Outline

1. Introduction
2. Consistent estimation - IV
3. Review of GMM estimation
4. Consistent estimation - GMM
5. Production function example redux
Consistent estimation

- **Balestra-Nerlove** (for fixed effects): IV on the within model using within X’s (current and lagged) as instruments for lagged y and the X’s.
- **Anderson-Hsiao**: IV on the differenced model using y lagged twice and differenced X’s as instruments.
- **Arellano-Bond**: GMM on the differenced model using a full set of valid lags as instruments.
- **Arellano-Bover/Blundell-Bond**: GMM on the differenced model using lagged levels as instruments and the level model using lagged differences as instruments.
Balestra-Nerlove (197?)

Within (LSDV) estimation, so fixed effects have been removed.

Appropriate when X’s are strictly exogenous.

\[ \tilde{y}_{it} = \delta \tilde{y}_{i,t-1} + \tilde{x}_{it}' \beta + \tilde{\epsilon}_{it} \quad t = 2, ..., T; i = 1, ..., N \]

*instruments*: \( \tilde{x}_{it}, \tilde{x}_{i,t-1} \)

If the X’s are uncorrelated with past, current, and future disturbances, they have the correct properties to serve as instruments for themselves and lagged y.
Anderson-Hsiao (1982)

First differenced estimation to remove the fixed effect
Appropriate when Xs are weakly exogenous (predetermined)

\[ y_{it} - y_{i,t-1} = \delta(y_{i,t-1} - y_{i,t-2}) + (x_{it} - x_{i,t-1})' \beta + \varepsilon_{it} - \varepsilon_{i,t-1} \]

\[ \Delta y_{it} = \delta\Delta y_{i,t-1} + \Delta x_{it} ' \beta + \Delta \varepsilon_{it} \]

*instruments*: \( y_{i,t-2}, \Delta x_{it} \) or \( y_{i,t-2}, \Delta x_{i,t-1} \)

Choice of instruments depends on assumptions about the simultaneity between X and disturbances.
Production function estimates
Balanced panel of 582 firms for 10 years (1986-95)
Dependent variable: S/L

<table>
<thead>
<tr>
<th></th>
<th>Within</th>
<th>Within (IV)*</th>
<th>FD</th>
<th>FD (IV)***</th>
<th>FD(IV)**</th>
</tr>
</thead>
<tbody>
<tr>
<td>log (L)</td>
<td>-.090</td>
<td>-.091</td>
<td>-.280</td>
<td>-.314</td>
<td>-.098</td>
</tr>
<tr>
<td>log (K/L)</td>
<td>0.159</td>
<td>0.273</td>
<td>0.180</td>
<td>0.189</td>
<td>2.54</td>
</tr>
<tr>
<td>log (S/L) (-1)</td>
<td>0.414</td>
<td>-.431</td>
<td>-.140</td>
<td>0.391</td>
<td>.266</td>
</tr>
<tr>
<td>&quot;Sargan&quot; test</td>
<td>31.8</td>
<td>0.000</td>
<td>--</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>D-W</td>
<td>1.68</td>
<td>0.62</td>
<td>1.65</td>
<td>2.39</td>
<td>1.82</td>
</tr>
<tr>
<td>s.e.</td>
<td>0.119</td>
<td>0.167</td>
<td>0.126</td>
<td>0.148</td>
<td>0.436</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.948</td>
<td>0.898</td>
<td>0.232</td>
<td>0.065</td>
<td>0.092</td>
</tr>
<tr>
<td>First stage R2</td>
<td>0.882</td>
<td>0.065</td>
<td>0.042</td>
<td>.06,.02,.22</td>
<td></td>
</tr>
<tr>
<td># obs.</td>
<td>4,656</td>
<td>4,656</td>
<td>4,656</td>
<td>4,656</td>
<td>4,656</td>
</tr>
</tbody>
</table>

*Balestra-Nerlove estimator with x(-1) as instruments.
**Anderson-Hsiao estimator with y(-2) and dx as instruments.
***Anderson-Hsiao estimator with y(-2) and dx(-1) as instruments.
GMM estimation (overview)

An estimator based on the method of moments.

\[ \text{theoretical moments} = \text{empirical moments} \]
\[ \text{(functions of parameters)} = \text{(functions of the data)} \]

Solve for the parameters as a function of the data.

