Generalized method of moments estimation in Stata 11

David M. Drukker

StataCorp

Stata Conference
Washington, DC 2009

Outline

1. A quick introduction to GMM

2. `gmm` examples
   - Ordinary least squares
   - Two-stage least squares
   - Cross-sectional Poisson with endogenous covariates
   - Fixed-effects Poisson regression

Method of Moments (MM)

- We estimate the mean of a distribution by the sample mean, the variance by the sample variance, etc
- We want to estimate $\mu = E[y]$
  - We use $\widehat{\mu} = (1/N)\sum_{i=1}^{N} y_i$
  - This estimator has nice properties because it solves the sample moment condition

$$(1/N)\sum_{i=1}^{N}(y_i - \mu) = 0$$

    which is the sample analog of the population moment condition

$$E[y - \mu] = 0$$

- Estimators that solve sample moment equations to produce estimates are called method-of-moments (MM) estimators
  - This method dates back to Pearson (1895)

Generalized method-of-moments (GMM)

- The MM only works when the number of moment conditions equals the number of parameters to estimate
    - If there are more moment conditions than parameters, the system of equations is algebraically over identified and cannot be solved
    - Generalized method-of-moments (GMM) estimators choose the estimates that minimize a quadratic form of the sample moment conditions
        - GMM gets as close to solving the over-identified system of sample moment equations as possible
        - GMM reduces to MM when the number of parameters equals the number of moment conditions
- Hansen (1982) produced many of the key results; Wooldridge (2002); Cameron and Trivedi (2005) provide good introductions

Definition of GMM estimator

- Our research question implies q population moment conditions

$$E[\mathbf{m}(\mathbf{w}_i, \boldsymbol{\theta})] = \mathbf{0}$$

  - $\mathbf{m}$ is $q \times 1$ vector of functions whose expected values are zero in the population
  - $\mathbf{w}_i$ is the data on person i
  - $\boldsymbol{\theta}$ is $k \times 1$ vector of parameters, $k \leq q$

- The sample moments that correspond to the population moments are

$$\overline{\mathbf{m}}(\boldsymbol{\theta}) = (1/N) \sum_{i=1}^{N} \mathbf{m}(\mathbf{w}_i, \boldsymbol{\theta})$$

- When $k < q$, GMM chooses the parameters that are as close as possible to solving the over-identified system of moment equations

$$\widehat{\boldsymbol{\theta}}_{GMM} \equiv \arg\min_{\boldsymbol{\theta}} \quad \overline{\mathbf{m}}(\boldsymbol{\theta})'\mathbf{W}\overline{\mathbf{m}}(\boldsymbol{\theta})$$

Some properties of the GMM estimator

$$\widehat{\boldsymbol{\theta}}_{GMM} \equiv \arg\min_{\boldsymbol{\theta}} \quad \overline{\mathbf{m}}(\boldsymbol{\theta})'\mathbf{W}\overline{\mathbf{m}}(\boldsymbol{\theta})$$

- When $k = q$, the MM estimator solves $\overline{\mathbf{m}}(\boldsymbol{\theta})$ exactly so $\overline{\mathbf{m}}(\theta)'\mathbf{W}\overline{\mathbf{m}}(\theta) = \mathbf{0}$
- $\mathbf{W}$ only affects the efficiency of the GMM estimator
  - Setting $\mathbf{W} = \mathbf{I}$ yields consistent, but inefficient estimates
  - Setting $\mathbf{W} = \text{Cov}[\overline{\mathbf{m}}(\theta)]^{-1}$ yields an efficient GMM estimator
  - We can take multiple steps to get an efficient GMM estimator
    1. Let $\mathbf{W} = \mathbf{I}$ and get
    $$\widehat{\boldsymbol{\theta}}_{GMM1} \equiv \arg\min_{\boldsymbol{\theta}} \quad \overline{\mathbf{m}}(\boldsymbol{\theta})'\overline{\mathbf{m}}(\boldsymbol{\theta})$$
    2. Use $\widehat{\boldsymbol{\theta}}_{GMM1}$ to get $\widehat{\mathbf{W}}$, which is an estimate of $\text{Cov}[\overline{\mathbf{m}}(\theta)]^{-1}$
    3. Get
    $$\widehat{\boldsymbol{\theta}}_{GMM2} \equiv \arg\min_{\boldsymbol{\theta}} \quad \overline{\mathbf{m}}(\boldsymbol{\theta})'\widehat{\mathbf{W}}\overline{\mathbf{m}}(\boldsymbol{\theta})$$
    4. Repeat steps 2 and 3 using $\widehat{\boldsymbol{\theta}}_{GMM2}$ in place of $\widehat{\boldsymbol{\theta}}_{GMM1}$

