathcal{O}_p(1)$ and for the uniform truncated kernel (11):

$$\pi_t^{SGEL}(\hat{\theta}_T) = \frac{\rho_1\left(\lambda_T(\hat{\theta}_T)' g_{tT}(\hat{\theta}_T)\right)}{\sum_{t=1}^T \rho_1\left(\lambda_T(\hat{\theta}_T)' g_{tT}(\hat{\theta}_T)\right)}$$

where $\lambda_T(\hat{\theta}_T) = \arg\max_{\lambda \in \hat{\Lambda}(\theta)} \sum_{t=1}^T \frac{\left[\rho(\lambda' g_{tT}(\hat{\theta}_T)) - \rho_0\right]}{T}$ for $t = 1, \ldots, T$. Using a Taylor expansion $\lambda_T(\hat{\theta}_T)$ around 0 yields (uniformly in $t = 1, \ldots, T$):

$$\pi_t^{SGEL}(\hat{\theta}_T) \quad = \quad \frac{1}{T} + \frac{1}{T}\lambda_T(\hat{\theta}_T)' g_{tT}(\hat{\theta}_T) - \frac{1}{T^2}\sum_{t=1}^T \lambda_T(\hat{\theta}_T)' g_{tT}(\hat{\theta}_T) + R_1(\hat{\theta}_T) \tag{A.2}$$

where the remainder term $\|R_1(\hat{\theta}_T)\| = \mathcal{O}_p\left(\|\lambda_T(\hat{\theta}_T)' g_{tT}(\hat{\theta}_T)\|^2/T\right)$. Under weaker assumptions than **Assumptions A**, Smith (2011, Theorem 2.2) shows that (i) $g_{tT}(\hat{\theta}_T) = \mathcal{O}_p\left((2K_T+1)^{-1/2}\right)$ for $t = 1, \ldots, T$ and (ii) $\|\lambda_T(\hat{\theta}_T)\| = \mathcal{O}_p\left[\left((T/(2K_T+1)^2)^{-1/2}\right]\right)$ for any efficient first-order estimator.[1] This yields that $\|R_1(\hat{\theta}_T)\| = \mathcal{O}_p\left(\|\lambda_T(\hat{\theta}_T)' g_{tT}(\hat{\theta}_T)\|^2/T\right) = \frac{1}{T}\mathcal{O}_p\left(\|\lambda_T(\hat{\theta}_T)\|^2 \|g_{tT}(\hat{\theta}_T)\|^2\right) = \mathcal{O}_p\left(\frac{2K_T+1}{T^2}\right)$.
Now, using the FOC with respect to the vector of Lagrange multipliers and a Taylor expansion around 0 leads to:

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^T \rho_1\left(\lambda_T(\hat{\theta}_T)' g_{tT}(\hat{\theta}_T)\right) g_{tT}(\hat{\theta}_T) \quad &= \quad -\frac{1}{T}\sum_{t=1}^T g_{tT}(\hat{\theta}_T) - \frac{1}{T}\sum_{t=1}^T g_{tT}(\hat{\theta}_T) g_{tT}(\hat{\theta}_T)' \lambda_T(\hat{\theta}_T) \\ &+ \quad \frac{1}{T}\sum_{t=1}^T g_{tT}(\hat{\theta}_T) \sum_{j=2}^{\infty} \frac{1}{j!}\rho_{j+1}(0)\left(\lambda_T(\hat{\theta}_T)' g_{tT}(\hat{\theta}_T)\right)^j.\end{aligned}$$

Therefore

$$\frac{1}{T}\sum_{t=1}^T \rho_1\left(\lambda_T(\hat{\theta}_T)' g_{tT}(\hat{\theta}_T)\right) g_{tT}(\hat{\theta}_T) \quad = \quad -\frac{1}{T}\sum_{t=1}^T g_{tT}(\hat{\theta}_T) - \frac{1}{T}\sum_{t=1}^T g_{tT}(\hat{\theta}_T) g_{tT}(\hat{\theta}_T)' \lambda_T(\hat{\theta}_T) + R_2(\hat{\theta}_T)$$

