Inequality restricted maximum entropy estimation in Stata

Randall Campbell†, R. Carter Hill‡

Mississippi State University†, Louisiana State University‡

Stata Conference New Orleans July 18, 2013
Generalized Maximum Entropy Estimation

- GME estimator developed by Golan, Judge, and Miller (1996)
- Campbell and Hill (2006) impose inequality restrictions on GME estimator in a linear regression model
- We develop new STATA commands to obtain GME parameter estimates, with or without inequality restrictions
- Our commands utilizes the optimize() function in MATA
Maximum Entropy Problem

- Entropy is the amount of uncertainty represented by a discrete probability distribution

- Entropy is measured as $H(p) = -\sum p_k \ln(p_k)$

- Jaynes’ dice problem: Estimate unknown probabilities of rolling each value on a die

- Maximize $H(p)$ subject to: $\sum k \times p_k = y$ and $\sum p_k = 1$,
  where $k$ is the number of sides on die and $y$ is the mean from prior rolls
Maximum Entropy Solution

\begin{itemize}
  \item \( L = - \sum p_k \ln(p_k) + \lambda (y - \sum k \cdot p_k) + \gamma (1 - \sum p_k) \)
  \item \( dL/dp_k = -1 - \ln(p_k) - \lambda k - \gamma = 0 \)
  \item Which implies \( \hat{p}_k = \frac{\exp(-\hat{\lambda}k)}{\sum \exp(-\hat{\lambda}k)} \)
  \item Substitute into Lagrangian and minimize (Golan, Judge, and Miller (1996)):
    \[ L(\lambda) = - \sum \hat{p}_k(\hat{\lambda}) \ln(\hat{p}_k(\hat{\lambda})) + \hat{\lambda}(y - \sum k \cdot \hat{p}_k(\hat{\lambda})) \]
\end{itemize}
Example: Estimated Probabilities for 6-Sided Die

<table>
<thead>
<tr>
<th>$y$</th>
<th>$\hat{p}_1$</th>
<th>$\hat{p}_2$</th>
<th>$\hat{p}_3$</th>
<th>$\hat{p}_4$</th>
<th>$\hat{p}_5$</th>
<th>$\hat{p}_6$</th>
<th>$H(\hat{p})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.5</td>
<td>0.1667</td>
<td>0.1667</td>
<td>0.1667</td>
<td>0.1667</td>
<td>0.1667</td>
<td>0.1667</td>
<td>1.792</td>
</tr>
<tr>
<td>4.0</td>
<td>0.1031</td>
<td>0.1227</td>
<td>0.1462</td>
<td>0.1740</td>
<td>0.2072</td>
<td>0.2468</td>
<td>1.749</td>
</tr>
<tr>
<td>4.5</td>
<td>0.0544</td>
<td>0.0788</td>
<td>0.1142</td>
<td>0.1655</td>
<td>0.2398</td>
<td>0.3475</td>
<td>1.614</td>
</tr>
<tr>
<td>5.0</td>
<td>0.0205</td>
<td>0.0385</td>
<td>0.0723</td>
<td>0.1357</td>
<td>0.2548</td>
<td>0.4781</td>
<td>1.368</td>
</tr>
<tr>
<td>5.5</td>
<td>0.0029</td>
<td>0.0086</td>
<td>0.0255</td>
<td>0.0755</td>
<td>0.2238</td>
<td>0.6637</td>
<td>0.953</td>
</tr>
<tr>
<td>6.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
STATA MEDICE Command

- The part that evaluates the function:
  ```c
  void mydice_eval(todo, lam, x, k, y, L, g, H)
  {
    a = J(k,1,.)
    b = J(k,1,.)
    for (i=1; i<=k; i++)
    {
      a[i] = exp(-x[i]*lam)
    }
    L = lam*y + ln(sum(a))
  }
  ```
Example

- `scalar k = 6`
- `scalar y = 4.5`
- `medice k y`

```
r(phat)[6,1]  
c1  
r1      .05435317
r2      .07877155
r3      .11415998
r4      .1654468
r5      .23977444
r6      .34749406
entropy function = 1.6135811
```
Model Specification for GME Estimation

- Model: \( y = X\beta + e \)

- Set up such that unknown parameters in form of probabilities (GJM, 1996):

\[
\beta = Zp = \begin{bmatrix}
z_1' & 0 & \ldots & 0 \\
0 & z_2' & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & z_k'
\end{bmatrix} \begin{bmatrix}
p_1 \\
p_2 \\
\vdots \\
p_k
\end{bmatrix}
\]

