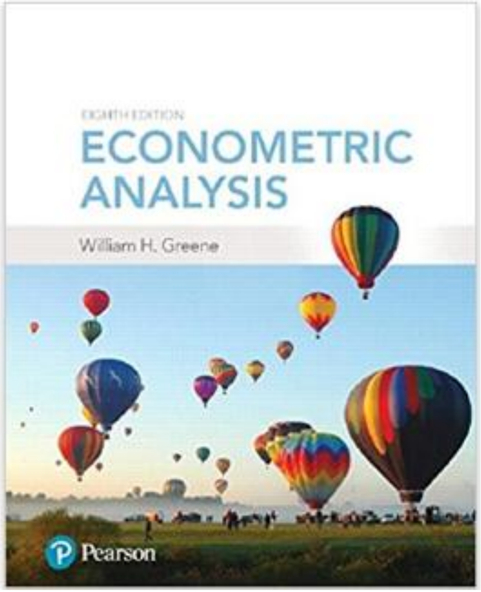
Econometrics I

Professor William Greene
Stern School of Business
Department of Economics



Econometrics I

Part 21 – Generalized Method of Moments

I also have a questions about nonlinear GMM - which is more or less nonlinear IV technique I suppose.

I am running a panel non-linear regression (non-linear in the parameters) and I have L parameters and K exogenous variables with L>K.

In particular my model looks kind of like this: $Y = b_1 * X^{\wedge} b_2 + e$, and so I am trying to estimate the extra b_2 that don't usually appear in a regression.
From what I am reading, to run nonlinear GMM I can use the K exogenous variables to construct the orthogonality conditions but what should I use for the extra, b_2 coefficients?
Just some more possible IVs (like lags) of the exogenous variables?
I agree that by adding more IVs you will get a more efficient estimation, but isn't it only the case when you believe the IVs are truly uncorrelated with the error term?
So by adding more "instruments" you are more or less imposing more and more restrictive assumptions about the model (which might not actually be true).

I am asking because I have not found sources comparing nonlinear GMM/IV to nonlinear least squares. If there is no homoscadesticity/serial correlation what is more efficient/give tighter estimates?

I'm trying to minimize a nonlinear program with the least square under nonlinear constraints. It's first introduced by Ané & Geman (2000). It consisted on the minimization of the sum of squared difference between the moment generating function and the theoretical moment generating function

(*The method was suggested by Quandt and Ramsey in the 1970s*.)

Method of Moment Generating Functions

For the normal distribution, the MGF is

$$M(t|\mu,\sigma)=E[\exp(tx)]=\exp[t\mu + \tfrac{1}{2}t^2\sigma^2]$$

Moment Equations: $\frac{1}{n}\sum_{i=1}^{n} \exp(t_j x_i) = \exp[t_j\mu + \tfrac{1}{2}t_j^{\,2}\sigma^2]$, j = 1, 2.

Choose two values of t and solve the two moment equations for μ and σ.

Mixture of Normals Problem:

$$f(x|\lambda,\mu_1,\sigma_1,\mu_2,\sigma_2) = \lambda N[\mu_1,\sigma_1] + (1-\lambda)N[\mu_2,\sigma_2]$$

Use the method of moment generating functions with 5 values of t.

$$M(t|\mu_1,\sigma_1,,\mu_2,\sigma_2)=E[\exp(tx)]=\lambda\exp[t\mu_1 + \tfrac{1}{2}t^2\sigma_1^{\,2}] + (1-\lambda)\exp[t\mu_2 + \tfrac{1}{2}t^2\sigma_2^{\,2}]$$

Finding the solutions to the moment equations: Least squares

$$\hat{M}(t_1) = \frac{1}{n}\sum_{i=1}^{n} \exp(t_1 x_i), \text{ and likewise for } t_2, \ldots$$

$$\text{Minimize}(\lambda, \mu_1, \sigma_1, \mu_2, \sigma_2)$$

$$\sum_{j=1}^{5} \left[\hat{M}(t_j) - \left(\lambda \exp[t\mu_1 + \tfrac{1}{2}t^2\sigma_1^{\,2}] + (1-\lambda)\exp[t\mu_2 + \tfrac{1}{2}t^2\sigma_2^{\,2}]\right)\right]^2$$

