Generalized Methods of Moments (GMM) Estimation with Applications using STATA

David Guilkey
Review of Recursive Simultaneous Equations Models

The GMM estimator is typically used to correct for bias caused by endogenous explanatory variables. So we start with a brief review of the standard recursive simultaneous equations model:

Simple recursive structural equation model with endogenous regressors:

\[ Y_{i1} = \beta X_i + \varepsilon_{i1} \]

\[ Y_{i2} = \alpha Y_{i1} + \varepsilon_{i2} \]

OLS estimation of second equation will lead to biased and inconsistent results if \( E(\varepsilon_{i1}\varepsilon_{i2}) \neq 0 \) i.e. if there is overlap in the set of unobservable variables that affect the two outcomes – sometimes called unobserved heterogeneity.

Examples:

The dependent variables are education and wage rate. Highly motivated individuals may both get more education and receive higher wages – the effect of education on wages is biased upwards (positive effect of education is overstated).

Second example: contraceptive use and fertility. If the unobservable is fecundity, highly fecund women may be more likely to use contraception and also more likely to have children – biasing the effect of contraceptive use towards zero (negative effect of contraceptive use biased towards zero).
Two Stage Least Squares

1. Estimate the first equation by OLS to get $\hat{\beta}$.
2. Replace $Y_{i1}$ with $\tilde{Y}_{i1} = \hat{\beta}X_i$ and run OLS:

Note that this model can be estimated since we assume that $X_i$ does not have a direct effect on $Y_{i2}$.

Suppose that it did:

$$Y_{i2} = \alpha Y_{i1} + \delta X_i + \epsilon_{i2}$$

Then, if we make the two stage least squares substitution, we get:

$$Y_{i2} = \alpha(\hat{\beta}X_i) + \delta X_i + \epsilon_{i2}$$

The perfect collinearity between the regressors implies that OLS cannot separately identify the two coefficients.

Note that if the second equation takes the following form:

$$Y_{i2} = \alpha Y_{i1} + \delta X_i^2 + \epsilon_{i2}$$

Then the model would be identified by functional form.

Overidentification occurs if there are more exclusion restrictions than needed – typically the higher the level of overidentification, the more efficient the estimator.
How do you find appropriate instruments (some examples).

1. Effect of physical activity on BMI.

Could use the presence of recreational facilities in community where person lives. Must be careful – individuals may self select into communities.

2. Effect of contraceptive use on fertility.

Presence of family planning facilities in the community. Again, must be careful if the government targets the placement of facilities. In some cases, this is not an issue – MATLAB in Bangladesh.


The exogenous shock of twins on fertility.

4. Effect of fertility on economic outcomes has used twins and sex composition as identifying variables.

5. Effect of time spent on homework on results on achievement tests.

One author used a dummy variable for whether or not the student’s roommate had video games.


Angrist used the draft lottery number.
7. Effect of past fertility on contraceptive use

Backdate the availability of services to the beginning of the woman’s child bearing years.

**Review of Generalized Least Squares (GLS)**

Consider the basic multivariate model:

\[ Y = X \beta + \varepsilon \]

Where \( Y \) is N x 1, \( X \) is N x K, and \( \varepsilon \) is N x 1

Ideal error term assumption is \( \varepsilon \sim N(0, \sigma^2 I) \) where I is an N x N identity matrix.

Leads to OLS estimator:

\[ \hat{\beta} = (X' X)^{-1} X' Y \]

Now suppose that \( \varepsilon \sim N(0, \Omega) \) -- unequal diagonal elements means that we have heteroskedastic errors and non-zero off diagonal elements means that we have autocorrelation.

Under these conditions, OLS is still unbiased but the standard errors are incorrect. Correct covariance matrix is:

\[ Cov(\hat{\beta}) = (X' X)^{-1} X' \Omega X(XX). \]

Sometimes referred to as sandwich estimator.

Can approximate this covariance matrix using robust option in STATA for heteroskedasticity or the Newy-West option for autocorrelation and heteroskedasticity.
An improved estimator is the GLS estimator:

\[ \hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y \]

This estimator is more efficient than OLS since it uses more information.

Example for heteroskedasticity:

\[ Y_i = X_i\beta + \varepsilon_i \]

Where all terms are scalars.

OLS: \[ \hat{\beta} = \frac{\sum_{i=1}^{N} X_iY_i}{\sum_{i=1}^{N} X_i^2} \]

GLS: \[ \hat{\beta} = \frac{\sum_{i=1}^{N} X_iY_i / \sigma_i^2}{\sum_{i=1}^{N} X_i^2 / \sigma_i^2} \]

Observations with high error variance down weighted relative two observations with low error variance.
Generalized Method of Moments

Suppose we have the following simple model:

\[ Y_i = X_i \beta + \epsilon_i \text{ where it is assumed that } E(X_i \epsilon_i) = 0 \]

The method of moments estimator sets the sample counterpart of this condition to zero:

\[ \frac{1}{N} \sum_{i=1}^{N} X_i (Y_i - X_i \beta) = 0 \]

which leads to \[ \hat{\beta} = \frac{\sum X_i Y_i}{\sum X_i^2} \]

Which is the OLS estimator.

Now suppose \( E(X_i \epsilon_i) \neq 0 \) but \( E(Z_i \epsilon_i) = 0 \) where \( Z \) is some other variable. If we set the sample counterpart of this moment condition to zero we get:

\[ \hat{\beta}_{IV} = \frac{\sum Z_i Y_i}{\sum Z_i X_i} \text{ which is the same as the instrumental variables estimator.} \]

Note that if the variables are standardized, the denominator is the correlation between \( X \) and \( Z \) -- so the estimator is not defined (not identified) if the correlation is zero.