- Do same thing with error term
Generalized Entropy Function

- Reparameterized Model: $y = XZp + Vw$

- $Z$ and $V$ are parameter and error 'support matrices', respectively

- $\text{Max } H(p, w) = -p' \ln(p) - w' \ln(w)$, subject to

- $y = XZp + Vw$

- $(I_K \otimes i'_M)p = i_K$

- $(I_N \otimes i'_J)w = i_N$, where $M$ is # support points for each parameter; $J$ is # support points for errors
GME Solution in Linear Regression Model

- $L = -p' \ln(p) - w' \ln(w) + \lambda' (XZp + Vw - y) + \\
  \gamma' [i_K - (l_K \otimes i_M)p] + \delta' [i_N - (l_N \otimes i_J)w]$

Solution:

- $\hat{p}_{km} = \exp(z_{km}x_k' \hat{\lambda}) / \sum_{m=1}^{M} \exp(z_{km}x_k' \hat{\lambda}),$ and

- $\hat{w}_{nj} = \exp(v_{nj} \hat{\lambda}_n) / \sum_{j=1}^{J} \exp(v_{nj} \hat{\lambda}_n)$
As in the dice problem, we can substitute these solutions into Lagrangian and solve for $\lambda$.

GME parameter estimates given by $\hat{\beta}_{GME} = Z\hat{\rho}$

The .ado code GMEREG specifies supports as

\[ z_k' = [-5b_k, -2.5b_k, 0, 2.5b_k, 5b_k] \quad \text{and} \]

\[ v_i' = [-3\hat{\sigma}, -1.5\hat{\sigma}, 0, 1.5\hat{\sigma}, 3\hat{\sigma}] \quad (\text{Pukelsheim, 1994}) \]
Example using Generated Data

- With the automated command, the user types gmereg y x1 x2 x3 ... 

- As an example, we generate 7 standard normal random variables (x1-x7) and a standard normal error

\[ y = 2x_1 + 1.2x_2 + 0.35x_3 + 0.4x_4 - 0.05x_5 + 0.8x_6 - 3x_7 + e \]
Results

Results obtained using REG and GMEREGERG in STATA:

<table>
<thead>
<tr>
<th></th>
<th>true</th>
<th>reg</th>
<th>gmereg</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>2.000</td>
<td>1.974</td>
<td>1.948</td>
</tr>
<tr>
<td>$x_2$</td>
<td>1.200</td>
<td>1.149</td>
<td>1.160</td>
</tr>
<tr>
<td>$x_3$</td>
<td>0.350</td>
<td>0.218</td>
<td>0.203</td>
</tr>
<tr>
<td>$x_4$</td>
<td>0.400</td>
<td>0.134</td>
<td>0.108</td>
</tr>
<tr>
<td>$x_5$</td>
<td>-0.050</td>
<td>0.060</td>
<td>0.016</td>
</tr>
<tr>
<td>$x_6$</td>
<td>0.800</td>
<td>0.822</td>
<td>0.791</td>
</tr>
<tr>
<td>$x_7$</td>
<td>-3.000</td>
<td>-3.198</td>
<td>-3.194</td>
</tr>
<tr>
<td>const</td>
<td>0.000</td>
<td>-0.125</td>
<td>-0.110</td>
</tr>
</tbody>
</table>
STATA GMEREGM Command

- GMEREG automatically sets up the parameter support matrix based on initial OLS estimates.

- A second version, GMEREGM, allows the user to specify their own support matrix.

- This allows the user to specify wider or narrower bounds and, as we will show, to impose cross-parameter restrictions.
Specifying the Parameter Support Matrix

Suppose we wish to impose wide, uninformative bounds; allow each parameter to fall between -20 and 20.