[1] See also Kitamura and Stutzer (1997).

where $\|R_2(\hat{\theta}_T)\| = \mathcal{O}_p\left((2K+1)^{-3/2}\|\lambda_T(\hat{\theta}_T)\|^2\right) = \mathcal{O}_p\left((2K+1)^{1/2}/T\right)$. By using a consistent estimator $\Omega_T(\hat{\theta}_T) = \frac{2K+1}{T}\sum_{t=1}^{T} g_{tT}(\hat{\theta}_T)g_{tT}(\hat{\theta}_T)'$, one obtains:

$$0 = \frac{1}{T}\sum_{t=1}^{T} g_{tT}(\hat{\theta}_T) + \frac{1}{2K+1}\Omega_T(\hat{\theta}_T)\lambda_T(\hat{\theta}_T) + R_2(\hat{\theta}_T)$$

and thus

$$\lambda_T(\hat{\theta}_T) \quad = \quad -(2K_T+1)\Omega_T(\hat{\theta}_T)^{-1}\overline{g}_T(\hat{\theta}_T) + \mathcal{O}_p\left((2K_T+1)^{3/2}/T\right). \tag{A.3}$$

Replacing the expression (A.3) in Eq. (A.2) evaluated at $\hat{\theta}_T$ yields the following SGEL implied probabilities uniformly over $t = 1, \ldots, T$:

$$\begin{aligned}\pi_t^{SGEL}(\hat{\theta}_T) \quad &= \quad \frac{1}{T} - \frac{1}{T}\left[(2K_T+1)\left[g_{tT}(\hat{\theta}_T) - \overline{g}_T(\hat{\theta}_T)\right]'\Omega_T(\hat{\theta}_T)^{-1}\overline{g}_T(\hat{\theta}_T)\right] \\ &+ \quad \frac{1}{T}\left[g_{tT}(\hat{\theta}_T) - \overline{g}_T(\hat{\theta}_T)\right]'\mathcal{O}_p\left((2K_T+1)^{3/2}/T\right) + \mathcal{O}_p\left(\frac{2K_T+1}{T^2}\right).\end{aligned}$$

We obtain:

$$\pi_t^{SGEL}(\hat{\theta}_T) = \frac{1}{T} - \frac{1}{T}(2K_T+1)\left[g_{tT}(\hat{\theta}_T) - \overline{g}_T(\hat{\theta}_T)\right]'\Omega_T(\hat{\theta}_T)^{-1}\overline{g}_T(\hat{\theta}_T) + \mathcal{O}_p\left((2K_T+1)/T^2\right) \tag{A.4}$$

uniformly over $t = 1, \ldots, T$.

On the other hand, the closed-form expression of the SEEL implied probabilities evaluated at $\hat{\theta}_T$ (uniformly for $t = 1, \ldots, T$) using the uncentered estimator of the long-run covariance matrix of the moment conditions is given by:

$$\pi_t^{SEEL}(\hat{\theta}_T) = \frac{1}{T} - \frac{1}{T}(2K_T+1)g_{tT}(\hat{\theta}_T)'\Omega_T(\hat{\theta}_T)^{-1}\overline{g}_T(\hat{\theta}_T) \tag{A.5}$$

(see Antoine, Bonnal and Renault 2007 in an i.i.d. context).

The expression $\frac{1}{T}(2K_T+1)\overline{g}_T(\hat{\theta}_T)'\hat{\Omega}_T^{-1}\overline{g}_T(\hat{\theta}_T)$ in Eq. (A.4) is $\mathcal{O}_p\left((2K_T+1)/T^2\right)$ since $\overline{g}_T(\hat{\theta}_T)$ is $\mathcal{O}_p\left(T^{-1/2}\right)$ (see Smith 2011, Lemma A.7). Finally, putting together Eq. (A.5) and Eq. (A.4) for the smoothed empirical likelihood (SEL) yields:

$$\pi_t^{SEL}(\hat{\theta}_T) = \pi_t^{SEEL}(\hat{\theta}_T) + \mathcal{O}_p\left((2K_T+1)/T^2\right)$$

uniformly over $t = 1, \ldots, T$.