Two conditions for an instrumental variable:

1. Correlated with the variable that needs to be instrumented
2. Uncorrelated with the error term.

Now suppose that that \( X \) is \( N \times K \) and \( Z \) is \( N \times L \) and that \( L > K \) – we have more instruments available than regressors – over identification.
The model in matrix form is:

\[ Y = X\beta + \epsilon \quad \text{where} \quad E(X'\epsilon) \neq 0 \]

The moment conditions are:

\[ E(Z'\epsilon) = 0 \]

Since we have more moment conditions than parameters we could throw out extra moment conditions but this would not lead to a unique \( \hat{\beta} \).

So what is done is instead is to find that \( \beta \) that minimizes:

\[
\min[(Z'\epsilon)'W(Z'\epsilon)]
\]

Where \( W \) is an \( L \times L \) weighting matrix.

If you substitute for \( \epsilon \):

\[
\min[(Z'(Y - X\beta)'W(Z'(Y - X\beta))]
\]

The solution for the optimal \( \beta \) is:

\[
\hat{\beta} = (X'ZWZ'X)^{-1} X'ZWZY
\]
How do you choose $W$?

Using an identity matrix will do – however the estimator ignores useful information and is not optimal. However note that:

$$\text{var}(Z'\varepsilon) = \sigma^2_\varepsilon (Z'Z)$$

So if we choose $W = (Z'Z)^{-1}$, we give higher weight to moment conditions with lower variance (just like the GLS example above).

This leads to:

$$\hat{\beta}_{IV} = \left[(X'Z(Z'Z)^{-1}Z'X)^{-1}\right]X'Z(Z'Z)^{-1}Z'Y$$

Some books refer to this as “generalized instrumental variables” as opposed to GMM and save the term GMM for models with general functional forms.

It is easy to show that this estimator is the same as two stage least squares.

Now suppose that $\varepsilon \sim N(0, \Omega)$, then $\text{var}(Z'\varepsilon) = \sigma^2_\varepsilon (Z'\Omega Z)$

And:

$$\hat{\beta}_{IV} = \left[(X'Z(Z'\Omega Z)^{-1}Z'X)^{-1}\right]X'Z(Z'\Omega Z)^{-1}Z'Y$$
Important specification tests:

1. It is important to show that your excluded variables have explanatory power for the fight-hand-side endogenous variables.
2. It is important to show that the instruments are uncorrelated with the error term and that the exclusion restrictions are valid.
3. Important to show that there are endogenous explanatory variables in the model since GMM is inefficient relative to OLS if all variables are exogenous.

Leads to (for 2):

\[ H_0 : Z'\varepsilon = 0 \]
\[ H_a : Z'\varepsilon \neq 0 \]

This test is easy to implement in STATA – a post estimation command. Note that the model must be over identified for the test to work. If the model is exactly identified, then \( Z'\hat{\beta} = 0 \) regardless of whether \( Z \) contains endogenous variables or not.

Example using Data from Tanzania

```
regress idealnum  gradef  city agef  goodsan goodwat fpmess
Source |       SS       df       MS              Number of obs =    7175
-------------+------------------------------           F(  6,  7168) =  231.69
Model |  7935.78153     6  1322.63026           Prob > F      =  0.0000
Residual |  40918.6068  7168  5.70851099           R-squared     =  0.1624
-------------+------------------------------           Adj R-squared =  0.1617
Total |  48854.3883  7174   6.8099231           Root MSE      =  2.3892
------------------------------------------------------------------------------
idealnum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
gradef |  -.1657421    .009817   -16.88   0.000    -.1849863   -.1464978
city |  -.3628387    .096505    -3.76   0.000    -.5520169   -.1736604
agef |   .0470808   .0034304    13.72   0.000     .0403562    .0538055
goodsan |  -.4821974   .1976966    -2.44   0.015    -.8697411   -.0946538
goodwat |  -.6871424   .0697211    -9.86   0.000    -.8238164   -.5504684
fpmess |  -.3810917   .0590844    -6.45   0.000    -.4969145   -.2652689
   _cons |   6.006125    .120121    50.00   0.000     5.770652    6.241597     
------------------------------------------------------------------------------
```
ivregress 2sls idealnum gradef city agef goodsan goodwat (fpmess= radio tv lisradio)

Instrumental variables (2SLS) regression
Number of obs = 7101
Wald chi2(6) = 1230.52
Prob > chi2 = 0.0000
R-squared = 0.0573
Root MSE = 2.5343

------------------------------------------------------------------------------
| idealnum | Coef. | Std. Err. | z | P>|z| | 95% Conf. Interval |
|----------|-------|-----------|---|-----|-------------------|
| fpmess   | -2.160707 | .3515359 | -6.15 | 0.000 | -2.849705, -1.47171 |
| gradef   | -.1108485 | .0147842 | -7.50 | 0.000 | -.1398251, -.0818719 |
| city     | -.1508169 | .1110777 | -1.36 | 0.175 | -.3685252, .0668915 |
| agef     | .0616064 | .0045227 | 13.62 | 0.000 | .0527422, .0704706 |
| goodsan  | -.591477 | .2107448 | -2.81 | 0.005 | -1.004529, -.1784248 |
| goodwat  | -.5264795 | .0804746 | -6.54 | 0.000 | -.6842068, -.3687522 |
| _cons    | 6.05979 | .1291835 | 46.91 | 0.000 | 5.806595, 6.312985 |
------------------------------------------------------------------------------
Instrumented: fpmess
Instruments: gradef city agef goodsan goodwat radio tv lisradio

ivregress 2sls idealnum gradef city agef goodsan goodwat (fpmess= radio tv lisradio), robust

Instrumental variables (2SLS) regression
Number of obs = 7101
Wald chi2(6) = 1294.87
Prob > chi2 = 0.0000
R-squared = 0.0573
Root MSE = 2.5343