Generalized – more moments than parameters;
use an estimate of their variance as weights and minimize a chi-squared statistic.

Properties – under regularity conditions, estimator is consistent asymptotically normal with an estimable covariance.
GMM using OCs (review)

Define:

\[ Z_i = \text{a vector of instruments for firm } i \text{ (may include } X_i) \]
\[ u_i = \text{a vector of disturbance (residual) functions} \]
\[ \text{e.g., } u_{it} = y_{it} - x_{it} \beta \]

The \( u_{it} \)'s are assumed to be mean zero conditional on the \( Z_i \).
The sample orthogonality conditions are

\[ g(\beta) = \frac{1}{N} \sum_{i=1}^{N} u_i^t \otimes Z_i^t \]

The following estimator is consistent for \( \beta \):

\[ \hat{\beta} = \arg \min_{\beta} g(\beta) A_N g(\beta)' \]
GMM estimation (review)

Any positive definite $A_N$ will yield consistent estimates; standard errors can be estimated using the robust (Eicker-White) formula.

Optimal choice of $A_N$ equal in expectation to the inverse of the covariance of the orthogonality conditions. In the sample, this means

$$\hat{A}_N^* = \left\{ \frac{1}{N} \sum_{i=1}^{N} \left[ \hat{\mu}_i \hat{\mu}_i ' \otimes Z_i Z_i ' \right] \right\}^{-1}$$

Instead of the Kronecker product form for the OCs, the $Z_{it}$ may be different for each period $t$.

For panel data, they usually will be different.

Can be implemented by applying a transformation matrix consisting of zeros and ones to the original instrument list $Z_i$. 

6/10/2003 Economics 244 - Spring 2002
GMM estimation (review)

Specification testing (Sargan test, Hansen’s J-statistic)
If there are $m$ orthogonality conditions and $k$ unknown parameters, then under the null
\[ g(\hat{\beta})\hat{V}_N^{-1}g(\hat{\beta})' \sim \chi^2(m - k) \]
where $\hat{V}_N$ is a consistent estimate of $V(g(\beta))$

Use as an omnibus test for the validity of the moment conditions when the model is overidentified ($m>k$).
Tests can be constructed for the validity of restrictions in a model with $k_1$ parameters that is nested within a model with $k_2$ parameters ($k_2>k_1$) simply by differencing this statistic (and using $k_2-k_1$ as d.f.)

$m = k$: model exactly identified, and the choice of weighting matrix $A_N$ does not affect the estimates.
Dynamic Panels

Right hand side variables of a panel data model are predetermined rather than strictly exogenous:

=> Estimate the fixed effects model using first differenced estimation (rather than within) with lagged RHS variables as instruments.

GMM estimation:

• (quasi-)efficient
• different number of instruments for each time period (depending on the number of lags available).
• Allows for heteroskedasticity and arbitrary serial correlation.
• Easy to generalize to a nonlinear model with additive disturbances.

**WARNING:** if you plan to use GMM methods, be aware that this is a developing area of research because of the occasional poor performance of GMM for panel data in finite samples. See Blundell and Bond (1999), Bond (2002), on the reading list.
GMM for dynamic panels

Recall the importance of the following assumption:

\[ E[\varepsilon_{it}|x_i, \alpha_i] = 0 \Rightarrow E[\varepsilon_{it}|x_i] = 0 \]

Now consider a model with a lagged dependent variable and \( T=3 \):

\[ y_{it} = \alpha_0 + \beta y_{it-1} + \alpha_i + \varepsilon_{it} \quad i = 1,\ldots,N; \quad t = 1,3 \]

Suppose we difference the data to remove the fixed effect:

\[ y_{i3} - y_{i2} = \beta(y_{i2} - y_{i1}) + (\varepsilon_{i3} - \varepsilon_{i2}) \]

Clearly this model fails to satisfy the assumption necessary for consistency, because

\[ E[(\varepsilon_{i3} - \varepsilon_{i2})|(y_{i2} - y_{i1})] = - E[\varepsilon_{i2}|(y_{i2} - y_{i1})] \neq 0 \]
GMM for dynamic panels

*The Instrumental Variables solution:* the assumption

\[ E[(\varepsilon_{i3} - \varepsilon_{i2})| (y_{i2} - y_{i1})] = 0 \]

has failed.