Proof of Proposition 2

The proof is based on Theorem 1 in Robinson (1988) which allows to evaluate the order of magnitude for the stochastic difference between two alternative estimators. The sketch of the proof for the smoothed 3SW-EEL is adapted from the one in Antoine et al. (2007) but with smoothed moment conditions. The proof for the smoothed 3S-EEL estimator follows.

Under **Assumptions A**, the p equations corresponding to the FOC of the SEL are given by:

$$f_T(\hat{\theta}_T^{SEL}) = \left[\sum_{t=1}^{T} \pi_t^{SEL}\left(\hat{\theta}_T^{SEL}\right) G_{tT}\left(\hat{\theta}_T^{SEL}\right)\right]' \left[\widetilde{\Omega}_T^{SEL}(\hat{\theta}_T^{SEL})\right]^{-1} \frac{1}{T}\sum_{t=1}^{T} g_{tT}(\hat{\theta}_T^{SEL}) = 0$$

where $\widetilde{\Omega}_T^{SEL}(\hat{\theta}_T^{SEL}) = (2K_T+1)\sum_{t=1}^{T} \pi_t^{SEL}\left(\hat{\theta}_T^{SEL}\right) g_{tT}(\hat{\theta}_T^{SEL}) g_{tT}(\hat{\theta}_T^{SEL})'$. The smoothed 3SW-EEL estimator is given by (using a second step efficient estimator $\hat{\theta}_T$):

$$h_T(\hat{\theta}_T^{S3SW}) = \left[\sum_{t=1}^{T} \pi_t^{SEEL}\left(\hat{\theta}_T^{S3SW}\right) G_{tT}\left(\hat{\theta}_T^{S3SW}\right)\right]' \left[\widetilde{\Omega}_T^{SEEL}(\hat{\theta}_T)\right]^{-1} \frac{1}{T}\sum_{t=1}^{T} g_{tT}(\hat{\theta}_T^{S3SW}) = 0$$

where $\widetilde{\Omega}_T^{SEEL}(\hat{\theta}_T) = (2K_T+1)\sum_{t=1}^{T} \pi_t^{SEEL}\left(\hat{\theta}_T\right) g_{tT}(\hat{\theta}_T) g_{tT}(\hat{\theta}_T)'$.

The objective is to show that $\hat{\theta}_T^{S3SW} - \hat{\theta}_T^{SEL} = \mathcal{O}_p\left((2K_T+1)/T^{3/2}\right)$. In doing so, we apply Theorem 1 in Robinson (1988). In that respect, two assumptions need to be fulfilled. Firstly, Assumption A1 in Robinson (1988) is directly verified since $\hat{\theta}_T^{SEL} = \theta_0 + o_p(1)$. Secondly, Assumption A2 in Robinson (1988) also holds since (i) $\hat{\theta}_T^{S3SW} = \theta_0 + o_p(1)$ and (ii) the derivative of $h_T(\theta)$ with respect to θ is uniformly continuous (for large T) with a probability arbitrarily close to one in the neighborhood of θ_0 by virtue of Assumption **A4** above. We also need to show that $H_T(\hat{\theta}_T^{S3SW}) = H(\theta_0) + o_p(1)$ where $H_T(\theta) = \frac{\partial h_T(\theta)}{\partial \theta}$ and $H(\theta_0) = G'\Omega^{-1}Eg_t(\theta_0) + G'\Omega^{-1}G$ where G and Ω are defined in Assumption **A5**. In this regard,

