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: $\sum k * p_k = y$ 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

- $L = -\sum p_k \ln(p_k) + \lambda(y - \sum k * p_k) + \gamma(1 - \sum p_k)$
- $dL/dp_k = -1 - \ln(p_k) - \lambda k - \gamma = 0$
- Which implies $\widehat{p}_k = \frac{\exp(-\widehat{\lambda} k)}{\sum \exp(-\widehat{\lambda} k)}$
- Substitute into Lagrangian and minimize (Golan, Judge, and Miller (1996)):

$$L(\lambda) = -\sum \widehat{p}_k(\widehat{\lambda}) \ln(\widehat{p}_k(\widehat{\lambda})) + \widehat{\lambda}(y - \sum k * \widehat{p}_k(\widehat{\lambda}))$$

Example: Estimated Probabilities for 6-Sided Die

y	$\widehat{p}_1$	$\widehat{p}_2$	$\widehat{p}_3$	$\widehat{p}_4$	$\widehat{p}_5$	$\widehat{p}_6$	$H(\widehat{p})$
3.5	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667	1.792
4.0	0.1031	0.1227	0.1462	0.1740	0.2072	0.2468	1.749
4.5	0.0544	0.0788	0.1142	0.1655	0.2398	0.3475	1.614
5.0	0.0205	0.0385	0.0723	0.1357	0.2548	0.4781	1.368
5.5	0.0029	0.0086	0.0255	0.0755	0.2238	0.6637	0.953
6.0	0	0	0	0	0	1	0

STATA MEDICE Command

- The part that evaluates the function:

```
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):

$$\beta = Zp = \begin{bmatrix} z_1' & 0 & \dots & 0 \\ 0 & z_2' & \dots & 0 \\ . & . & . & . \\ 0 & 0 & \dots & z_k' \end{bmatrix} . \begin{bmatrix} p_1 \\ p_2 \\ . \\ p_k \end{bmatrix}$$

- Do same thing with error term

Generalized Entropy Function

- Reparameterized Model: $y = XZp + Vw$
- Z and V are parameter and error 'support matrices', respectively
- Max $H(p, w) = -p^{'} \ln(p) - w^{'} \ln(w)$, subject to
- $y = XZp + Vw$
- $(I_K \bigotimes i_M^{'})p = i_K$
- $(I_N \bigotimes i_J^{'})w = i_N$,
  where M is # support points for each parameter; J is # support points for errors

GME Solution in Linear Regression Model

- $L = -p^{'}\ln(p) - w^{'}\ln(w) + \lambda^{'}(XZp + Vw - y) + \gamma^{'}[i_K - (I_K \bigotimes i_M^{'})p] + \delta^{'}[i_N - (I_N \bigotimes i_J^{'})w]$
- Solution:
- $\widehat{p}_{km} = \exp(z_{km}x_k^{'}\widehat{\lambda}) / \sum_{m=1}^{M} \exp(z_{km}x_k^{'}\widehat{\lambda})$, and

$$\widehat{w}_{nj} = \exp(v_{nj}\widehat{\lambda}_n) / \sum_{j=1}^{J} \exp(v_{nj}\widehat{\lambda}_n)$$

STATA GMEREG Command

- As in the dice problem, we can substitute these solutions into Lagrangian and solve for λ.
- GME parameter estimates given by $\widehat{\beta}_{GME} = Z\widehat{p}$
- The .ado code GMEREG specifies supports as

$z_k^{'} = [-5b_k, -2.5b_k, 0, 2.5b_k, 5b_k]$ and

$v_i^{'} = [-3\widehat{\sigma}, -1.5\widehat{\sigma}, 0, 1.5\widehat{\sigma}, 3\widehat{\sigma}]$ (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 GMEREG in STATA:

	true	reg	gmereg
x_1	2.000	1.974	1.948
x_2	1.200	1.149	1.160
x_3	0.350	0.218	0.203
x_4	0.400	0.134	0.108
x_5	-0.050	0.060	0.016
x_6	0.800	0.822	0.791
x_7	-3.000	-3.198	-3.194
const	0.000	-0.125	-0.110

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 x 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:

```
. matrix zmat = (-5,-5,-1,-1,-1,-1,-5,-5 \
                 -2.5,-2.5,-0.5,-0.5,-0.5,-0.5,-2.5,-2.5 \
                 0,0,0,0,0,0,0,0 \
                 2.5,2.5, 0.5,0.5,0.5,0.5,2.5,2.5 \
                 5,5,1,1,1,1,5,5)
```

- Or one that is not symmetric about zero:

```
. matrix zmat = (-10,-10,-1,-1,-1,-1,-15,-10 \
                 0,0,0,0,-0.5,0,–10,-5 \
                 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:

	true	zmat1	zmat2	zmat3
x_1	2.000	1.942	1.905	1.936
x_2	1.200	1.161	1.113	1.167
x_3	0.350	0.241	0.215	0.286
x_4	0.400	0.175	0.084	0.236
x_5	-0.050	0.050	0.010	0.047
x_6	0.800	0.821	0.636	0.845
x_7	-3.000	-3.186	-3.195	-3.183
const	0.000	-0.138	-0.168	-0.129

Sign Restrictions

- Our estimates all took correct signs except the estimate for β_5
- Parameter sign restrictions are easily accomplished through support matrix. For example,

```
. 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 $\widehat{\beta}_4 > \widehat{\beta}_3$
- 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 $\widehat{\beta}_4 > \widehat{\beta}_3$

GME Solution with Cross-Parameter Restrictions

- The problem becomes harder computationally; the GME solution is
- $\widehat{p}_{km} = \dfrac{\exp(z_{km}x_1'\widehat{\lambda}+z_{km}x_2'\widehat{\lambda}+\ldots+z_{km}x_K'\widehat{\lambda})}{\sum_{m=1}^{M}\exp(z_{km}x_1'\widehat{\lambda}+z_{km}x_2'\widehat{\lambda}+\ldots+z_{km}x_K'\widehat{\lambda})}$
- 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)

```
. matrix z0 = (0\0\0\0\0)
. matrix z1 = (0\5\10\15\20)
. matrix z2 = z1
. matrix z3 = z1
. matrix z4 = (0\0.5\1\1.5\2)
. matrix z5 = (-20\-15\-10\-5\0)
. matrix z6 = z1
. matrix z7 = z5
. matrix z8 = (-20\-10\0\10\20)
```

STATA GMEREGMR Command

- 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: $\widehat{\beta}_4 = \widehat{\beta}_3 + z_4' p_4 > \widehat{\beta}_3$; $\widehat{\beta}_6 = \widehat{\beta}_3 + \widehat{\beta}_4 + z_6' p_6 > \widehat{\beta}_3 + \widehat{\beta}_4$; and $\widehat{\beta}_2 = \widehat{\beta}_3 + \widehat{\beta}_6 + z_2' p_2 > \widehat{\beta}_3 + \widehat{\beta}_6$

Results

- Results based on different support matrices:

	true	sign only	single	multiple
x_1	2.000	1.961	2.009	2.067
x_2	1.200	1.217	1.250	1.324
x_3	0.350	0.363	0.165	0.006
x_4	0.400	0.347	0.385	0.224
x_5	-0.050	-0.150	-0.021	-0.041
x_6	0.800	0.849	0.952	0.860
x_7	-3.000	-3.220	-3.186	-3.194
const	0.000	-0.125	-0.139	-0.163

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, censroed regression, ..