Alternative estimator: Maximum Likelihood

$$L(\lambda, \mu_1, \sigma_1, \mu_2, \sigma_2) = \sum_{i=1}^{n} \log\left\{\lambda N[x_i \mid \mu_1, \sigma_1] + (1-\lambda) N[x_i \mid \mu_2, \sigma_2]\right\}$$

The Method of Moments

Estimating Parameters of Distributions Using Moment Equations

Population Moment

$$\mu_k = E[x^k] = f_k(\theta_1, \theta_2, ..., \theta_K)$$

Sample Moment

$m_k = \frac{1}{n}\Sigma_{i=1}^{n} x_i^k$. m_k may also be $\frac{1}{n}\Sigma_{i=1}^{n} h_k(x_i)$, need not be powers

Law of Large Numbers

$$\text{plim } m_k = \mu_k = f_k(\theta_1, \theta_2, ..., \theta_K)$$

'Moment Equation' (k = 1,...,K)

$$m_k = \frac{1}{N}\Sigma_{i=1}^{N} x_i^k = f_k(\theta_1, \theta_2, ..., \theta_K)$$

Method of Moments applied by inverting the moment equations.

$$\hat{\theta}_k = g_k(m_1, ..., m_K),\ k = 1, ..., K$$

Estimating a Parameter

- Mean of Poisson
  - $p(y)=\exp(-\lambda)\, \lambda^y / y!$
  - $E[y]= \lambda$.
    $\text{plim } (1/n)\Sigma_i y_i = \lambda$.
    This is the estimator
- Mean of Exponential
  - $p(y) = \alpha \exp(- \alpha\, y)$
  - $E[y] = 1/ \alpha$.
    $\text{plim } (1/n)\Sigma_i y_i = 1/ \alpha$

Mean and Variance of a Normal Distribution

$$p(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$$

Population Moments

$$E[y] = \mu, \quad E[y^2] = \sigma^2 + \mu^2$$

Moment Equations

$$\tfrac{1}{n}\Sigma_{i=1}^{n} y_i = \mu, \quad \tfrac{1}{n}\Sigma_{i=1}^{n} y_i^2 = \sigma^2 + \mu^2$$

Method of Moments Estimators

$$\hat{\mu} = \overline{y}, \quad \hat{\sigma}^2 = \tfrac{1}{n}\Sigma_{i=1}^{n} y_i^2 - (\overline{y}^2) = \tfrac{1}{n}\Sigma_{i=1}^{n} (y_i - \overline{y})^2$$

Gamma Distribution

$$p(y) = \frac{\lambda^P \exp(-\lambda y) y^{P-1}}{\Gamma(P)}$$

$$E[y] = \frac{P}{\lambda}$$

$$E[y^2] = \frac{P(P+1)}{\lambda^2} \quad \left(E[y^3] = \frac{P(P+1)(P+2)}{\lambda^3} \text{ and so on} \right)$$

$$E[1/y] = \frac{\lambda}{P-1}$$

$$E[\log y] = \Psi(P) - \log\lambda, \ \Psi(P)=d\ln\Gamma(P)/dP$$

(Each pair gives a different answer. Is there a 'best' pair? Yes, the ones that are 'sufficient' statistics. E[y] and E[logy]. For a different course....)

The Linear Regression Model

Population

$y_i = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i$

Population Expectation

$E[\varepsilon_i x_{ik}] = 0,\ k = 1,...,K$

Moment Equations

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - x_{i1}\beta_1 - x_{i2}\beta_2 - ... - x_{iK}\beta_K)x_{i1} = 0$$

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - x_{i1}\beta_1 - x_{i2}\beta_2 - ... - x_{iK}\beta_K)x_{i2} = 0$$

...

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - x_{i1}\beta_1 - x_{i2}\beta_2 - ... - x_{iK}\beta_K)x_{iK} = 0$$

Solution: Linear system of K equations in K unknowns. Least Squares

Instrumental Variables

Population

$y_i = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i$

Population Expectation

$E[\varepsilon_i z_{ik}] = 0$ for instrumental variables z_1 ... z_K.

Moment Equations

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - x_{i1}\beta_1 - x_{i2}\beta_2 - \ldots - x_{iK}\beta_K)z_{i1} = 0$$

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - x_{i1}\beta_1 - x_{i2}\beta_2 - \ldots - x_{iK}\beta_K)z_{i2} = 0$$

...

