The Empirical Saddlepoint Likelihood Estimator Applied to Two-Step GMM

Fallaw Sowell

Tepper School of Business
Carnegie Mellon University

May 2009
Overview

• New estimator empirical saddlepoint likelihood (ESPL)
• More accurate parameter estimates for models specified by moment conditions.
• Same asymptotic distribution as efficient two-step GMM.
• Smaller higher order bias than other (nonbias corrected) estimators (Newey and Smith (2004) and Schennach (2007)).
• New test for overidentifying restrictions.
• Better behavior is simulations.
How the ESPL exploits additional information

\[ \sqrt{n} \, G_n(\theta_0) \sim N \left( 0, \Sigma_g(\theta_0) \right) \]

NLS \hspace{1cm} \argmin_{\theta} \ G_n(\theta)' \ G_n(\theta) \hspace{1cm} \text{consistent}

GLS, GMM \hspace{1cm} \argmin_{\theta} \ G_n(\theta)' \hat{\Sigma}^{-1} \ G_n(\theta) \hspace{1cm} \text{consistent, efficient}

ML, ESPL \hspace{1cm} \argmax_{\theta} \left| \Sigma_g(\theta) \right|^{-1/2} \exp \left\{ -\frac{n}{2} \ G_n(\theta)' \Sigma_g^{-1}(\theta) \ G_n(\theta) \right\} \hspace{1cm} \text{consistent, efficient and smaller higher order bias}
Talk outline

- Two-Step GMM notation.
- Moment conditions to estimations equations (FOC’s).
- Introduce the EPSL estimator and new tests.
- Higher order bias intuition.
- Simulation comparison with other estimators and tests.
Two-step GMM

\[ E \left[ g(x_i, \theta_0) \right] = 0, \quad \theta \quad k \leq m, \quad g_i(\theta) \equiv g(x_i, \theta) \]

\[ M_0 = E \left[ \frac{\partial g_i(\theta_0)}{\partial \theta'} \right] \quad \hat{M}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial g_i(\theta)}{\partial \theta'} \]

\[ \sqrt{n} G_n(\theta_0) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_i(\theta_0) \sim a \ N \left( 0, \Sigma_g \right) \]

\[ \hat{\theta}_{gmm} \equiv \arg\min_{\theta} G_n(\theta)' \hat{\Sigma}_g^{-1} G_n(\theta) \]

\[ \hat{M}_n(\hat{\theta}_{gmm})' \hat{\Sigma}_g^{-1} G_n(\hat{\theta}_{gmm}) = 0 \]

\[ J = n G_n(\hat{\theta}_{gmm})' \hat{\Sigma}_g^{-1} G_n(\hat{\theta}_{gmm}) \sim \chi^2_{m-k} \]
Estimation Equations from moment conditions

The saddlepoint density needs a just identified system.

\[
\hat{M}_n(\hat{\theta}_{gmm})'\hat{\Sigma}_g^{-1}G_n(\hat{\theta}_{gmm}) = 0_{k \times 1}
\]

\[
\left(\hat{M}_n(\hat{\theta}_{gmm})'\hat{\Sigma}_g^{-1/2}\right)\left(\hat{\Sigma}_g^{-1/2}G_n(\hat{\theta}_{gmm})\right) = 0
\]

\[
C_1(\hat{\theta}_{gmm})'\left(\hat{\Sigma}_g^{-1/2}G_n(\hat{\theta}_{gmm})\right) = 0
\]

\(C_1(\hat{\theta}_{gmm})\) is an orthonormal basis for the span of \(\hat{\Sigma}_g^{-1/2}\hat{M}_n(\hat{\theta}_{gmm})\).