The gmm command

- The new command gmm estimates parameters by GMM
- gmm is similar to nl, you specify the sample moment conditions as substitutable expressions
- Substitutable expressions enclose the model parameters in braces {}

The interactive syntax of gmm

- For many models, the population moment conditions have the form

$$E[\mathbf{z}e(\beta)] = \mathbf{0}$$

where $\mathbf{z}$ is a $q \times 1$ vector of instrumental variables and $e(\beta)$ is a scalar function of the data and the parameters β

- The corresponding syntax of *gmm* is

gmm (*eb_expression*) [*if*] [*in*] [*weight*] ,
    instruments(*instrument_varlist*) [*options*]

where some options are

onestep	use one-step estimator (default is two-step estimator)
winitial(*wmtype*)	initial weight-matrix $\mathbf{W}$
wmatrix(*witype*)	weight-matrix $\mathbf{W}$ computation after first step
vce(*vcetype*)	*vcetype* may be robust, cluster, bootstrap, hac

Ordinary least squares (OLS) is an MM estimator

- We know that OLS estimates the parameters of the conditional expectation of $y_i = \mathbf{x}_i\boldsymbol{\beta} + \epsilon_i$ under the assumption that $E[\epsilon|\mathbf{x}] = 0$
- Standard probability theory implies that

$$E[\epsilon|\mathbf{x}] = 0 \Rightarrow E[\mathbf{x}\epsilon] = \mathbf{0}$$

  So the population moment conditions for OLS are

$$E[\mathbf{x}(y - \mathbf{x}\boldsymbol{\beta})] = \mathbf{0}$$

- The corresponding sample moment conditions are

$$(1/N)\sum_{i=1}^{N} \mathbf{x}_i(y_i - \mathbf{x}_i\boldsymbol{\beta}) = \mathbf{0}$$

  Solving for $\boldsymbol{\beta}$ yields

$$\widehat{\boldsymbol{\beta}}_{OLS} = \left(\sum_{i=1}^{N} \mathbf{x}_i'\mathbf{x}_i\right)^{-1} \sum_{i=1}^{N} \mathbf{x}_i' y_i$$

Modeling crime data I

- We have (fictional) data on crime in 3,000 communities

```
. use cscrime2, clear

. describe

Contains data from cscrime2.dta
  obs:         3,000
 vars:             5                          29 Jul 2009 12:02
 size:       132,000 (98.7% of memory free)   (_dta has notes)
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
policepc        double %10.0g                 police officers per thousand
arrestp         double %10.0g                 arrests/crimes
convictp        double %10.0g                 convictions/arrests
legalwage       double %10.0g                 legal wage index 0-20 scale
crime           double %10.0g                 property-crime index 0-50 scale
-------------------------------------------------------------------------------
Sorted by:
```

Modeling crime data II

- We specify that

$$\texttt{crime}_i = \texttt{policepc}_i\beta_1 + \texttt{legalwage}_i\beta_2 + \beta_3 + \epsilon_i$$

- We want to model

$$E[\texttt{crime}|\texttt{policepc}, \texttt{legalwage}] = \texttt{policepc}\beta_1 + \texttt{legalwage}\beta_2 + \beta_3$$