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - x_{i1}\beta_1 - x_{i2}\beta_2 - \ldots - x_{iK}\beta_K)z_{iK} = 0$$

Solution : Also a linear system of K equations in K unknowns.

$$b_{IV} = (\mathbf{Z'X} / n)^{-1}(\mathbf{Z'y} / n)$$

An extension : What is the solution if there are M > K IVs?

Maximum Likelihood

Log likelihood function, logL = $\frac{1}{n}\Sigma_{i=1}^{n}\log f(y_i\,|\,x_i,\theta_1,...,\theta_K)$

Population Expectations

$$E\left[\frac{\partial \log L}{\partial \theta_k}\right]=0,\ k = 1,...,K$$

Sample Moments

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \log f(y_i\,|\,x_i,\theta_1,...,\theta_K)}{\partial \theta_k}=0$$

Solution: K nonlinear equations in K unknowns.

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \log f(y_i\,|\,x_i,\hat{\theta}_{1,MLE},...,\hat{\theta}_{K,MLE})}{\partial \hat{\theta}_{k,MLE}}=0$$

Behavioral Application

Life Cycle Consumption (text, pages 488-489)

$$E_t\left[(1+r)\left(\frac{1}{1+\delta}\right)\left(\frac{c_{t+1}}{c_t}\right)^{-\alpha}-1|\Omega_t\right]=0$$

$\delta =$ discount rate

$c_t =$ consumption

$\Omega_t =$ information at time t

Let β=1/(1+δ), $R_{t+1} = c_{t+1} / c_t$, λ=-α

$$E_t[\beta(1+r)R_{t+1}^{\lambda}-1|\,\Omega_t]=0$$

What is in the information set? Each piece of 'information' provides a moment equation for estimation of the two parameters.

$$\left[\sum_{t=1}^{T}\left((1+r)\left(\frac{1}{1+\delta}\right)\left(\frac{c_{t+1}}{c_t}\right)^{-\alpha}-1\right)w_{tk}\right]=0,\ k=1,...,K$$

Identification

- Can the parameters be estimated?
- Not a sample ‘property’
- Assume an infinite sample
  - Is there sufficient information in a sample to reveal consistent estimators of the parameters
  - Can the ‘moment equations’ be solved for the population parameters?

Identification

- Exactly Identified Case: K population moment equations in K unknown parameters.
  - Our familiar cases, OLS, IV, ML, the MOM estimators
  - Is the counting rule sufficient?
  - What else is needed?
- Overidentified Case
  - Instrumental Variables
- Underidentified Case
  - Multicollinearity
  - Variance parameter in a probit model
  - Shape parameter in a loglinear model

Overidentification

Population

$y_i = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i,\ \beta_1,...,\beta_K$

Population Expectation

$E[\varepsilon_i z_{ik}] = 0$ for instrumental variables z_1 ... z_M M > K.

There are M > K Moment Equations - more than necessary

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - x_{i1}\beta_1 - x_{i2}\beta_2 - ... - x_{iK}\beta_K)z_{i1} = 0$$

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - x_{i1}\beta_1 - x_{i2}\beta_2 - ... - x_{iK}\beta_K)z_{i2} = 0$$

...

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - x_{i1}\beta_1 - x_{i2}\beta_2 - ... - x_{iK}\beta_K)z_{iM} = 0$$

Solution: A linear system of M equations in K unknowns.

Overidentification

Two Equation Covariance Structures Model

Country 1: $\mathbf{y}_1 = \mathbf{X}_1\boldsymbol{\beta} + \boldsymbol{\varepsilon}_1$

Country 2: $\mathbf{y}_2 = \mathbf{X}_2\boldsymbol{\beta} + \boldsymbol{\varepsilon}_2$

Two Population Moment Conditions:

$E[(1/T)\ \mathbf{X}_1'(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\beta})] = \mathbf{0}$

$E[(1/T)\ \mathbf{X}_2'(\mathbf{y}_2 - \mathbf{X}_2\boldsymbol{\beta})] = \mathbf{0}$

(1) How do we combine the two sets of equations?

(2) Given two OLS estimates, $\mathbf{b}_1$ and $\mathbf{b}_2$, how do we reconcile them?

Note: There are even more. $E[(1/T)\ \mathbf{X}_1'(\mathbf{y}_2 - \mathbf{X}_2\boldsymbol{\beta})] = \mathbf{0}$.