In our example, we have $K = 8$ parameters to estimate and have $M = 5$ support points. We specify the $K \times M$ support matrix:

```
.matrix zmat = (-20,-20,-20,-20,-20,-20,-20,-20
                  -10,-10,-10,-10,-10,-10,-10,-10
                  0,0,0,0,0,0,0,0
                 10,10,10,10,10,10,10,10
                20,20,20,20,20,20,20,20)
```

```
. gmeregm y x1 x2 x3 x4 x5 x6 x7
```

Alternative Specifications

- Suppose we wish a more informative prior:
  
  \[
  \text{matrix zmat = (-5,-5,-1,-1,-1,-5,-5) -2.5,-2.5,-0.5,-0.5,-0.5,-2.5,0.5,0.5,0.5,2.5,2.5,2.5,5,5,1,1,1,5,5)}
  \]

- Or one that is not symmetric about zero:

  \[
  \text{matrix zmat = (-10,-10,-1,-1,-1,-1,-15,-10) -0,0,0,-0.5,0,-10,5,5,1,1,0,1,-5,0 10,10, 2,2,0.5,2,0,5 15,15,3,3,1,3,10,10)}
  \]
### Results

Results based on different support matrices:

<table>
<thead>
<tr>
<th></th>
<th>true</th>
<th>zmat1</th>
<th>zmat2</th>
<th>zmat3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>2.000</td>
<td>1.942</td>
<td>1.905</td>
<td>1.936</td>
</tr>
<tr>
<td>$x_2$</td>
<td>1.200</td>
<td>1.161</td>
<td>1.113</td>
<td>1.167</td>
</tr>
<tr>
<td>$x_3$</td>
<td>0.350</td>
<td>0.241</td>
<td>0.215</td>
<td>0.286</td>
</tr>
<tr>
<td>$x_4$</td>
<td>0.400</td>
<td>0.175</td>
<td>0.084</td>
<td>0.236</td>
</tr>
<tr>
<td>$x_5$</td>
<td>-0.050</td>
<td>0.050</td>
<td>0.010</td>
<td>0.047</td>
</tr>
<tr>
<td>$x_6$</td>
<td>0.800</td>
<td>0.821</td>
<td>0.636</td>
<td>0.845</td>
</tr>
<tr>
<td>$x_7$</td>
<td>-3.000</td>
<td>-3.186</td>
<td>-3.195</td>
<td>-3.183</td>
</tr>
<tr>
<td>const</td>
<td>0.000</td>
<td>-0.138</td>
<td>-0.168</td>
<td>-0.129</td>
</tr>
</tbody>
</table>
Sign Restrictions

- Our estimates all took correct signs except the estimate for $\beta_5$.

- Parameter sign restrictions are easily accomplished through support matrix. For example,

```plaintext
. matrix zmat = (0,0,0,0,-0.4,0,-10,-10
               2.5,1,0.25,0.25,-0.3,0.5,-7.5,-5
               5,2,0.5,0.5,-0.2,1,-5,0
               7.5,3,0.75,0.75,-0.1,1.5,-2.5,5
               10,4,1,1,0,2,0,10)
```
Cross-Parameter Restrictions

- Note that while the true $\beta_4 > \beta_3$, this does not hold for our estimates.
- Suppose we wish to restrict $\hat{\beta}_4 > \hat{\beta}_3$.
- Campbell and Hill (2006) impose such restrictions through the support matrix:
  \[
  \begin{bmatrix}
  \beta_3 \\
  \beta_4
  \end{bmatrix} = Z^* \begin{bmatrix}
  p_3 \\
  p_4
  \end{bmatrix} = \begin{bmatrix}
  z'_3 & 0 \\
  z_3 & z_4'
  \end{bmatrix} \begin{bmatrix}
  p_3 \\
  p_4
  \end{bmatrix}
  \]

- Specifying $z'_4$ to take only non-negative values ensures that $\hat{\beta}_4 > \hat{\beta}_3$. 

The problem becomes harder computationally; the GME solution is

\[ \hat{p}_{km} = \frac{\exp(z_{km}x_1'\hat{\lambda} + z_{km}x_2'\hat{\lambda} + \ldots + z_{km}x_K'\hat{\lambda})}{\sum_{m=1}^M \exp(z_{km}x_1'\hat{\lambda} + z_{km}x_2'\hat{\lambda} + \ldots + z_{km}x_K'\hat{\lambda})} \]

When parameter support is block diagonal, cross-product terms drop out.
Restricted GME Estimation