The other information in the moments is contained in

\[
C_2(\hat{\theta}_{gmm})'\left(\hat{\Sigma}_g^{-1/2}G_n(\hat{\theta}_{gmm})\right) = \hat{\lambda}_{(m-k) \times 1}
\]

\[
J = nG_n(\hat{\theta}_{gmm})'\hat{\Sigma}_g^{-1}G_n(\hat{\theta}_{gmm})
\]

\[
= nG_n(\hat{\theta}_{gmm})'\hat{\Sigma}_g^{-1/2}C_2(\hat{\theta}_{gmm})C_2(\hat{\theta}_{gmm})'\hat{\Sigma}_g^{-1/2}G_n(\hat{\theta}_{gmm})
\]

\[
= n\hat{\lambda}'\hat{\lambda} \sim \chi^2_{m-k}
\]
Just identified estimation equations

• Jointly estimate \( \theta \) and \( \lambda \) with estimation equations

\[
\Psi_n(\alpha) \equiv \hat{\Sigma}_g^{-1/2} G_n(\theta) - C_2(\theta) \lambda = 0_{m \times 1}
\]

where \( \alpha = [ \theta', \lambda' ]' \).
The empirical saddlepoint density

\[ \hat{f}_s(\alpha) = \left( \frac{n}{2\pi} \right)^{\frac{m}{2}} \left| \sum_{i=1}^{n} \exp \left\{ \tau_n' \psi_i(\alpha) \right\} \psi_i(\alpha) \psi_i(\alpha)' \right|^{\frac{1}{2}} \times \left| \sum_{i=1}^{n} \exp \left\{ \tau_n' \psi_i(\alpha) \right\} \frac{\partial \psi_i(\alpha)}{\partial \alpha'} \right|^{-1} \times \exp \left\{ \left( n - \frac{m}{2} \right) \ln \left( \frac{1}{n} \sum_{i=1}^{n} \exp \left\{ \tau_n' \psi_i(\alpha) \right\} \right) \right\} \]

where \( \tau_n \) solves

\[ \sum_{i=1}^{n} \psi_i(\alpha) \exp \left\{ \tau' \psi_i(\alpha) \right\} = 0 \]
The ESPL estimator

The ESPL estimator is the parameter value where $\hat{f}_s(\alpha)$ takes its maximum.
Alternatively where its natural log is maximized

$$\alpha_{espl} \equiv \arg\max_{\alpha} L_n(\alpha)$$

$$L_n(\alpha) = \frac{1}{2n} \ln \left( \left| \sum_{i=1}^{n} \exp \{ \tau'_n \psi_i(\alpha) \} \psi_i(\alpha) \psi_i(\alpha)' \right| \right)$$

$$- \frac{1}{n} \ln \left( \left| \sum_{i=1}^{n} \exp \{ \tau'_n \psi_i(\alpha) \} \frac{\partial \psi_i(\alpha)}{\partial \alpha'} \right| \right)$$

$$+ \left( 1 - \frac{m}{2n} \right) \ln \left( \frac{1}{n} \sum_{i=1}^{n} \exp \{ \tau'_n \psi_i(\alpha) \} \right)$$
Conditional ESPL (CESPL) estimator

- Recall $\alpha = [\theta', \lambda']'$ and $\lambda$ judges the overidentifying restrictions.
- Estimate $\theta$ conditional on $\lambda = 0$, i.e. the overidentifying restrictions being true.

$$\hat{\theta}_{cespl} \equiv \arg\max_{\theta} \mathcal{L}_n(\theta, \lambda = 0)$$
New tests for overidentifying restrictions

- **LR** \[2n \left( \mathcal{L}_n( \hat{\theta}_{espl}, \hat{\lambda}_{espl}, \hat{\tau}_{espl} ) - \mathcal{L}_n( \hat{\theta}_{cespl}(0), 0, \hat{\tau}_{espl} ) \right)\]
- **Wald** \[n \hat{\lambda}'_{espl} \hat{\lambda}_{espl}\]
- **Score** \[n \frac{\partial \mathcal{L}_n( \hat{\theta}_{cespl}(0), 0, \hat{\tau}_{cespl} )}{\partial \lambda'} \frac{\partial \mathcal{L}_n( \hat{\theta}_{cespl}(0), 0, \hat{\tau}_{cespl} )}{\partial \lambda}\]
- **Tilting** \[n \hat{\tau}( \hat{\theta}_{cespl}(0), 0)' \Sigma_g \hat{\tau}( \hat{\theta}_{cespl}(0), 0).\]
- **Jr** \[n \psi_n( \hat{\theta}_{cespl}(0), 0)' \Sigma_g \psi_n( \hat{\theta}_{cespl}(0), 0)\]

If moment conditions are true \(\chi^2_{m-k}\).