$$\begin{aligned}
H_T(\hat{\theta}_T^{S3SW}) &= \frac{\partial\left[\sum_{t=1}^{T} \pi_t^{SEEL}\left(\hat{\theta}_T^{S3SW}\right) G_{tT}\left(\hat{\theta}_T^{S3SW}\right)\right]'}{\partial \theta} \left[\widetilde{\Omega}_T^{SEEL}(\hat{\theta}_T)\right]^{-1} \frac{1}{T}\sum_{t=1}^{T} g_{tT}(\hat{\theta}_T^{S3SW}) \\
&+ \left[\sum_{t=1}^{T} \pi_t^{SEEL}\left(\hat{\theta}_T^{S3SW}\right) G_{tT}\left(\hat{\theta}_T^{S3SW}\right)\right]' \left[\widetilde{\Omega}_T^{SEEL}(\hat{\theta}_T)\right]^{-1} \frac{1}{T}\sum_{t=1}^{T} \frac{\partial g_{tT}(\hat{\theta}_T^{S3SW})}{\partial \theta'} \\
&= G'\Omega^{-1}Eg(\theta_0) + G'\Omega^{-1}G + o_p(1) = H(\theta_0) + o_p(1)
\end{aligned}$$

since $\sum_{t=1}^{T} \pi_t^{SEEL}\left(\hat{\theta}_T^{S3SW}\right) G_{tT}\left(\hat{\theta}_T^{S3SW}\right) = \frac{1}{T}\sum_{t=1}^{T} G_{tT}\left(\hat{\theta}_T^{S3SW}\right) + o_p(1) = \frac{1}{T}\sum_{t=1}^{T} G_t\left(\hat{\theta}_T^{S3SW}\right) + o_p(1) = G +$ $o_p(1)$ taking that $\pi_t^{SEEL}\left(\hat{\theta}_T^{S3SW}\right) = \frac{1}{T}\left(1 + o_p(1)\right)$ uniformly in t and $\frac{1}{T}\sum_{t=1}^{T} G_{tT}(\theta) = \frac{1}{T}\sum_{t=1}^{T} G_t(\theta) + o_p(1)$ (see Smith 2011). Moreover, $\widetilde{\Omega}_T^{SEEL}(\hat{\theta}_T) \xrightarrow{p} \Omega$ and $\sum_{t=1}^{T} g_{tT}(\hat{\theta}_T^{S3SW}) \xrightarrow{p} Eg_t(\theta_0)$.

Under these assumptions, Theorem 1 in Robinson (1988) implies that:

$$\hat{\theta}_T^{S3SW} - \hat{\theta}_T^{SEL} = \mathcal{O}_p\left(\|h_T\left(\hat{\theta}_T^{SEL}\right) - f_T\left(\hat{\theta}_T^{SEL}\right)\|\right)$$

where

$$h_T(\hat{\theta}_T^{SEL}) = \left[\sum_{t=1}^{T} \pi_t^{SEEL}\left(\hat{\theta}_T^{SEL}\right) G_{tT}\left(\hat{\theta}_T^{SEL}\right)\right]' \left[\widetilde{\Omega}_T^{SEEL}(\hat{\theta}_T)\right]^{-1} \frac{1}{T}\sum_{t=1}^{T} g_{tT}(\hat{\theta}_T^{SEL}) = 0.$$

Taking Therorem 3.1 in Smith (2011), the estimator $\sum_{t=1}^{T} \pi_t^{SEEL}\left(\hat{\theta}_T^{SEL}\right) G_{tT}\left(\hat{\theta}_T^{SEL}\right)$ is an estimator of $G = E\partial g(z_t, \theta_0)/\partial\theta'$ that efficiently incorporates the moment information (1) for any SGEL estimator. In particular, the conclusion is valid for the SEL and the smoothed CUE. This implies that (using Proposition 1):