Underidentification – Model/Data

Consider the Mover - Stayer Model

Binary choice for whether an individual 'moves' or 'stays'

$$d_i = 1(\mathbf{z}_i'\alpha + u_i > 0)$$

Outcome equation for the individual, conditional on the state:

$$y_i \,|\, (d = 0) \;=\; \mathbf{x}_i'\beta_0 + \varepsilon_{i0}$$

$$y_i \,|\, (d = 1) \;=\; \mathbf{x}_i'\beta_1 + \varepsilon_{i1}$$

$$(\varepsilon_{i0}, \varepsilon_{i1}) \sim N[(0,0),(\sigma_0^2, \sigma_1^2, \rho\sigma_0\sigma_1)]$$

An individual either moves or stays, but not both (or neither).

The parameter ρ cannot be estimated with the observed data regardless of the sample size. It is unidentified.

Underidentification - Normalization

When a parameter is unidentified, the log likelihood is invariant to changes in it. Consider the logit binary choice model

$$\text{Prob}[y=0]=\frac{\exp(\beta_0 x)}{\exp(\beta_0 x)+\exp(\beta_1 x)} \qquad \text{Prob}[y=1]=\frac{\exp(\beta_1 x)}{\exp(\beta_0 x)+\exp(\beta_1 x)}$$

Probabilities sum to 1, are monotonic, etc. But, consider, for any $\delta \neq 0$,

$$\text{Prob}[y=0]=\frac{\exp[(\beta_0+\delta)x]}{\exp[(\beta_0+\delta)x]+\exp[(\beta_1+\delta)x]}=\frac{\exp(\delta x)\,[\exp(\beta_0 x)]}{\exp(\delta x)[\exp(\beta_0 x)+\exp(\beta_1 x)]}$$

$$\text{Prob}[y=1]=\frac{\exp[(\beta_1+\delta)x]}{\exp[(\beta_0+\delta)x]+\exp[(\beta_1+\delta)x]}=\frac{\exp(\delta x)\,[\exp(\beta_1 x)]}{\exp(\delta x)[\exp(\beta_0 x)+\exp(\beta_1 x)]}$$

$\exp(\delta x)$ always cancels out.

The parameters are unidentified. A normalization such as $\beta_0 = 0$ is needed.

Underidentification: Moments Nonlinear LS vs. MLE

$$y_i \sim \text{Gamma}(P,\lambda_i),\ \lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i)$$

$$f(y_i) = \frac{\lambda_i^P \exp(-\lambda_i y_i) y_i^{P-1}}{\Gamma(P)}$$

$$E[y_i \mid \mathbf{x}_i] = \frac{P}{\lambda_i}$$

We consider nonlinear least squares and maximum likelihood estimation of the parameters. We use the German health care data, where

y = income

x = 1,age,educ,female,hhkids,married

Nonlinear Least Squares

```
--> NAMELIST ; x = one,age,educ,female,hhkids,married $
--> Calc     ; k=col(x) $
--> NLSQ     ; Lhs = hhninc ; Fcn = p / exp(b1'x)
             ; labels = k_b,p ; start = k_0,1 ; maxit = 20$
Moment matrix has become nonpositive definite.
Switching to BFGS algorithm
Normal exit:  16 iterations. Status=0. F=    381.1028
-----------------------------------------------------------
User Defined Optimization.........................
Nonlinear    least squares regression ............
LHS=HHNINC   Mean                 =          .35208
             Standard deviation   =          .17691
             Number of observs.   =           27326
Model size   Parameters           =               7
             Degrees of freedom   =           27319
Residuals    Sum of squares       =       762.20551
             Standard error of e  =          .16701
--------+--------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]
--------+--------------------------------------------------
      B1|    1.39905       14319.39          .000   .9999 <======
      B2|     .00029         .00029          .986   .3242
      B3|    -.05527***      .00105       -52.809   .0000
      B4|    -.01843***      .00580        -3.180   .0015
      B5|     .05445***      .00665         8.184   .0000
      B6|    -.26424***      .00823       -32.109   .0000
       P|     .63239       9055.493          .000   .9999 <=======
--------+--------------------------------------------------
Nonlinear least squares did not work. That is the implication of the
infinite standard errors for B1 (the constant) and P.
```