------------------------------------------------------------------------------
| idealnum | Coef. | Std. Err. | z | P>|z| | 95% Conf. Interval |
|----------|-------|-----------|---|-----|-------------------|
| fpmess   | -2.160707 | .3482724 | -6.20 | 0.000 | -2.843309, -1.478106 |
| gradef   | -.1108485 | .01455 | -7.62 | 0.000 | -.1393659, -.0823311 |
| city     | -.1508169 | .098868 | -1.53 | 0.127 | -.3445946, .0429609 |
| agef     | .0616064 | .0046162 | 13.35 | 0.000 | .0525588, .070654 |
| goodsan  | -.591477 | .1483923 | -3.99 | 0.000 | -.8823206, -.3006335 |
| goodwat  | -.5264795 | .0800936 | -6.57 | 0.000 | -.68346, -.369499 |
| _cons    | 6.05979 | .133223 | 45.49 | 0.000 | 5.798678, 6.320903 |
------------------------------------------------------------------------------
Instrumented: fpmess
Instruments: gradef city agef goodsan goodwat radio tv lisradio

estat endogenous

Tests of endogeneity
Ho: variables are exogenous

Robust score chi2(1) = 31.232 (p = 0.0000)
Robust regression F(1,7093) = 31.4789 (p = 0.0000)
ivregress gmm idealnum gradef city agef goodsan goodwat (fpmess= radio tv lisradio)

Instrumental variables (GMM) regression

Number of obs = 7101
Wald chi2(6) = 1296.23
Prob > chi2 = 0.0000
R-squared = 0.0572

GMM weight matrix: Robust

Root MSE = 2.5345

|               Robust
|              Coef. Std. Err.  z   P>|z|   [95% Conf. Interval]
-------------+--------------------------------------------------------
idealnum | -2.161912   .3479156    -6.21   0.000   -2.843814   -1.480009
fpmess   | -.1109999   .0145499    -7.63   0.000   -.1395171   -.0824827
gradef   | -.155256   .0986711   -1.57   0.116   -.3486477   .0381358
city     | .0613938   .0046154   13.30   0.000   .0523479   .0704398
agef     | -.6016565   .1483434    -4.06   0.000   -.8924043   -.3109088
goodsan  | -.5241035   .080061    -6.55   0.000   -.6810202   -.3671868
goodwat  | 6.066762   .1331309   45.57   0.000   5.80583    6.327693
_cons    | 6.155256   .0986711   -1.57   0.116   -.3486477   .0381358
-------------+--------------------------------------------------------

Instruments for equation 1: gradef city agef goodsan goodwat radio tv lisradio

estat overid

Test of overidentifying restriction:

Hansen's J chi2(2) = 4.95188 (p = 0.0841)

gmm (idealnum-{b0}-{b1}*fpmess-{b2}*gradef-{b3}*city-{b4}*agef-{b5}*goodwat-(b6)*goodsan), instruments(gradef city agef goodsan goodwat radio tv lisradio)

GMM estimation

Number of parameters = 7
Number of moments = 9
Initial weight matrix: Unadjusted
Number of obs = 7101
GMM weight matrix: Robust

|              Coef. Std. Err.  z   P>|z|   [95% Conf. Interval]
-------------+--------------------------------------------------------
/b0 | 6.066762   .1331309   45.57   0.000   5.80583    6.327693
/b1 | -2.161912   .3479156    -6.21   0.000   -2.843814   -1.480009
/b2 | -.1109999   .0145499    -7.63   0.000   -.1395171   -.0824827
/b3 | -.155256   .0986711   -1.57   0.116   -.3486477   .0381358
/b4 | .0613938   .0046154   13.30   0.000   .0523479   .0704398
/b5 | -.6016565   .1483434    -4.06   0.000   -.8924043   -.3109088
/b6 | -.5241035   .080061    -6.55   0.000   -.6810202   -.3671868
_cons | 6.066762   .1331309   45.57   0.000   5.80583    6.327693
-------------+--------------------------------------------------------

Instruments for equation 1: gradef city agef goodsan goodwat radio tv lisradio _cons

estat overid

Test of overidentifying restriction:

Hansen's J chi2(2) = 4.95188 (p = 0.0841)
Extension to Limited Dependent Variables

Poisson Regression Model

Useful for count data:

Number of births
Number of deaths
Number of FP facilities in a community

Estimation method is maximum likelihood:

\[ P(Y_i = y_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i !} \]

Where:

\[ \lambda_i = e^{X_i \beta} \]

And

\[ E(y_i) = \lambda_i = e^{X_i \beta} \]

So we can define moment conditions:

\[ \sum_{i=1}^{N} (y_i - e^{X_i \beta}) X_i ' = 0 \]

In this case, it is easy to show that the first order conditions for the maximum likelihood estimator are:
\[
\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^{N} (y_i - e^{X_i \beta}) X_i' = 0
\]

Which means that maximum likelihood and method of moments will be the same in this case.

If we have endogenous regressors:

\[
\sum_{i=1}^{N} (y_i - e^{X_i \beta}) Z_i' = 0
\]

Example using Data from Tunisia:

<table>
<thead>
<tr>
<th>idealnum</th>
<th>Freq.</th>
<th>Percent</th>
<th>Cum.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>1</td>
<td>80</td>
<td>2.16</td>
<td>2.22</td>
</tr>
<tr>
<td>2</td>
<td>861</td>
<td>23.28</td>
<td>25.49</td>
</tr>
<tr>
<td>3</td>
<td>982</td>
<td>26.55</td>
<td>52.04</td>
</tr>
<tr>
<td>4</td>
<td>1,265</td>
<td>34.20</td>
<td>86.24</td>
</tr>
<tr>
<td>5</td>
<td>229</td>
<td>6.19</td>
<td>92.43</td>
</tr>
<tr>
<td>6</td>
<td>280</td>
<td>7.57</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Total | 3,699 | 100.00 |

Standard Maximum Likelihood

```
poisson idealnum births urban age15_25 age26_30 age31_35 educf dhspfp5
dclnf210 dclnf35
```