Need an instrument uncorrelated with \((\varepsilon_{i3} - \varepsilon_{i2})\) and correlated with \((y_{i2} - y_{i1})\). Use \(y_{i1}\), the dependent variable lagged twice.

GMM is generalized version of this IV:

- more than 2 time periods
- (possibly) predetermined right hand side variables
- \(E[\varepsilon_i \varepsilon_i'] = \Sigma_i\) rather than \(\sigma^2 I_T\)
Basic GMM for dynamic panels

\[ y_{it} = \alpha_0 + \beta x_{it} + \alpha_i + \varepsilon_{it} \quad i = 1, \ldots, N; \ t = 1, \ldots, T \]

or

\[ y_{it} = \alpha_0 + \beta x_{it} + u_{it} \]

But now

\[ E[\varepsilon_{it}|x_{iT}, \ldots, x_{i,t+1}, x_{it}, x_{i,t-1}, \ldots, x_{i1}] \neq 0 \]

although

\[ E[\varepsilon_{it}|x_{it}, x_{i,t-1}, \ldots, x_{i1}] = 0 \]

Difference the model to remove the effect:

\[ \Delta y_{it} = \beta \Delta x_{it} + \Delta u_{it} \quad i = 1, \ldots, N; \ t = 2, \ldots, T \]

If \( x \) is predetermined, then we have

\[ E[\Delta u_{it}|\Delta x_{it}] = E[\varepsilon_{it-1}-\varepsilon_{it-1}|x_{it}-x_{i,t-1}] \neq 0 \]

because of the potential correlation of \( \varepsilon_{it-1} \) and \( x_{it} \). However,

\[ E[\Delta u_{it}|x_{i,t-1}, x_{i,t-2}, \ldots, x_{i1}] = 0 \]

Therefore we can use all prior lags of \( x \) as instruments for the \( t \)th equation.
Implementation example

Simple model with one x.

T-1 equations for the $\Delta u$’s.
T possible instruments $x$.
$\Rightarrow$ T(T-1) potential moment conditions $\Delta u \otimes x$.

However, only these conditions are valid:

(1) $E[\Delta u_2 x_1] = 0$
(2) $E[\Delta u_3 x_2] = 0$  $E[\Delta u_3 x_1] = 0$

……

(T-1) $E[\Delta u_{T} x_{T-1}] = 0$ ....  $E[\Delta u_{T} x_{1}] = 0$

for a total of 1+2+…+(T-1) = (T-1)T/2 conditions.

One way to implement this is to use a diagonal selection matrix $S$ of zeroes and ones. The moment conditions become

$$g(\beta) = \frac{1}{N} \sum_{i=1}^{N} S(u_i \otimes Z_i)$$
Notes

• Measurement error or simultaneity in x can be handled by lagging the instruments once (dropping the lag 1 variables).
  – This allows for exogeneity tests (using the J-statistic).

• The validity of the instruments depends on the correlation pattern of the ε’ s over time. For example, to use \( E[\Delta \varepsilon_{it}\mid x_{i,t-1}, x_{i,t-2}, \ldots, x_{i1}] = 0 \), we need serially uncorrelated ε’s or Δε’s that are MA(1).
  – A-B suggest using LM tests for serial correlation as a test for instrument validity. Regress the residuals from GMM on their own values, lag once (LM(1)) or twice (LM(2)).
GMM estimates – prod. function

582 firms; 10 years; dep. var. = log(Y/L)