Maximum Likelihood

```
----------------------------------------------------------------------
Gamma (Loglinear) Regression Model
Dependent variable                  HHNINC
Log likelihood function        14293.00214
Restricted log likelihood       1195.06953
Chi squared [    6 d.f.]       26195.86522
Significance level                  .00000
McFadden Pseudo R-squared      -10.9599753 (4 observations with income = 0
Estimation based on N =  27322, K =    7  were deleted so logL was
computable.)
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+-------------------------------------------------------------
        |Parameters in conditional mean function
Constant|    3.40841***       .02154      158.213   .0000
     AGE|     .00205***       .00028        7.413   .0000      43.5272
    EDUC|    -.05572***       .00120      -46.496   .0000      11.3202
  FEMALE|    -.00542          .00545        -.995   .3198       .47881
  HHKIDS|     .06512***       .00618       10.542   .0000       .40272
 MARRIED|    -.26341***       .00692      -38.041   .0000       .75869
        |Scale parameter for gamma model
 P_scale|    5.12486***       .04250      120.594   .0000
--------+-------------------------------------------------------------
```

MLE apparently worked fine. Why did one method (nls) fail and another consistent estimator work without difficulty?

Moment Equations: NLS

$$E[y \mid \mathbf{x}] = P / \exp(\boldsymbol{\beta}'\mathbf{x}_i)$$

$$\mathbf{e}'\mathbf{e} = \sum_{i=1}^{n} \left(y_i - P / \exp(\boldsymbol{\beta}'\mathbf{x}_i)\right) = \sum_{i=1}^{n} e_i^2$$

$$\frac{\partial \mathbf{e}'\mathbf{e}}{\partial P} = \sum_{i=1}^{n} \frac{-2e_i}{\exp(\boldsymbol{\beta}'\mathbf{x}_i)} = 0$$

$$\frac{\partial \mathbf{e}'\mathbf{e}}{\partial \beta} = \sum_{i=1}^{n} \frac{2e_i P}{\exp(\boldsymbol{\beta}'\mathbf{x}_i)} \mathbf{x}_i = \mathbf{0}$$

Consider the term for the constant in the model, β_1. Notice that the first order condition for the constant term is

$\sum_{i=1}^{n} \frac{2e_i P}{\exp(\boldsymbol{\beta}'\mathbf{x}_i)} = 0$. This doesn't depend on P, since we can divide both sides of the equation by P. This means that we cannot find solutions for both β_1 and P. It is easy to see why NLS cannot distinguish P from β_1. E[y|x] = $\exp((\log P - \beta_1) - \ldots)$. There are an infinite number of pairs of (P,β_1) that produce the same constant term in the model.

Moment Equations MLE

The log likelihood function and likelihood equations are

$$\text{logL}=\sum_{i=1}^{n} P\log\lambda_i - \log\Gamma(P) - \lambda_i y_i + (P-1)\log y_i$$

$$\frac{\partial \log L}{\partial P} = \sum_{i=1}^{n} \left(\log\lambda_i - \Psi(P) + \log y_i\right) = 0,\ \ \Psi(P) = \frac{d\log\Gamma(P)}{dP}$$

$$\frac{\partial \log L}{\partial \beta_i} = \sum_{i=1}^{n} \left(\frac{P}{\lambda_i}\lambda_i - y_i\lambda_i\right);\ \text{using}\ \frac{\partial\lambda_i}{\partial \boldsymbol{\beta}} = \lambda_i \mathbf{x}_i.$$

Recall, the expected values of the derivatives of the log likelihood equal zero. So, a look at the first equation reveals that the moment equation in use for estimating P is $\text{E}[\text{logy}_i\,|\,\mathbf{x}_i] = \Psi(P) - \log\lambda_i$ and another K moment equations, $\text{E}\left[\left(\text{y}_i - \frac{P}{\lambda_i}\right)\mathbf{x}_i\right] = 0$ are also in use. So, the MLE uses K+1 functionally independent moment equations for K+1 parameters, while NLS was only using K independent moment equations for the same K+1 parameters.