$$\sum_{t=1}^{T}\left(\pi_t^{SEL}\left(\hat{\theta}_T^{SEL}\right) - \pi_t^{SEEL}\left(\hat{\theta}_T^{SEL}\right)\right) G_{tT}\left(\hat{\theta}_T^{SEL}\right) = \sum_{t=1}^{T} \mathcal{O}_p\left((2K_T+1)/T^2\right) G_{tT}\left(\hat{\theta}_T^{SEL}\right) = \mathcal{O}_p\left((2K_T+1)/T\right)$$

and thus

$$\sum_{t=1}^{T} \pi_t^{SEL}\left(\hat{\theta}_T^{SEL}\right) G_{tT}\left(\hat{\theta}_T^{SEL}\right) = \sum_{t=1}^{T} \pi_t^{SEEL}\left(\hat{\theta}_T^{SEL}\right) G_{tT}\left(\hat{\theta}_T^{SEL}\right) + o_p(1).$$

Consequently,

$$\begin{aligned}
&\hat{\theta}_T^{S3SW} - \hat{\theta}_T^{SEL} = \mathcal{O}_p\left(\|h_T\left(\hat{\theta}_T^{SEL}\right) - f_T\left(\hat{\theta}_T^{SEL}\right)\|\right) \\
&\leq \mathcal{O}_p\left(\left\|\left[\sum_{t=1}^{T} \pi_t^{SEL}\left(\hat{\theta}_T^{SEL}\right) G_{tT}\left(\hat{\theta}_T^{SEL}\right)\right]'\right\| \left\|\widetilde{\Omega}_T^{SEEL}(\hat{\theta}_T)^{-1} - \widetilde{\Omega}_T^{SEL}(\hat{\theta}_T^{SEL})^{-1}\right\| \left\|\frac{1}{T}\sum_{t=1}^{T} g_{tT}(\hat{\theta}_T^{SEL})\right\|\right).
\end{aligned}$$

For the first right-hand side term, $\sum_{t=1}^{T} \pi_t^{SEL}\left(\hat{\theta}_T^{SEL}\right) G_{tT}\left(\hat{\theta}_T^{SEL}\right) \xrightarrow{p} G$ and this term is $\mathcal{O}_p(1)$. The last term $\frac{1}{T}\sum_{t=1}^{T} g_{tT}(\hat{\theta}_T^{SEL})$ is shown to be $\mathcal{O}_p(1/\sqrt{T})$ by Smith (2011, Lemma A.7). Thus, to get the desired result, we only need to show that:

$$\left\|\widetilde{\Omega}^{SEEL}(\hat{\theta}_T)^{-1} - \widetilde{\Omega}^{SEL}(\hat{\theta}_T^{SEL})^{-1}\right\| = \mathcal{O}_p\left((2K_T+1)/T\right). \tag{A.6}$$

By virtue of the triangular inequality,

$$\left\|\widetilde{\Omega}^{SEEL}(\hat{\theta}_T) - \widetilde{\Omega}^{SEL}(\hat{\theta}_T^{SEL})\right\| \leq \left\|\widetilde{\Omega}^{SEEL}(\hat{\theta}_T) - \widetilde{\Omega}^{SEEL}(\hat{\theta}_T^{SEL})\right\| + \left\|\widetilde{\Omega}^{SEEL}(\hat{\theta}_T^{SEL}) - \widetilde{\Omega}^{SEL}(\hat{\theta}_T^{SEL})\right\|.$$

The first right-hand side expression is $\mathcal{O}_p(1/T)$ by an usual Taylor expansion using $\hat{\theta}_T - \hat{\theta}_T^{SEL} = \mathcal{O}_p(1/T)$ and the boundness Assumption **A4**. For the second expression, using Proposition 1, one has:

$$\left\|\widetilde{\Omega}^{SEEL}(\hat{\theta}_T^{SEL}) - \widetilde{\Omega}^{SEL}(\hat{\theta}_T^{SEL})\right\| = (2K_T+1)\sum_{t=1}^{T}\mathcal{O}_p\left((2K_T+1)/T^2\right) g_{tT}(\hat{\theta}_T^{SEL}) g_{tT}(\hat{\theta}_T^{SEL}) = \mathcal{O}_p\left((2K_T+1)/T\right).$$

This shows Eq. (A.6). The result follows by noticing that $M^{-1} - N^{-1} = M^{-1}(N-M)N^{-1}$.