- A third version, GMEREGMR, performs restricted estimation. However, must enter larger zmat (KM x K)
- Enter separate support column vector for each parameter (and a vector of 0s)

\[
\begin{align*}
\text{matrix } z_0 &= (0, 0, 0, 0, 0) \\
\text{matrix } z_1 &= (0, 5, 10, 15, 20) \\
\text{matrix } z_2 &= z_1 \\
\text{matrix } z_3 &= z_1 \\
\text{matrix } z_4 &= (0, 0.5, 1, 1.5, 2) \\
\text{matrix } z_5 &= (-20, -15, -10, -5, 0) \\
\text{matrix } z_6 &= z_1 \\
\text{matrix } z_7 &= z_5 \\
\text{matrix } z_8 &= (-20, -10, 0, 10, 20)
\end{align*}
\]
Now combine into zmat; matrix is block diagonal except for column 4 which include z3 and z4

```
. matrix zmat = (z1,z0,z0,z0,z0,z0,z0,z0  
                   z0,z2,z0,z0,z0,z0,z0,z0  
                   z0,z0,z3,z3,z0,z0,z0,z0  
                   z0,z0,z0,z4,z0,z0,z0,z0  
                   z0,z0,z0,z0,z5,z0,z0,z0  
                   z0,z0,z0,z0,z0,z6,z0,z0  
                   z0,z0,z0,z0,z0,z0,z7,z0  
                   z0,z0,z0,z0,z0,z0,z0,z8)
```

Now enter command  `gmeregmr = y x1 x2 x3 x4 x5 x6 x7`
Final Example - Multiple Restrictions

- Suppose we wish to impose $\beta_4 > \beta_3$ and $\beta_6 > \beta_3 + \beta_4$ and $\beta_2 > \beta_3 + \beta_6$
- The parameter support vectors we chose:

  - matrix z0 = (0 0 0 0 0)
  - matrix z1 = (0 5 10 15 20)
  - matrix z2 = (0 0.5 1 1.5 2)
  - matrix z3 = z1
  - matrix z4 = z2
  - matrix z5 = (-20 -15 -10 -5 0)
  - matrix z6 = z2
  - matrix z7 = z5
  - matrix z8 = (-20 -10 0 10 20)
Zmat - Multiple Restrictions

- The zmat to ensure restrictions hold
  . matrix zmat = (z1,z0,z0,z0,z0,z0,z0,z0 z0,z2,z0,z0,z0,z0,z0,z0 z0,3*z3,z3,z3,z0,2*z3,z0,z0 z0,z4,z0,z4,z0,z4,z0,z0 z0,z0,z0,z0,z5,z0,z0,z0 z0,z6,z0,z0,z0,z6,z0,z0 z0,z0,z0,z0,z0,z0,z7,z0 z0,z0,z0,z0,z0,z0,z0,z8)

- Note: \( \hat{\beta}_4 = \hat{\beta}_3 + z'_4 p_4 > \hat{\beta}_3 \); \( \hat{\beta}_6 = \hat{\beta}_3 + \hat{\beta}_4 + z'_6 p_6 > \hat{\beta}_3 + \hat{\beta}_4 \); and \( \hat{\beta}_2 = \hat{\beta}_3 + \hat{\beta}_6 + z'_2 p_2 > \hat{\beta}_3 + \hat{\beta}_6 \)
## Results

Results based on different support matrices:

<table>
<thead>
<tr>
<th></th>
<th>true</th>
<th>sign only</th>
<th>single</th>
<th>multiple</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>2.000</td>
<td>1.961</td>
<td>2.009</td>
<td>2.067</td>
</tr>
<tr>
<td>$x_2$</td>
<td>1.200</td>
<td>1.217</td>
<td>1.250</td>
<td>1.324</td>
</tr>
<tr>
<td>$x_3$</td>
<td>0.350</td>
<td>0.363</td>
<td>0.165</td>
<td>0.006</td>
</tr>
<tr>
<td>$x_4$</td>
<td>0.400</td>
<td>0.347</td>
<td>0.385</td>
<td>0.224</td>
</tr>
<tr>
<td>$x_5$</td>
<td>-0.050</td>
<td>-0.150</td>
<td>-0.021</td>
<td>-0.041</td>
</tr>
<tr>
<td>$x_6$</td>
<td>0.800</td>
<td>0.849</td>
<td>0.952</td>
<td>0.860</td>
</tr>
<tr>
<td>$x_7$</td>
<td>-3.000</td>
<td>-3.220</td>
<td>-3.186</td>
<td>-3.194</td>
</tr>
<tr>
<td>const</td>
<td>0.000</td>
<td>-0.125</td>
<td>-0.139</td>
<td>-0.163</td>
</tr>
</tbody>
</table>
Work to Do

- Our commands provide GME estimates and can be used with either no user input or a user-specified parameter support.

- We are working to output additional statistics such as prior mean and standard errors.

- We are also developing STATA GME commands for binary and multinomial choice models, censored regression, ..