GMM Agenda

- The Method of Moments. Solving the moment equations
  - Exactly identified cases
  - Overidentified cases
- Consistency. How do we know the method of moments is consistent?
- Asymptotic covariance matrix.
- Consistent vs. Efficient estimation
  - A weighting matrix
  - The minimum distance estimator
  - What is the efficient weighting matrix?
  - Estimating the weighting matrix.
- The Generalized method of moments estimator - how it is computed.
- Computing the appropriate asymptotic covariance matrix

The Method of Moments

Moment Equation: Defines a sample statistic that mimics a population expectation:

The population expectation – orthogonality condition:

E[$\mathbf{m}_i(\beta)$] = **0**. Subscript i indicates it depends on data vector indexed by 'i' (or 't' for a time series setting)

The Method of Moments - Example

Gamma Distribution Parameters

$$p(y_i) = \frac{\lambda^P \exp(-\lambda y_i) y_i^{P-1}}{\Gamma(P)}$$

Population Moment Conditions

$$E[y_i] = \frac{P}{\lambda}, \quad E[\log y_i] = \Psi(P) - \log\lambda$$

Moment Equations:

$$E[\bar{m}_1(\lambda,P)] = E[\{(1/n)\Sigma_{i=1}^{n} y_i\} - P/\lambda] = 0$$

$$E[\bar{m}_2(\lambda,P)] = E[\{(1/n)\Sigma_{i=1}^{n} \log y_i\} - (\Psi(P) - \log\lambda)] = 0$$

Application

	I	Y
1 »	1	20.5
2 »	2	31.5
3 »	3	47.7
4 »	4	26.2
5 »	5	44
6 »	6	8.28
7 »	7	30.8
8 »	8	17.2
9 »	9	19.9
10 »	10	9.96
11 »	11	55.8
12 »	12	25.2
13 »	13	29
14 »	14	85.5
15 »	15	15.1
16 »	16	28.5
17 »	17	21.4
18 »	18	17.7
19 »	19	6.42
20 »	20	84.9

Solving the moment equations

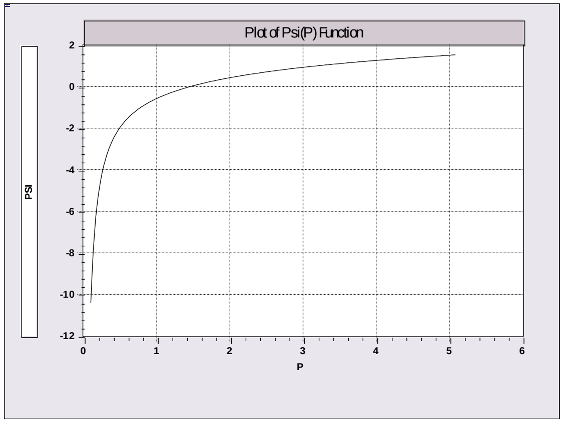
Use least squares:

Minimize $\{\bar{m}_1 - E[\bar{m}_1]\}^2 + \{\bar{m}_2 - E[\bar{m}_2]\}^2$

$$= (\bar{m}_1 - (P/\lambda))^2 + (\bar{m}_2 - (\Psi(P) - \log\lambda))^2$$

$$\bar{m}_1 = 31.278$$

$$\bar{m}_2 = 3.221387$$



Method of Moments Solution

```
create   ; y1=y       ;  y2=log(y)$
calc     ; m1=xbr(y1) ;  ms=xbr(y2)$
minimize; start = 2.0, .06 ; labels = p,l
        ; fcn   =   (l*m1-p)^2
                  + (ms - psi(p)+log(l)) ^2  $
+---------------------------------------------+
| User Defined Optimization                   |
| Dependent variable            Function      |
| Number of observations               1      |
| Iterations completed                 6      |
| Log likelihood function      .5062979E-13 |
+---------------------------------------------+
+---------+--------------+
|Variable | Coefficient  |
+---------+--------------+
 P             2.41060361
 L              .07707026
```

Nonlinear Instrumental Variables

There are K parameters, $\boldsymbol{\beta}$

$$y_i = f(\mathbf{x_i},\boldsymbol{\beta}) + \varepsilon_i.$$

There exists a set of K instrumental variables, $\mathbf{z}_i$ such that $E[\mathbf{z}_i \varepsilon_i] = 0.$

The sample counterpart is the **moment equation**

$$(1/n)\Sigma_i \mathbf{z_i} \varepsilon_i = (1/n)\Sigma_i \mathbf{z_i} [y_i - f(\mathbf{x_i},\boldsymbol{\beta})]$$

$$= (1/n)\Sigma_i \mathbf{m_i} (\boldsymbol{\beta}) = \overline{m}(\boldsymbol{\beta}) = \mathbf{0}.$$

The **method of moments estimator** is the solution to the moment equation(s).