What remains to be shown is the result for the smoothed 3S-EEL: $\hat{\theta}_T^{S3S} - \hat{\theta}_T^{SEL} = \mathcal{O}_p\left((2K_T+1)/T^{3/2}\right)$. As aforementioned, the estimator $\sum_{t=1}^{T}\pi_t^{SEL}\left(\hat{\theta}_T^{SEL}\right) G_{tT}\left(\hat{\theta}_T^{SEL}\right)$ is an estimator of $G = E\partial g(z_t,\theta_0)/\partial\theta'$ which efficiently incorporates the moment information (1) for any SGEL estimator. This also holds if the SEL estimator is replaced by any first order equivalent estimator (e.g., the 2-step GMM estimator). We get that:

$$\sum_{t=1}^{T}\pi_t^{SEL}\left(\hat{\theta}_T^{SEL}\right) G_{tT}\left(\hat{\theta}_T^{SEL}\right) = \sum_{t=1}^{T}\pi_t^{SEL}\left(\hat{\theta}_T\right) G_{tT}\left(\hat{\theta}_T\right) + o_p(1) = \sum_{t=1}^{T}\pi_t^{SEEL}\left(\hat{\theta}_T\right) G_{tT}\left(\hat{\theta}_T\right) + o_p(1)$$

by Smith (2011, Theorem 3.1) and Proposition 1. Using the derivation for Eq. (A.6), the result follows directly.

Appendix 2: Bias-corrected versions of the proposed estimators

Given that the smoothed 3S-EEL, 3SW-EEL and the SEL estimators have the same bias-order, namely $\mathcal{O}(T^{-1})$, the higher-order asymptotic derivations in Anatolyev (2005) allow us for proposing a bias-corrected version of these estimators. The next proposition gives the corresponding expression for the smoothed 3SW-EEL estimator. The same result applies for the smoothed 3S-EEL estimator.

Proposition 1 *Under* **Assumptions A**, *a consistent estimator of the asymptotic bias of order T^{-1} is given by:*

$$B\hat{i}as(\hat{\theta}_T^{S3SW}) = \hat{B}_{G\Xi g}/T + \hat{B}_{\partial^2 g}/T$$

where $\hat{B}_{G\Xi g}$ *and* $\hat{B}_{\partial^2 g}$ *are consistent estimators of:*

$$B_{G\Xi g} = \Xi \sum_{u=-\infty}^{\infty} E\left[g_{\theta,t}\Xi g_{t-u}\right]$$

$$B_{\partial^2 g} = \Xi \sum_{j=1}^{p} E\left[\frac{\partial g_{\theta,t}}{\partial \theta_j}\frac{\Sigma}{2}e_j\right]$$

and e_j *is the* j*th column of the identity matrix of order* p*,* $\Sigma = \left(G'\Omega^{-1}G\right)^{-1}$ *and* $\Xi = \Sigma G'\Omega^{-1}$*. The bias corrected smoothed three-step estimators* $\hat{\theta}_T^{S3SWc}$ *defined as* $\hat{\theta}_T^{S3SWc} = \hat{\theta}_T^{S3SW} - \hat{Bias}(\hat{\theta}_T^{S3SW})$ *are asymptotically unbiased up to order* T^{-1}*.*

Proof: Theorem 1 in Anatolyev (2005) provides the asymptotic bias of the SEL estimator. Taking Proposition 2, both smoothed three-step EEL estimators are asymptotically higher-order equivalent, i.e. the asymptotic bias of each estimator is the same as the one for the SEL estimator up to an order $\mathcal{O}_p\left((2K_T+1)/T^{3/2}\right)$. The first term appearing in the asymptotic bias of the SEL estimator (Theorem 1 in Anatolyev (2005)) is removed by the use of the uniform kernel proposed by Kitamura and Stutzer (1997). The asymptotic bias at order T^{-1} of the 3SW-EEL estimator is then given by: $B_{G\Xi g} + B_{\partial^2 g}$. Finally, both terms are removed by the correction.