- If $E[\epsilon|\texttt{policepc}, \texttt{legalwage}] = 0$, the population moment conditions are

$$E\left[\begin{pmatrix}\texttt{policepc}\\ \texttt{legalwage}\end{pmatrix}(\texttt{crime} - \texttt{policepc}\beta_1 - \texttt{legalwage}\beta_2 - \beta_3)\right] = \begin{pmatrix}0\\ 0\end{pmatrix}$$

OLS by GMM I

```
. gmm (crime - policepc*{b1} - legalwage*{b2} - {b3}),          ///
>         instruments(policepc legalwage) nolog
Final GMM criterion Q(b) =  2.62e-31
GMM estimation
Number of parameters =   3
Number of moments    =   3
Initial weight matrix: Unadjusted                    Number of obs   =     3000
GMM weight matrix:     Robust
```

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
/b1	-.4226003	.0100658	-41.98	0.000	-.4423289	-.4028716
/b2	-7.543894	.3969104	-19.01	0.000	-8.321824	-6.765964
/b3	27.79852	.0546507	508.66	0.000	27.69141	27.90563

Instruments for equation 1: policepc legalwage _cons

OLS by GMM II

```
. regress crime policepc legalwage, robust
Linear regression                                    Number of obs =     3000
                                                     F(  2,   2997) = 1384.95
                                                     Prob > F      =   0.0000
                                                     R-squared     =   0.6217
                                                     Root MSE      =   1.7972
```

crime	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
policepc	-.4226003	.0100709	-41.96	0.000	-.4423468	-.4028538
legalwage	-7.543894	.397109	-19.00	0.000	-8.322528	-6.765261
_cons	27.79852	.054678	508.40	0.000	27.69131	27.90573

IV and 2SLS

- For some variables, the assumption $E[\epsilon|x] = 0$ is too strong and we need to allow for $E[\epsilon|x] \neq 0$
- If we have q variables $\mathbf{z}$ for which $E[\epsilon|\mathbf{z}] = \mathbf{0}$ and the correlation between $\mathbf{z}$ and $\mathbf{x}$ is sufficiently strong, we can estimate β from the population moment conditions

$$E[\mathbf{z}(y - \mathbf{x}\beta)] = \mathbf{0}$$

- $\mathbf{z}$ are known as instrumental variables
- If the number of variables in $\mathbf{z}$ and $\mathbf{x}$ is the same ($q = k$), solving the the sample moment conditions yields the MM estimator known as the instrumental variables (IV) estimator
- If there are more variables in $\mathbf{z}$ than in $\mathbf{x}$ ($q > k$) and we let $\mathbf{W} = \left(\sum_{i=1}^{N} \mathbf{z}_i'\mathbf{z}_i\right)^{-1}$ in our GMM estimator, we obtain the two-stage least-squares (2SLS) estimator

2SLS on crime data I

- The assumption that $E[\epsilon|\texttt{policepc}] = 0$ is false if communities increase `policepc` in response an increase in crime (an increase in ϵ_i)
- The variables `arrestp` and `convictp` are valid instruments, if they measure some components of communities' toughness-on crime that are unrelated to ϵ but are related to `policepc`
- We will continue to maintain that $E[\epsilon|\texttt{legalwage}] = 0$

2SLS by GMM I

```
. gmm (crime - policepc*{b1} - legalwage*{b2} - {b3}),          ///
>         instruments(arrestp convictp legalwage ) nolog onestep
Final GMM criterion Q(b) =  .0001736
GMM estimation
Number of parameters =   3
Number of moments    =   4
Initial weight matrix: Unadjusted                      Number of obs   =    3000
```

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
/b1	-.9516683	.0785137	-12.12	0.000	-1.105552	-.7977844
/b2	-2.304205	.9648523	-2.39	0.017	-4.195281	-.4131291
/b3	29.88578	.3135637	95.31	0.000	29.2712	30.50035

```
Instruments for equation 1: arrestp convictp legalwage _cons
```