Imbens, Spady and Johnson (1998) test \(ET_r = n \hat{\tau}( \hat{\theta}_{et} )' \Sigma_g \hat{\tau}( \hat{\theta}_{et} ).\)
Higher Order Bias

The $O(n^{-1})$ part of the bias. The expectation of the first term ignored in the asymptotic distribution. (Newey and Smith (2004))

$$E(\hat{\theta}_{gmm} - \theta_0) = \frac{1}{n} (B_I + B_G + B_\Omega + B_W) + O(n^{-3/2})$$

$$E(\hat{\theta}_{el} - \theta_0) = \frac{1}{n} (B_I) + O(n^{-3/2})$$

$$E(\hat{\theta}_{espl} - \theta_0) = \frac{1}{n} (B_D + B_I) + O(n^{-3/2}).$$

For many models $B_D$ reduces to $-B_I$. 

Hall and Horowitz (1996) model

\[
g_i(\theta) = \begin{bmatrix}
\exp \{\mu - \theta (x_i + y_i) + 3y_i\} - 1 \\
y_i \left( \exp \{\mu - \theta (x_i + y_i) + 3y_i\} - 1 \right) \\
(z_{i3}^2 - 1) \left( \exp \{\mu - \theta (x_i + y_i) + 3y_i\} - 1 \right) \\
\vdots \\
(z_{im}^2 - 1) \left( \exp \{\mu - \theta (x_i + y_i) + 3y_i\} - 1 \right)
\end{bmatrix}
\]

where \( \theta_0 = 3 \),
\( x_i \) and \( y_i \) are iid from \( N(0, .16) \),
\( z_{ij} \) are iid \( N(0, 1) \) for \( j = 3, \ldots, m \).
\( \mu = -.72 \) is known.
Bias Comparison

<table>
<thead>
<tr>
<th>Model (m)</th>
<th>50</th>
<th>100</th>
<th>200</th>
<th>200</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.012</td>
<td>0.020</td>
<td>0.007</td>
<td>0.048</td>
<td>0.036</td>
</tr>
<tr>
<td>10</td>
<td>0.113</td>
<td>0.057</td>
<td>0.029</td>
<td>0.059</td>
<td>0.137</td>
</tr>
<tr>
<td>4</td>
<td>0.112</td>
<td>0.056</td>
<td>0.029</td>
<td>0.055</td>
<td>0.108</td>
</tr>
<tr>
<td>10</td>
<td>0.155</td>
<td>0.073</td>
<td>0.037</td>
<td>0.098</td>
<td>0.236</td>
</tr>
</tbody>
</table>

Table: The bias of the **ESPL**, **CESPL**, EL, ETEL and ET estimators for the Hall and Horowitz model. The sample size is denoted \(n\) and the number of moment conditions is \(m\).
Bias

Figure: Cumulative distribution functions for the ESPL, CESPL, EL, ETEL and ET estimators of $\theta$ in the Hall and Horowitz model. The sample size is denoted by $n$ and the number of moment conditions is $m$. These c.d.f.’s were calculated using 10000 simulated samples.
Tests of overidentifying restrictions

Figure: QQplots for tests of the overidentifying restrictions for the Hall and Horowitz model. The sample size is denoted by $n$ and the number of moment conditions is $m$. These were calculated using 10000 simulated samples. Vertical lines show the nominal .95 and .99 levels, and the 45 degree line would represent perfect agreement.
Extensions and future work

- Other sets of moment conditions.
- Errors with weak dependence.
- Large number of overidentifying restrictions.