<table>
<thead>
<tr>
<th>Instruments</th>
<th>lag 0</th>
<th>lag 1 to t</th>
<th>lag 1 to 3</th>
<th>lag 2 to t</th>
<th>lag 2 to 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FD with FD</td>
<td>FD with Lev</td>
<td>FD with Lev</td>
<td>FD with Lev</td>
<td>FD with Lev</td>
</tr>
<tr>
<td>log (L)</td>
<td>-.272 (.022)</td>
<td>0.096 (.033)</td>
<td>0.208 (.050)</td>
<td>-.058 (.037)</td>
<td>-.057 (.052)</td>
</tr>
<tr>
<td>log (K/L)</td>
<td>0.190 (.019)</td>
<td>0.259 (.033)</td>
<td>0.181 (.059)</td>
<td>0.183 (.048)</td>
<td>0.082 (.060)</td>
</tr>
<tr>
<td>Sargan</td>
<td>14.8</td>
<td>121.6</td>
<td>66.9</td>
<td>92.6</td>
<td>59.2</td>
</tr>
<tr>
<td>DF</td>
<td>14</td>
<td>86</td>
<td>44</td>
<td>70</td>
<td>40</td>
</tr>
<tr>
<td>chi-sq/DF</td>
<td>1.06</td>
<td>1.41</td>
<td>1.52</td>
<td>1.32</td>
<td>1.48</td>
</tr>
<tr>
<td>p-value</td>
<td>0.390</td>
<td>0.007</td>
<td>0.015</td>
<td>0.037</td>
<td>0.026</td>
</tr>
<tr>
<td>LM(1)</td>
<td>-3.90 (.000)</td>
<td>-6.76 (.000)</td>
<td>-6.86 (.000)</td>
<td>-6.27 (.000)</td>
<td>-6.27 (.000)</td>
</tr>
<tr>
<td>LM(2)</td>
<td>-0.25 (.803)</td>
<td>-0.43 (.664)</td>
<td>0.17 (.862)</td>
<td>-0.39 (.696)</td>
<td>-0.03 (.974)</td>
</tr>
</tbody>
</table>

Method of estimation is iterated GMM; robust standard errors in parentheses. P-values of LM tests for serial correlation in parentheses.
GMM estimates – prod. function

582 firms; 10 years; dep. var. = log(Y/L)

<table>
<thead>
<tr>
<th>Instruments</th>
<th>lag 0</th>
<th>lag 1 to t</th>
<th>lag 1 to 3</th>
<th>lag 1 to 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lev with lev</td>
<td>Lev with FD</td>
<td>Lev with FD</td>
<td>System</td>
</tr>
<tr>
<td>log (L)</td>
<td>-.028 (.008)</td>
<td>-.060 (.022)</td>
<td>-.042 (.027)</td>
<td>-.065 (.030)</td>
</tr>
<tr>
<td>log (K/L)</td>
<td>0.332 (.021)</td>
<td>0.336 (.034)</td>
<td>0.331 (.040)</td>
<td>1.165 (.009)</td>
</tr>
<tr>
<td>Sargan</td>
<td>18.0</td>
<td>82.1</td>
<td>53.8</td>
<td>192.6</td>
</tr>
<tr>
<td>DF</td>
<td>16</td>
<td>70</td>
<td>40</td>
<td>73</td>
</tr>
<tr>
<td>chi-sq/DF</td>
<td>1.13</td>
<td>1.17</td>
<td>1.35</td>
<td>2.64</td>
</tr>
<tr>
<td>p-value</td>
<td>0.323</td>
<td>0.153</td>
<td>0.071</td>
<td>0.000</td>
</tr>
<tr>
<td>LM(1)</td>
<td>35.3 (.000)</td>
<td>32.4 (.000)</td>
<td>31.7 (.000)</td>
<td>35.4 (.000)</td>
</tr>
<tr>
<td>LM(2)</td>
<td>32.5 (.000)</td>
<td>29.4 (.000)</td>
<td>28.9 (.000)</td>
<td>31.8 (.000)</td>
</tr>
</tbody>
</table>
Other applications

- Returns to R&D using production functions
- Labor demand (Arellano-Bond)
- Investment equations
  - (Mairesse, Hall, and Mulkay 2000) - comparison of methods
  - (Mulkay, Mairesse, and Hall 2001) - uses ADL (2,2) error corrected accelerator model to test for liquidity constraints in R&D and investment