2SLS by GMM II

```
. ivregress 2sls crime legalwage (policepc = arrestp convictp) , robust
Instrumental variables (2SLS) regression          Number of obs =     3000
                                                  Wald chi2(2)  =   696.63
                                                  Prob > chi2   =   0.0000
                                                  R-squared     =        .
                                                  Root MSE      =   3.0516
```

crime	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
policepc	-.9516683	.0785137	-12.12	0.000	-1.105552	-.7977844
legalwage	-2.304205	.9648523	-2.39	0.017	-4.195281	-.4131291
_cons	29.88578	.3135637	95.31	0.000	29.2712	30.50035

```
Instrumented:  policepc
Instruments:   legalwage arrestp
               convictp
```

Poisson with endogenous covariates

- We want to model to $E[y_i|\mathbf{x}_i, \nu_i] = \exp(\mathbf{x}_i\boldsymbol{\beta})\nu_i$
- This setup allows the distribution of ν_i to depend on $\mathbf{x}_i$
- Mullahy (1997) showed that we can use instrumental variables $\mathbf{z}_i$ and the population moment conditions

$$E\left[\mathbf{z}_i(y_i\exp(\mathbf{x}_i\boldsymbol{\beta}) - 1)\right] = \mathbf{0}$$

  to estimate $\boldsymbol{\beta}$

```
. use accident2, clear
. describe
Contains data from accident2.dta
  obs:           948
 vars:             6                        29 Jul 2009 11:59
 size:        26,544 (99.7% of memory free)
--------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
--------------------------------------------------------------------------------
kids            float  %9.0g
cvalue          float  %9.0g
tickets         float  %9.0g
traffic         float  %9.0g
male            float  %9.0g
accidents       float  %9.0g
--------------------------------------------------------------------------------
Sorted by:
```

- `traffic` and `male` are exogenous variables
- `tickets` is an endogenous variable
- `kids` and `cvalue` are instrumental variables

```
. gmm (accidents*exp(-tickets*{b1} - traffic*{b2} - male*{b3} - {b4}) - 1), ///
>     instruments(kids cvalue traffic male) onestep nolog
Final GMM criterion Q(b) =  .0109217
GMM estimation
Number of parameters =   4
Number of moments    =   5
Initial weight matrix: Unadjusted                    Number of obs  =     948
```

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
/b1	1.745919	.1984268	8.80	0.000	1.357009	2.134828
/b2	.1216527	.0421674	2.88	0.004	.0390061	.2042993
/b3	4.693161	.5129505	9.15	0.000	3.687797	5.698526
/b4	-11.51383	1.208924	-9.52	0.000	-13.88327	-9.144379

```
Instruments for equation 1: kids cvalue traffic male _cons
```

More complicated moment conditions

- The structure of the moment conditions for some models is too complicated to fit into the interactive syntax used thus far
- For example, Wooldridge (1999, 2002); Blundell, Griffith, and Windmeijer (2002) discuss estimating the fixed-effects Poisson model for panel data by GMM.
- In the Poisson panel-data model we are modeling

$$E[y_{it}|\mathbf{x}_{it},\eta_i] = \exp(\mathbf{x}_{it}\boldsymbol{\beta} + \eta_i)$$

- Hausman, Hall, and Griliches (1984) derived a conditional log-likelihood function when the outcome is assumed to come from a Poisson distribution with mean $\exp(\mathbf{x}_{it}\boldsymbol{\beta} + \eta_i)$ and η_i is an observed component that is correlated with the $\mathbf{x}_{it}$

- Wooldridge (1999) showed that you could estimate the parameters of this model by solving the sample moment equations

$$\sum_i \sum_t \mathbf{x}_{it} \left(y_{it} - \mu_{it} \frac{\overline{y}_i}{\overline{\mu}_i} \right) = \mathbf{0}$$

- These moment conditions do not fit into the interactive syntax because the term $\overline{\mu}_i$ depends on the parameters
- Need to use moment-evaluator program syntax

Moment-evaluator program syntax

- An abbreviated form of the program syntax for gmm is

gmm *moment_program* [*if*] [*in*] [*weight*] ,

equations(*moment_cond_names*)

parameters(*parameter_names*)

[instruments() options]

- The *moment_program* is an ado-file of the form

```
program gmm_eval
     version 11
     syntax varlist if, at(name)
     quietly {
       <replace elements of varlist with error
        part of moment conditions>
     }
end
```