Poisson regression

Number of obs = 3699  
LR chi2(9) = 339.89  
Prob > chi2 = 0.0000  

Log likelihood = -6261.6451  
Pseudo R2 = 0.0264

| idealnum | Coef.  | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|-----------|--------|-----------|-------|------|----------------------|
| births    | .045359| .0044232  | 10.25 | 0.000| .0366898 .0540282    |
| urban     | -.12156| .0238621  | -5.09 | 0.000| -.1683284 -.0747908  |
| age15_25  | .070940| .0331187  | 2.14  | 0.032| .0060294 .1358525    |
| age26_30  | .076045| .0285344  | 2.67  | 0.008| .0201184 .1319712    |
| age31_35  | .061751| .0247624  | 2.49  | 0.013| .0132173 .110284     |
| educf     | -.008356| .0027712 | -3.02 | 0.003| -.0137881 -.0029251  |
| dhspfp5   | -.038812| .0236606 | -1.64 | 0.101| -.0851862 .0075618   |
| dclnf210  | -.019130| .0208825 | -0.92 | 0.360| -.0600598 .0217981   |
| dclnf35   | -.035497| .0203251 | -1.75 | 0.081| -.0753334 .0043397   |
| _cons     | 1.142399| .0399497 | 28.60 | 0.000| 1.064099 1.220699    |
```
GMM with No Correction for Endogeneity

global xb
"{b1}*births+{b2}*urban+{b3}*age15_25+{b4}*age26_30+{b5}*age31_35+{b6}* educf+
{b7}*dhspfp5+{b8}*dclnf210+{b9}* dclnf35+{b0}"

gmm (idealnum-exp($xb)),instruments( births urban age15_25 age26_30 age31_35
educf dhspfp5 dclnf210 dclnf35)
warning: 268 missing values returned for equation 1 at initial values

GMM estimation

Number of parameters = 10
Number of moments = 10
Initial weight matrix: Unadjusted  Number of obs = 3699
GMM weight matrix: Robust

|                      | Coef.    | Std. Err. |     z  | P>|z|  | [95% Conf. Interval] |
|----------------------|----------|-----------|--------|------|---------------------|
| /b1                  | 0.045359 | 0.0025938 | 17.49  | 0.000| 0.0402753 0.0504427 |
| /b2                  | -0.12156 | 0.013797  | -8.81  | 0.000| -0.1486006 -0.0945187 |
| /b3                  | 0.070941 | 0.0185094 | 3.83   | 0.000| 0.0346633 0.1072186 |
| /b4                  | 0.076045 | 0.0161104 | 4.72   | 0.000| 0.0444689 0.1076207 |
| /b5                  | 0.061751 | 0.0143083 | 4.32   | 0.000| 0.033707 0.0897944 |
| /b6                  | -0.008357| 0.001387  | -5.63  | 0.000| -0.0112646 -0.0054486 |
| /b7                  | -0.038812| 0.0134263 | -2.89  | 0.004| -0.0651273 -0.012497 |
| /b8                  | -0.019131| 0.012083 | -1.58  | 0.113| -0.0428129 0.0045513 |
| /b9                  | -0.035497| 0.011773 | -3.01  | 0.003| -0.0585725 -0.0124211 |
| /b0                  | 1.142399 | 0.0233899 | 48.84  | 0.000| 1.096556 1.188243  |

Instruments for equation 1: births urban age15_25 age26_30 age31_35 educf
dhspfp5 dclnf210 dclnf35 _cons
GMM with Instruments for Births

\[
\text{global } \mathbf{x}_b = (b_1) \text{births} + (b_2) \text{urban} + (b_3) \text{age15_25} + (b_4) \text{age26_30} + (b_5) \text{age31_35} + (b_6) \text{educf} + (b_7) \text{dhspfp5} + (b_8) \text{dclnf210} + (b_9) \text{dclnf35} + (b_0)
\]

\[
\text{gmm (idealnum-exp($\mathbf{x}_b$)), instruments(urban age15_25 age26_30 age31_35 educf hosp55 clnc355 clnc255 dhspfp5 dclnf210 dclnf35)}
\]

GMM estimation

Number of parameters = 10
Number of moments = 12
Initial weight matrix: Unadjusted
GMM weight matrix: Robust

|          | Coef.  | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|----------|--------|-----------|-------|------|----------------------|
| /b1      | 0.178753| 0.0775377 | 2.31  | 0.021 | 0.0267819, 0.3307241  |
| /b2      | -0.0445122| 0.0506145 | -0.88 | 0.379 | -0.1437149, 0.0546904 |
| /b3      | 0.7156728| 0.3993534 | 1.79  | 0.073 | -0.0670454, 1.498391  |
| /b4      | 0.5368065| 0.2891844 | 1.86  | 0.063 | -0.0299846, 1.103597  |
| /b5      | 0.3105718| 0.1603059 | 1.94  | 0.053 | -0.003622, 0.6247656  |
| /b6      | 0.0082527| 0.0103891 | 0.79  | 0.427 | -0.0121095, 0.0286149 |
| /b7      | 0.0024966| 0.0341734 | 0.07  | 0.942 | -0.064482, 0.0694752  |
| /b8      | -0.0051071| 0.0176397 | -0.29 | 0.772 | -0.0396804, 0.0294661 |
| /b9      | 0.0410424| 0.0505894 | 0.81  | 0.417 | -0.0581111, 0.1401959 |
| /b0      | 0.1108901| 0.6380012 | 0.17  | 0.862 | -1.139569, 1.361349   |

Instruments for equation 1: urban age15_25 age26_30 age31_35 educf hosp55 clnc355 clnc255 dhspfp5 dclnf210 dclnf35 _cons

\text{estat overid}

Test of overidentifying restriction:

Hansen's J \text{ chi2(2) } = 5.61925 \ (p = 0.0602)
Probit

The dependent variable is now dichotomous and the model takes the following form:

\[ Y_i^* = X_i \beta + \varepsilon_i \text{ where } \varepsilon_i \sim N(0,1) \]

The observed dependent variable is \( Y_i \) which is the sign for the latent variable \( Y_i^* \).

We can show that:

\[ E(Y_i) = \Phi(X_i \beta) \]

So we can define moment conditions:

\[ \sum_{i=1}^{N} (Y_i - \Phi(X_i \beta))X_i' = 0 \]

And that the first order conditions for the maximum likelihood estimator are:

\[ \frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^{N} \frac{(Y_i - \Phi(X_i \beta))\phi(X_i \beta)X_i'}{\Phi(X_i \beta)(1 - \Phi(X_i \beta))} = 0 \]

So method of moments and maximum likelihood will not be the same in this case. Unless we define the residual to be:

\[ \frac{(Y_i - \Phi(X_i \beta))\phi(X_i \beta)}{\Phi(X_i \beta)(1 - \Phi(X_i \beta))} \]
Which have been referred to as “generalized residuals” see Gourieroux, Monfort, Renault, and Trognon (Journal of Econometrics 1987)

If we have endogenous regressors and a set of instruments, we can define the following moment conditions:

\[
\sum_{i=1}^{N} (Y_i - \Phi(X_i \beta))Z_i' = 0
\]

And

\[
\sum_{i=1}^{N} \frac{(Y_i - \Phi(X_i \beta))\phi(X_i \beta)Z_i'}{\Phi(X_i \beta)(1 - \Phi(X_i \beta))} = 0
\]

Examples using Data from Zimbabwe:

Simple Probit

