Generalized method of moments (GMM) estimation in Stata 11

David M. Drukker

StataCorp

Encuentro de Usarios de Stata en México 2010

Outline

What is GMM?

- The generalize method of moments (GMM) is a general framework for deriving estimators
- Maximum likelihood (ML) is another general framework for deriving estimators.

GMM and ML

- ML estimators use assumptions about the specific families of distributions for the random variables to derive an objective function
  - We maximize this objective function to select the parameters that are most likely to have generated the observed data
- GMM estimators use assumptions about the moments of the random variables to derive an objective function
  - The assumed moments of the random variables are known as the population moments
  - The data provide the sample moments
  - We minimize the objective function to select the parameters that yield the smallest differences between the population moments and the sample moments
- ML is a special case of GMM

What is generalized about GMM?

- For each assumed population moment, we obtain a population moment condition
- For each population moment condition, there is a sample moment condition
- In the method of moments (MM), we have the same number of sample moment conditions as we have parameters
- In the generalized method of moments (MM), we more sample moment conditions than we have parameters

Method of Moments (MM)

- We estimate the mean of a distribution by the sample, the variance by the sample variance, etc
- We want to estimate $\mu = E[y]$
  - The population moment condition is $E[y] - \mu = 0$
  - The sample moment condition is

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

  - Our estimator is obtained by solving the sample moment condition for the parameter
- Estimators that solve sample moment equations to produce estimates are called method-of-moments (MM) estimators
  - This method dates back to Pearson (1895)

Ordinary least squares (OLS) is an MM estimator

- We know that OLS estimates the parameters of the condtional 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 condtions are

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

  Solving for $\boldsymbol{\beta}$ yields

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

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 moment conditions
    - GMM gets a close to solving the over-identified system as possible
    - GMM reduces to MM when the number of parameters equals the number of moment condtions

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 parmeters, $k \leq q$

- 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 $k < q$, the GMM choses 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

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

- 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 inefficent 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
    $$\widehat{\boldsymbol{\theta}}_{GMM1} \equiv \text{arg min}_{\boldsymbol{\theta}} \quad \overline{\mathbf{m}}(\boldsymbol{\theta})'\overline{\mathbf{m}}(\boldsymbol{\theta})$$
    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
    $$\widehat{\boldsymbol{\theta}}_{GMM2} \equiv \text{arg min}_{\boldsymbol{\theta}} \quad \overline{\mathbf{m}}(\boldsymbol{\theta})'\widehat{\mathbf{W}}\overline{\mathbf{m}}(\boldsymbol{\theta})$$
    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 paramters 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 syntax of gmm I

- For many models, the population moment conditions have the form
$$E[\mathbf{z}e(\boldsymbol{\beta})] = \mathbf{0}$$
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 $\boldsymbol{\beta}$
- 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

Modeling crime data I

- We have data

```
. use cscrime, clear

. describe

Contains data from cscrime.dta
  obs:        10,000
 vars:             5                        24 May 2008 17:01
 size:       440,000 (95.8% 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

$$\texttt{crime}_i = \beta_0 + \texttt{policepc}_i\beta_1 + \texttt{legalwage}_i\beta_2 + \epsilon_i$$

- We want to model

$$E[\texttt{crime}|\texttt{policepc}, \texttt{legalwage}] = \beta_0 + \texttt{policepc}\beta_1 + \texttt{legalwage}\beta_2$$

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

$$E\left[\begin{pmatrix}\texttt{policepc}\\ \texttt{legalwage}\end{pmatrix}\epsilon\right] = \begin{pmatrix}0\\0\end{pmatrix}$$

hold

OLS by GMM I

```
. gmm (crime - policepc*{b1} - legalwage*{b2} - {b3}),          ///
>         instruments(policepc legalwage) nolog
Final GMM criterion Q(b) =  6.61e-32
GMM estimation
Number of parameters =   3
Number of moments    =   3
Initial weight matrix: Unadjusted                      Number of obs   =    10000
GMM weight matrix:     Robust

------------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         /b1 |  -.4203287   .0053645   -78.35   0.000    -.4308431   -.4098144
         /b2 |  -7.365905   .2411545   -30.54   0.000    -7.838559   -6.893251
         /b3 |   27.75419   .0311028   892.34   0.000     27.69323    27.81515
------------------------------------------------------------------------------
Instruments for equation 1: policepc legalwage _cons
```

OLS by GMM I

```
. regress crime policepc legalwage, robust

Linear regression                               Number of obs =   10000
                                                F(  2,  9997) = 4422.19
                                                Prob > F      =  0.0000
                                                R-squared     =  0.6092
                                                Root MSE      =  1.8032
```

crime	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
policepc	-.4203287	.0053653	-78.34	0.000	-.4308459	-.4098116
legalwage	-7.365905	.2411907	-30.54	0.000	-7.838688	-6.893123
_cons	27.75419	.0311075	892.20	0.000	27.69321	27.81517

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 contions yield 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) =   .001454
GMM estimation
Number of parameters =   3
Number of moments    =   4
Initial weight matrix: Unadjusted                   Number of obs   =    10000

------------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         /b1 |  -1.002431   .0455469   -22.01   0.000    -1.091701   -.9131606
         /b2 |  -1.281091   .5890977    -2.17   0.030    -2.435702   -.1264811
         /b3 |    30.0494   .1830541   164.16   0.000     29.69062    30.40818
------------------------------------------------------------------------------
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 =    10000
                                                  Wald chi2(2)  = 1891.83
                                                  Prob > chi2   =  0.0000
                                                  R-squared     =       .
                                                  Root MSE      =   3.216
```

crime	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
policepc	-1.002431	.0455469	-22.01	0.000	-1.091701	-.9131606
legalwage	-1.281091	.5890977	-2.17	0.030	-2.435702	-.1264811
_cons	30.0494	.1830541	164.16	0.000	29.69062	30.40818

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

More complicated moment conditions

- The structure of the moment conditions for some model 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 abreviated form of the syntax for gmm is

  gmm *moment_progam* [*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
```

```
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

```
. use xtaccidents

. by id: egen max_a = max(accidents )

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

. 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)

------------------------------------------------------------------------------
             |               Robust
             |      Coef.   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.

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.

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.