Panel Accident data

```
. use xtaccidents

. describe

Contains data from xtaccidents.dta
  obs:         5,000
 vars:             7                          31 May 2008 19:50
 size:       160,000 (98.5% of memory free)
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
id              float  %9.0g
male            float  %9.0g
t               float  %9.0g
kids            float  %9.0g
cvalue          float  %9.0g
tickets         float  %9.0g
accidents       float  %9.0g
-------------------------------------------------------------------------------
Sorted by:  id  t

. by id: egen max_a = max(accidents )

. drop if max_a ==0
(3750 observations deleted)
```

```
program xtfe
    version 11
    syntax varlist if, at(name)
    quietly {
        tempvar mu mubar ybar
        generate double `mu' = exp(kids*`at'[1,1]    ///
            + cvalue*`at'[1,2]                       ///
            + tickets*`at'[1,3]) `if'
        egen double `mubar'  = mean(`mu') `if', by(id)
        egen double `ybar'   = mean(accidents) `if', by(id)
        replace `varlist'    = accidents             ///
                             - `mu'*`ybar'/`mubar' `if'
    }
end
```

FE Poisson by gmm

```
. gmm xtfe , equations(accidents) parameters(kids cvalue tickets)     ///
>         instruments(kids cvalue tickets, noconstant)                ///
>         vce(cluster id) onestep nolog
Final GMM criterion Q(b) =  1.50e-16
GMM estimation
Number of parameters =   3
Number of moments    =   3
Initial weight matrix: Unadjusted                    Number of obs   =    1250
                                         (Std. Err. adjusted for 250 clusters in id)
```

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
/kids	-.4506245	.0969133	-4.65	0.000	-.6405711	-.2606779
/cvalue	-.5079946	.0615506	-8.25	0.000	-.6286315	-.3873577
/tickets	.151354	.0873677	1.73	0.083	-.0198835	.3225914

```
Instruments for equation 1: kids cvalue tickets
```

FE Poisson by xtpoisson, fe

```
. xtpoisson accidents kids cvalue tickets, fe nolog
Conditional fixed-effects Poisson regression      Number of obs      =      1250
Group variable: id                                Number of groups   =       250

                                                  Obs per group: min =         5
                                                                 avg =       5.0
                                                                 max =         5

                                                  Wald chi2(3)       =    104.31
Log likelihood  = -351.11739                      Prob > chi2        =    0.0000
```

accidents	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
kids	-.4506245	.0981448	-4.59	0.000	-.6429848	-.2582642
cvalue	-.5079949	.0549888	-9.24	0.000	-.615771	-.4002188
tickets	.151354	.0825006	1.83	0.067	-.0103442	.3130521

Bibliography

Blundell, Richard, Rachel Griffith, and Frank Windmeijer. 2002. "Individual effects and dynamics in count data models," *Journal of Econometrics*, 108, 113–131.

Cameron, A. Colin and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and applications*, Cambridge: Cambridge University Press.

Hansen, L. P. 1982. "Large-sample properties of Generalized Method-of-Moment Estimators," *Econometrica*, 1029–1054.

Hausman, Jerry A., Bronwyn H. Hall, and Zvi Griliches. 1984. "Econometric Models for Count Data with an Application to the Patents-R&D Relationship," *Econometrica*, 52(4), 909–938.

Mullahy, J. 1997. "Instrumental variable estimation of Poisson Regression models: Application to models of cigarette smoking behavior," *Review of Economics and Statistics*, 79, 586–593.

Pearson, Karl. 1895. "Contributions to the mathematical theory of evolution—II. Skew variation in homogeneous material," *Philosophical Transactions of the Royal Society of London, Series A*, 186, 343–414.

Wooldridge, Jeffrey. 2002. *Econometric Analysis of Cross Section and Panel Data*, Cambridge, Massachusetts: MIT Press.

Wooldridge, Jeffrey M. 1999. "Distribution-free estimation of some nonlinear panel-data models," *Journal of Econometrics*, 90, 77–90.