```
probit curuse ageyr educf city
Probit regression                                               Number of obs  =       4044
LR chi2(3)          =     123.15
Prob > chi2         =     0.0000
Log likelihood      = -2473.4264                       Pseudo R2      =     0.0243
------------------------------------------------------------------------------
curse |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ageyr |   .0247129   .0024259    10.19   0.000     .0199582    .0294676
educf |   .0218258   .0067044     3.26   0.001     .0086853    .0349662
city |   .1519386   .0461074     3.30   0.001     .0615698    .2423074
_cons |  -1.353872   .0936086   -14.46   0.000    -1.537342   -1.170403
------------------------------------------------------------------------------
```
Method of Moments

\textbf{global \( xb \) \( \{b0}+{b1}* \text{ageyr}+{b2}* \text{educf}+{b3}* \text{city} \)}
\textbf{global \( \Phi \) \text{normal}(\{xb\})}
\textbf{global \( \phi \) \text{normalden}(\{xb\})}

\textbf{gmm (curuse-$\Phi$),instruments( ageyr educf city)}

\textbf{Step 1}

\textbf{GMM estimation}

Number of parameters = 4
Number of moments = 4
Initial weight matrix: Unadjusted
GMM weight matrix: Robust

<table>
<thead>
<tr>
<th>( b0 )</th>
<th>( b1 )</th>
<th>( b2 )</th>
<th>( b3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.33728</td>
<td>0.0240179</td>
<td>0.0226185</td>
<td>0.1515532</td>
</tr>
<tr>
<td>0.0917991</td>
<td>0.0022603</td>
<td>0.0069999</td>
<td>0.0458515</td>
</tr>
<tr>
<td>-14.57</td>
<td>10.63</td>
<td>3.23</td>
<td>3.31</td>
</tr>
<tr>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>-1.517203 -1.157357</td>
<td>0.0195878 0.0284481</td>
<td>0.0088989 0.0363381</td>
<td>0.0616859 0.2414204</td>
</tr>
</tbody>
</table>

Instruments for equation 1: ageyr educf city _cons

Method of Moments with Generalized Residuals

\textbf{gmm (curuse*$\phi$/\($\Phi$-(1-curuse)*$\phi$/\(1-\Phi\)),instruments( ageyr educf city)}

\textbf{GMM estimation}

Number of parameters = 4
Number of moments = 4
Initial weight matrix: Unadjusted
GMM weight matrix: Robust

<table>
<thead>
<tr>
<th>( b0 )</th>
<th>( b1 )</th>
<th>( b2 )</th>
<th>( b3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.353872</td>
<td>0.0247129</td>
<td>0.0218258</td>
<td>0.1519386</td>
</tr>
<tr>
<td>0.0919552</td>
<td>0.0022603</td>
<td>0.0069999</td>
<td>0.0458515</td>
</tr>
<tr>
<td>-14.72</td>
<td>10.63</td>
<td>3.23</td>
<td>3.31</td>
</tr>
<tr>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>-1.534101 -1.173643</td>
<td>0.020199 0.0292269</td>
<td>0.0083076 0.0353439</td>
<td>0.0619983 0.2418789</td>
</tr>
</tbody>
</table>

Instruments for equation 1: ageyr educf city _cons
Model with Endogenous Regressor

Simple Probit

probit curuse ageyr educf city idealnum

Probit regression
Number of obs  =  3432
LR chi2(4)     =  129.53
Prob > chi2    =  0.0000
Pseudo R2      =  0.0290

------------------------------------------------------------------------------
curse |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ageyr |  .0275553   .0027264    10.11   0.000     .0222116     .032899
educf |  .0158365   .0075078     2.11   0.035     .0011215    .0305514
city |  .1585321   .0498385     3.18   0.001     .0608505    .2562138
idealnum | -.0093307   .0112343    -0.83   0.406    -.0313495    .0126881
   _cons | -1.251713   .1103242   -11.35   0.000    -1.467944   -1.035482
------------------------------------------------------------------------------

Joint Estimation of Two Equations by Maximum Likelihood

ivprobit curuse ageyr educf city (idealnum= goodwat goodsan),first

Probit model with endogenous regressors
Number of obs  =  3432
Wald chi2(4)    =   536.69
Prob > chi2     =  0.0000

Log likelihood = -9314.1061

------------------------------------------------------------------------------
curse |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
idealnum | -.4120529   .0420308    -9.80   0.000    -.4944318   -.3296739
ageyr |  .0470123   .0023144    20.31   0.000     .0424763    .0515484
educf | -.0573738   .0107820    -5.32   0.000    -.0785061   -.0362415
city | -.1813323   .0631177    -2.87   0.004    -.3050408   -.0576238
   _cons |  .9205355   .3242921     2.84   0.005     .2849347    1.556136
-------------+----------------------------------------------------------------
idealnum     |
   ageyr |  .0742471   .0038626    19.22   0.000     .0666766    .0818175
   educf | -.1636876   .0111069   -14.74   0.000    -.1854567   -.1419186
city | -.2835769   .1141850    -2.55   0.011    -.5019533    -.0652006
goodwat | -.4396987   .0916246    -4.80   0.000    -.6192796   -.2601206
goodsan | -.1043804   .0716666    -1.46   0.145    -.2448443    .0360834
   _cons |  4.255872   .1497554    28.42   0.000     3.962356    4.549387
-------------+----------------------------------------------------------------
   /athrho |  1.093695   .2272441     4.81   0.000     .6483046    1.539085
   /lnsigma |  .6675622   .1207199     55.30   0.000     .4439017    .8912226
-------------+----------------------------------------------------------------
rho |  .7982227   .0824534     9.75   0.000     .6344284    .9620170
sigma |  1.949479   .0235339     83.80   0.000     1.903895    1.996155
-------------+----------------------------------------------------------------
Instrumented:  idealnum
Instruments:  ageyr educf city goodwat goodsan

Wald test of exogeneity (/athrho = 0): chi2(1) =  23.16  Prob > chi2 = 0.0000
Method of Moments Using Generalized Residuals

\begin{verbatim}