It is worth noting that the two terms, $B_{G\Xi g}$ and $B_{\partial^2 g}$, correspond to the asymptotic bias of the infeasible GMM estimator using the optimal linear combination of the moment conditions. Consistent estimators of these two terms are then obtained by replacing the moment conditions or their derivatives with their respective smoothed counterparts (see Lemma 2 and Lemma 3 in Anatolyev, 2005). Hence, $\widetilde{G}_T = \sum_{t=1}^{T}\pi_t^{SEEL}(\hat{\theta}_T^{S3SW})G_{tT}(\hat{\theta}_T^{S3SW})$, $\widetilde{\Omega}_T = (2K_T+1)\sum_{t=1}^{T}\pi_t^{SEEL}(\hat{\theta}_T^{S3SW})g_{tT}(\hat{\theta}_T^{S3SW})g_{tT}(\hat{\theta}_T^{S3SW})'$ and a consistent estimator of $\sum_{u=-\infty}^{\infty}E\left[g_{\theta,t}\Xi g_{t-u}\right]$ is given by:

$$(2K_T+1)\sum_{t=1}^{T}\pi_t^{SEEL}(\hat{\theta}_T^{S3SW})G_{tT}(\hat{\theta}_T^{S3SW})\widetilde{\Xi}_T g_{tT}(\hat{\theta}_T^{S3SW})$$

where $\widetilde{\Xi}_T = \left(\widetilde{G}_T'\widetilde{\Omega}_T^{-1}\widetilde{G}_T\right)^{-1}\widetilde{G}_T'\widetilde{\Omega}_T^{-1}$ (see Lemma 3b in Anatolyev (2005)). Finally, a consistent estimator of the second bias term is obtained with $\widetilde{\Sigma}_T = \left(\widetilde{G}_T'\widetilde{\Omega}_T^{-1}\widetilde{G}_T\right)^{-1}$ and a consistent estimate of the second-order partial derivative of the smoothed moment conditions with respect to the parameter vector θ. As a final remark, note that that the bias terms may also be estimated with the uniform weights, $1/T$, instead of the constrained implied probabilities.

Appendix 3: The reduced-form and the concentration parameter

The full model of the endogenous variable, y_t, rests on the following system:

$$\begin{aligned} y_t &= \gamma_b y_{t-1} + \gamma_f E_t y_{t+1} + \lambda x_t + \epsilon_t \\ x_t &= \rho_1 x_{t-1} + \rho_2 x_{t-2} + v_t. \end{aligned}$$

Using the companion form of the second equation, we have:

$$\begin{aligned} y_t &= \gamma_b y_{t-1} + \gamma_f E_t y_{t+1} + \begin{pmatrix} \lambda & 0 \end{pmatrix} \begin{pmatrix} x_t \\ x_{t-1} \end{pmatrix} + \epsilon_t \\ \begin{pmatrix} x_t \\ x_{t-1} \end{pmatrix} &= \Lambda \begin{pmatrix} x_{t-1} \\ x_{t-2} \end{pmatrix} + \begin{pmatrix} v_t \\ 0 \end{pmatrix} \end{aligned}$$

where:

$$\Lambda = \begin{pmatrix} \rho_1 & \rho_2 \\ 1 & 0 \end{pmatrix}.$$

Determining the roots of the characteristic equation, we get:

$$y_t = \delta_1 y_{t-1} + \frac{1}{\delta_2 \gamma_f} \begin{pmatrix} \lambda & 0 \end{pmatrix} \left(I_2 - \delta_2^{-1}\Lambda\right)^{-1} \begin{pmatrix} x_t \\ x_{t-1} \end{pmatrix} + \frac{1}{\delta_2 \gamma_f}\epsilon_t$$

where:

$$\delta_1 = \frac{1 - \sqrt{1 - 4\gamma_f\gamma_b}}{2\gamma_f} \qquad \text{and} \qquad \delta_2 = \frac{1 + \sqrt{1 - 4\gamma_f\gamma_b}}{2\gamma_f}.$$

It is straightforward to show that the reduced-form is:

$$\begin{aligned} y_t &= \delta_1 y_{t-1} + \alpha_0 x_t + \alpha_1 x_{t-1} + \alpha_\epsilon \epsilon_t \\ x_t &= \rho_1 x_{t-1} + \rho_2 x_{t-2} + v_t \end{aligned}$$

where $\alpha_0 = \frac{\lambda}{\Delta \delta_2 \gamma_f}$, $\alpha_1 = \alpha_0 \frac{\rho_2}{\delta_2}$, $\alpha_\epsilon = \frac{1}{\delta_2 \gamma_f}$, $\alpha_\epsilon = \frac{1}{\delta_2 \gamma_f}$ and Δ is given by $\rho\left(\frac{1}{\delta_2}\right) \equiv 1 - \frac{\rho_1}{\delta_2} - \frac{\rho_1}{\delta_2^2}$.

Following Mavroeidis (2004), the first-stage regression for the endogenous regressors is given by:

$$Y_t = \Pi' Z_t + \Phi' X_t + V_t$$

where $Y_t = (y_{t+1}, x_t)'$, $Z_t = (x_{t-1}, x_{t-2})'$, $X_t = y_{t-1}$, $V_t = (\eta_{t+1}, v_t)'$ and:

$$\begin{aligned}
\Pi' &= \begin{pmatrix} \tilde{\alpha}_0 & \tilde{\alpha}_1 \\ \rho_1 & \rho_2 \end{pmatrix}, \Phi' = \begin{pmatrix} \delta_1^2 \\ 0 \end{pmatrix} \\
\eta_{t+1} &= \alpha_\epsilon \epsilon_{t+1} + \delta_1 \alpha_\epsilon \epsilon_t + \alpha_0 v_{t+1} + (\alpha_0 \rho_1 + \alpha_1 + \delta_1 \alpha_0) v_t
\end{aligned}$$

with $\tilde{\alpha_0} = \rho_1 (\alpha_0 \rho_1 + \alpha_1 + \delta_1 \alpha_0) + \alpha_0 \rho_2 + \delta_1 \alpha_1$, and $\tilde{\alpha}_1 = (\alpha_0 \rho_1 + \alpha_1 + \delta_1 \alpha_0) \rho_2$.

The concentration parameter is the minimum eigenvalue of the following matrix:

$$T \Sigma_{VV}^{-1/2} \Pi' \Omega \Pi \Sigma_{VV}^{-1/2}$$

where $\Omega = \Sigma_{ZZ} - \Sigma_{ZX} \Sigma_{XX}^{-1} \Sigma_{ZX}'$, Σ_{ZZ}, Σ_{XX}, and Σ_{ZX} are respectively the variance-covariance of Z_t and X_t, and the covariance matrix between Z_t and X_t, $\Sigma_{VV} = A \Sigma A' + B \Sigma B'$, with:

$$\begin{aligned}
\Sigma &= \begin{pmatrix} \sigma_\epsilon^2 & \sigma_{\epsilon v} \\ \sigma_{\epsilon v} & \sigma_v^2 \end{pmatrix} \\
A &= \begin{pmatrix} \delta_1 \alpha_\epsilon & \alpha_0 (\delta_1 + \rho_1) + \alpha_1 \\ 0 & 1 \end{pmatrix} \\
B &= \begin{pmatrix} \alpha_\epsilon & \alpha_0 \\ 0 & 0 \end{pmatrix}.
\end{aligned}$$

Supplementary material additional reference

Robinson, P.M., 1988, “The Stochastic Difference between Econometric Statistics”, *Econometrica* 56: 531-548.

.