global xb "(b0)+(b1)* ageyr+(b2)* educf+(b3)* city+(b4)*idealnum"
global Phi "normal($xb)"
global phi "normalden($xb)"
gmm (curuse*$phi/$Phi-(1-curuse)*$phi/(1-$Phi)),instruments( ageyr educf city goodwat goodsan)
\end{verbatim}

GMM estimation

Number of parameters = 5
Number of moments = 6
Initial weight matrix: Unadjusted
GMM weight matrix: Robust

<table>
<thead>
<tr>
<th></th>
<th>Robust</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
<td>Std. Err.</td>
<td>z</td>
</tr>
<tr>
<td>----------------------</td>
<td>--------</td>
<td>-----------</td>
<td>------</td>
</tr>
<tr>
<td>/b0</td>
<td>1.925231</td>
<td>1.660998</td>
<td>1.16</td>
</tr>
<tr>
<td>/b1</td>
<td>0.0921017</td>
<td>0.0373101</td>
<td>2.47</td>
</tr>
<tr>
<td>/b2</td>
<td>-0.0956154</td>
<td>0.0563338</td>
<td>-1.70</td>
</tr>
<tr>
<td>/b3</td>
<td>-0.3823969</td>
<td>0.2995131</td>
<td>-1.28</td>
</tr>
<tr>
<td>/b4</td>
<td>-0.8756558</td>
<td>0.4847308</td>
<td>-1.81</td>
</tr>
</tbody>
</table>

Instruments for equation 1: ageyr educf city goodwat goodsan _cons

\texttt{estat overid}

Test of overidentifying restriction:

\texttt{Hansen's J chi2(1) = .439423 (p = 0.5074)}
Logit

We can show that:

\[ E(Y_i) = \frac{e^{X_i\beta}}{1 + e^{X_i\beta}} \]

So we can define moment conditions:

\[ \sum_{i=1}^{N} (Y_i - E(Y_i))X_i' = 0 \]

Which are the same as the first order conditions for the maximum likelihood estimator.

If we have endogenous regressors, we can define the following moment conditions:

\[ \sum_{i=1}^{N} (Y_i - E(Y_i))Z_i' = 0 \]

Example using data from Tanzania:

Simple Logit

logit curuse agef gradef city idealnum

<table>
<thead>
<tr>
<th>Logistic regression</th>
<th>Number of obs = 6420</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR chi2(4) = 293.40</td>
<td>Prob &gt; chi2 = 0.0000</td>
</tr>
<tr>
<td>Log likelihood = -1839.6667</td>
<td>Pseudo R2 = 0.0739</td>
</tr>
</tbody>
</table>

| curuse | Coef. | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|--------|-------|-----------|-------|-----|----------------------|
| agef   | .0441804 | .0053186 | 8.31  | 0.000 | .0337561 | .0546046 |
| gradef | .205559  | .0161523 | 12.73 | 0.000 | .1739011 | .2372169 |
| city   | .5334482 | .1087663 | 4.90  | 0.000 | .3202702 | .7466263 |
| idealnum | -.0521602 | .0209566 | -2.49 | 0.013 | -.0932343 | -.0110861 |
| _cons  | -4.355425 | .240694 | -18.10 | 0.000 | -4.827177 | -3.883674 |
GMM with Exogenous Regressors

global xb "(b0)+(b1)* agef+(b2)* gradef+(b3)* city+(b4)*idealnum"
global ey "exp($xb)/(1+exp($xb))"
gmm (curuse-$ey), instruments( agef gradef city idealnum)

GMM estimation

Number of parameters  =  5  
Number of moments     =  5
Initial weight matrix: Unadjusted                     Number of obs  =    6420
GMM weight matrix:     Robust

--------------------------------------------------------------------
|        | Coef.  | Robust Std. Err. |     z  |   P>|z|  [95% Conf. Interval] |
|        |       |                   |        |      |               |               |
|/b0     |  -4.355425 |    .2311157 |    -18.85 | 0.000 |   -4.808404 |   -3.902447 |
|/b1     |   .0441804 |    .0047281 |      9.34 | 0.000 |     .0349135 |     .0534472 |
|/b2     |  .205559  |     .01677  |     12.26 | 0.000 |    .1726903 |    .2384276  |
|/b3     |  .5334482 |     .1104231|      4.83 | 0.000 |    .317023  |    .7498735  |
|/b4     |  -.0521602|     .0207588|     -2.51 | 0.012 |   -.0928467|   -.0114737  |
--------------------------------------------------------------------

Instruments for equation 1: agef gradef city idealnum _cons

GMM with Idealnum Endogenous

global xb "(b0)+(b1)* agef+(b2)* gradef+(b3)* city+(b4)*idealnum"
global ey "exp($xb)/(1+exp($xb))"
gmm (curuse-$ey), instruments( agef gradef city goodsan goodwat)

GMM estimation

Number of parameters  =  5  
Number of moments     =  6
Initial weight matrix: Unadjusted                     Number of obs  =    6420
GMM weight matrix:     Robust

--------------------------------------------------------------------
|        | Coef.  | Robust Std. Err. |     z  |   P>|z|  [95% Conf. Interval] |
|        |       |                   |        |      |               |               |
|/b0     |  -.7549698|    2.294859 |    -0.33 | 0.742 |   -5.25281  |    3.742871 |
|/b1     |   .0724285|    .0256051 |      2.83 | 0.005 |     .0222435|    .1226135  |
|/b2     |  .1431967|     .0205348|      6.97 | 0.000 |     .1029493|    .1834441  |
|/b3     |  .1939602|     .201301 |      0.96 | 0.335 |   -.2005824 |    .5885028  |
|/b4     |  -.9054825|     .7181797|     -1.26 | 0.207 |  -2.313089  |    .5021237  |
--------------------------------------------------------------------

Instruments for equation 1: agef gradef city goodsan goodwat _cons

estat overid

Test of overidentifying restriction:

Hansen's J chi2(1) = .153577 (p = 0.6951)
Extensions

Seemingly Unrelated Regression Model

Consider a two equation model:

\[ Y_{i1} = X_{i1}\beta_1 + \varepsilon_{i1} \]
\[ Y_{i2} = X_{i2}\beta_2 + \varepsilon_{i2} \]

Where we assume that \( E(\varepsilon_{i1}\varepsilon_{i2}) = \sigma_{i2} \) -- contemporaneous correlation between the errors of the two equations. For example if these are two equations for an individual, we assume that there is overlap in the unobservables affected the two outcomes. The equations are seemingly unrelated because the the association is through unobservables.

Note that if \( X_{i1} = X_{i2} \), then OLS yields the same results as the seemingly unrelated regression estimator.
Example for Tunisia:

```plaintext
sureg (idealnum urban age15_25 age26_30 age31_35 educf dhspfp5 dclnf35) (births urban age15_25 age26_30 age31_35 educf hosp55 c > lnc255 clnc355), corr

Seemingly unrelated regression

<table>
<thead>
<tr>
<th></th>
<th>Obs</th>
<th>Parms</th>
<th>RMSE</th>
<th>&quot;R-sq&quot;</th>
<th>chi2</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>idealnum</td>
<td>3699</td>
<td>8</td>
<td>1.098898</td>
<td>0.1511</td>
<td>646.08</td>
<td>0.0000</td>
</tr>
<tr>
<td>births</td>
<td>3699</td>
<td>8</td>
<td>1.966098</td>
<td>0.4528</td>
<td>3057.12</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

| Equation       | Coef. | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|----------------|-------|-----------|------|------|-----------------------|
| idealnum       |       |           |      |      |                       |
| urban | -.5341832 | .04836 | -11.05 | 0.000 | -.6289671 | -.4393993 |
| age15_25 | -.4340123 | .052471 | -8.27 | 0.000 | -.5368536 | -.331171 |
| age26_30 | -.2219301 | .0501645 | -4.42 | 0.000 | -.3202506 | -.1236096 |
| age31_35 | -.0421442 | .0487635 | -0.86 | 0.387 | -.1377188 | .0534304 |
| educf | -.0426635 | .0051863 | -8.23 | 0.000 | -.0528284 | -.0324986 |
| dhspfp5 | -.1439378 | .0444092 | -3.24 | 0.001 | -.2309783 | -.0568973 |
| dclnf210 | -.0822242 | .0412718 | -1.99 | 0.046 | -.1631155 | -.0013329 |
| dclnf35 | -.1438506 | .0401266 | -3.58 | 0.000 | -.2224973 | -.065204 |
| _cons | 4.208579 | .0503266 | 83.63 | 0.000 | 4.109941 | 4.307217 |
| births       |       |           |      |      |                       |
| urban | -.472903 | .0929165 | -5.09 | 0.000 | -.6550159 | -.2907901 |
| age15_25 | -4.146174 | .0937669 | -44.22 | 0.000 | -.6329954 | -.3562394 |
| age26_30 | -2.875245 | .089736 | -32.04 | 0.000 | -.3051124 | -.2618966 |
| age31_35 | -1.491799 | .0872567 | -17.10 | 0.000 | -.1662819 | -.1320778 |
| educf | -.1171785 | .0092421 | -12.68 | 0.000 | -.1352927 | -.0990642 |
| hosp55 | -.2233258 | .1038585 | -2.15 | 0.032 | -.4268846 | -.019767 |
| clnc255 | -.0075894 | .1029784 | -0.07 | 0.941 | -.2094241 | .1942452 |
| clnc355 | -.223498 | .0657493 | -3.40 | 0.001 | -.3523643 | -.0946318 |
| _cons | 6.543402 | .0790779 | 82.75 | 0.000 | 6.388412 | 6.698392 |

Correlation matrix of residuals:

<table>
<thead>
<tr>
<th></th>
<th>idealnum</th>
<th>births</th>
</tr>
</thead>
<tbody>
<tr>
<td>idealnum</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>births</td>
<td>0.2901</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Breusch-Pagan test of independence: chi2(1) = 311.228, Pr = 0.0000
GMM

```
gmm (eq1: idealnum - (b0)-(xb1:urban age15_25 age26_30 age31_35 educf dhspfp5 dclnf210 dclnf35)) (eq2:births- (c0)-(xb2: urban age15_25 age
> 26_30 age31_35  educf dhspfp5 dclnf210 dclnf35)) instruments(eq1:urban age15_25 age26_30 age31_35 educf dhspfp5 dclnf210 dclnf35) instruments (eq2:urban age15_25 age26_
> 30 age31_35 educf hosp55 clnc255 clnc355) winitial(unadjusted,independent) wmatrix(unadjusted) twostep;

GMM estimation
Number of parameters = 18
Number of moments = 18
Initial weight matrix: Unadjusted Number of obs = 3699
GMM weight matrix: Unadjusted

|     | Coef.  | Std. Err. | z     | P>|z|    | [95% Conf. Interval] |
|-----|--------|-----------|-------|--------|---------------------|
| /b0 | 4.259347 | 0.0512122 | 83.17 | 0.000  | 4.158973 - 4.359721 |
| /xb1_urban | -0.5224261 | 0.0487445 | -10.72 | 0.000  | -0.6179636 - -0.4268885 |
| /xb1_age1-25 | -0.43919 | 0.0524799 | -8.37 | 0.000  | -0.5420487 - -0.3363313 |
| /xb1_age2-30 | -0.225901 | 0.050169 | -4.50 | 0.000  | -0.3242304 - -0.1275715 |
| /xb1_age3-35 | -0.044741 | 0.0487651 | -0.91 | 0.362  | -0.1400519 - 0.051037 |
| /xb1_educf | -0.0413129 | 0.0051912 | -7.96 | 0.000  | -0.0514876 - -0.0311383 |
| /xb1_dhspfp5 | -0.1716952 | 0.0458525 | -3.74 | 0.000  | -0.2615644 - -0.0818261 |
| /xb1_dclnf210 | -0.0878041 | 0.0426432 | -2.06 | 0.039  | -0.1713832 - -0.004225 |
| /xb1_dclnf35 | -0.209072 | 0.0417471 | -5.01 | 0.000  | -0.2908949 - -0.1272492 |
| /c0 | 6.552747 | 0.0798257 | 82.09 | 0.000  | 6.396292 - 6.709203 |
| /xb2_urban | -0.4345315 | 0.0938955 | -4.63 | 0.000  | -0.6185632 - -0.2504998 |
| /xb2_age1-25 | -4.146713 | 0.0937755 | -44.22 | 0.000  | -4.33051 - -3.962916 |
| /xb2_age2-30 | -2.877945 | 0.089742 | -32.07 | 0.000  | -3.053836 - -2.702054 |
| /xb2_age3-35 | -1.493905 | 0.0872593 | -17.12 | 0.000  | -1.66493 - -1.322879 |
| /xb2_educf | -1.163168 | 0.092479 | -12.58 | 0.000  | -1.344424 - -0.981912 |
| /xb2_hosp55 | -0.3579445 | 0.1077349 | -3.22 | 0.001  | -0.569101 - -0.146788 |
| /xb2_clnc255 | 0.0673702 | 0.0682729 | 0.63 | 0.528  | 0.0528 - 0.1418735 |
| /xb2_clnc355 | -0.234723 | 0.0682729 | -3.43 | 0.001  | -0.3681847 - -0.100559 |

Instruments for equation 1: urban age15_25 age26_30 age31_35 educf dhspfp5 dclnf210 dclnf35 _cons
Instruments for equation 2: urban age15_25 age26_30 age31_35 educf hosp55 clnc255 clnc355 _cons

26
Panel Data Models

STATA has separate packages for instrumental variables regressions for panel data models xtivreg and one for dynamic panel data models – xtdpdsys.

Topic for an entire talk. However, there are a couple of interesting points:

Instrumental Variables and Panel Data:

Consider a simple panel data model with only time varying regressors:

\[ Y_{ti} = X_{ti} \beta + \mu_i + \epsilon_{ti} \]

Where there are K regressors and \( \beta \) is K x 1. Now suppose that:

\[ E(X_{ti} \epsilon_{ti}) \neq 0 \text{ and } E(X_{ti} \mu_i) \neq 0 \]

So OLS, random effects, fixed effects, and first differencing are not consistent estimators.

However, suppose a set of L instruments exists \( Z_{ti} \) that are uncorrelated with the errors but correlated with \( X_{ti} \) then instrumental variables methods can be used. On the surface, it appears that identification requires \( L \geq K \) for identification but this is not necessarily the case. To see this, suppose \( T=2 \):

\[ Y_{1i} = X_{1i} \beta + \mu_i + \epsilon_{1i} \]

\[ Y_{2i} = X_{2i} \beta + \mu_i + \epsilon_{2i} \]
We can think of this as two equations where we have restricted the effect of the explanatory variables to be the same. Now $Z_{2i}$ is an instrument for $X_{2i}$ -- however – so is $Z_{1i}$. In addition, $Z$’s from both time periods are valid instruments $X_{li}$ (assuming strong exogeneity of the $Z$’s).

Note that as we add time periods, the number of $\beta$’s to be estimated stays the same (unless we start adding time interactions, for example) and so the level of over identification increases. Higher levels of identification typically increase the efficiency of the resulting estimator.

Leads to what is referred to as “GMM style” instrument set.
Example:

“Standard identification” or sometimes called summation assumption:

Let $\eta_{ti} = \mu_i + \varepsilon_{ti}$

\[
\eta_i = \begin{bmatrix}
\eta_{1i} \\
\eta_{2i} \\
\vdots \\
\eta_{Ti}
\end{bmatrix} \quad \text{and} \quad Z_i = \begin{bmatrix}
Z'_{1i} \\
Z'_{2i} \\
\vdots \\
Z'_{Ti}
\end{bmatrix}
\]

Then $E(Z'_i \eta_i) = 0$ requires that $L \geq K$ for identification.

Contemporaneous exogeneity assumption redefines:

\[
Z_i = \begin{bmatrix}
Z'_{1i} & 0 & \ldots & 0 \\
0 & Z'_{2i} \\
\vdots & \ddots & \ddots & \vdots \\
0 & \ldots & 0 & Z'_{Ti}
\end{bmatrix}
\]
Which means that we have TxL orthogonality conditions – but we can get even more if we go to strong endogeneity (Z’s and η’s from different time periods orthogonal)