(How the solution is obtained is not always obvious, and varies from problem to problem.)

The MOM Solution

There are K equations in K unknowns in $\bar{\mathbf{m}}(\beta)=\mathbf{0}$

If there is a solution, there is an exact solution

At the solution, $\bar{\mathbf{m}}(\beta)=\mathbf{0}$, and $[\bar{\mathbf{m}}(\beta)]'[\bar{\mathbf{m}}(\beta)] = 0$

Since $[\bar{\mathbf{m}}(\beta)]'[\bar{\mathbf{m}}(\beta)] \geq 0$, the solution can be found by solving the programming problem

Minimize wrt β : $[\bar{\mathbf{m}}(\beta)]'[\bar{\mathbf{m}}(\beta)]$

For this problem,

$$[\bar{\mathbf{m}}(\beta)]'[\bar{\mathbf{m}}(\beta)] = [(1/n)\boldsymbol{\varepsilon}'\mathbf{Z}] \times [(1/n)\mathbf{Z}'\boldsymbol{\varepsilon}]$$

The solution is defined by

$$\frac{\partial[\bar{\mathbf{m}}(\beta)]'[\bar{\mathbf{m}}(\beta)]}{\partial\beta} = \frac{\partial[(1/n)\boldsymbol{\varepsilon}'\mathbf{Z}] \times [(1/n)\mathbf{Z}'\boldsymbol{\varepsilon}]}{\partial\beta}$$

MOM Solution

$$\frac{\partial[(1/n)\boldsymbol{\varepsilon}'\mathbf{Z}] \times [(1/n)\mathbf{Z}'\boldsymbol{\varepsilon}]}{\partial\beta} = -2\left[(1/n)\mathbf{G}'\mathbf{Z}\right][(1/n)\mathbf{Z}'\boldsymbol{\varepsilon}]$$

$\mathbf{G}$ = $n \times K$ matrix with row i equal to $\mathbf{g}_i = \dfrac{\partial f(\mathbf{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}$

For the classical linear regression model,

$f(\mathbf{x}_i, \boldsymbol{\beta}) = \mathbf{x}_i'\boldsymbol{\beta}$, $\mathbf{Z} = \mathbf{X}$, $\mathbf{G} = \mathbf{X}$, and the FOC are

$$-2[(1/n)(\mathbf{X}'\mathbf{X})]\,[(1/n)\mathbf{X}'\boldsymbol{\varepsilon}] = 0$$

which has unique solution $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$

Variance of the Method of Moments Estimator

The MOM estimator solves $\bar{m}(\boldsymbol{\beta})=\mathbf{0}$

$\bar{\mathbf{m}}(\boldsymbol{\beta})=\frac{1}{n}\sum_{i=1}^{n}\mathbf{m}_i(\boldsymbol{\beta})$ so the variance is $\frac{1}{n}\boldsymbol{\Omega}$ for some $\boldsymbol{\Omega}$

Generally, $\boldsymbol{\Omega} = E[\mathbf{m}_i(\boldsymbol{\beta})\mathbf{m}_i(\boldsymbol{\beta})']$

The asymptotic covariance matrix of the estimator is

$$\text{Asy.Var}[\boldsymbol{\beta}_{\mathbf{MOM}}]=(\mathbf{G})^{-1}\left(\frac{1}{n}\boldsymbol{\Omega}\right)(\mathbf{G}')^{-1} \text{ where } \mathbf{G}=\frac{\partial\bar{\mathbf{m}}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}'}$$

Example 1: Gamma Distribution

$$\bar{m}_1 = \tfrac{1}{n}\Sigma_{i=1}^{n}(y_i - \tfrac{P}{\lambda})$$

$$\bar{m}_2 = \tfrac{1}{n}\Sigma_{i=1}^{n}(\log y_i - \Psi(P) + \log\lambda)$$

$$\frac{1}{n}\boldsymbol{\Omega} = \frac{1}{n}\begin{bmatrix} \text{Var}(y_i) & \text{Cov}(y_i, \log y_i) \\ \text{Cov}(y_i, \log y_i) & \text{Var}(\log y_i) \end{bmatrix}$$

$$\mathbf{G} = \frac{1}{n}\sum_{i=1}^{N}\begin{bmatrix} -\frac{1}{\lambda} & \frac{P}{\lambda^2} \\ -\Psi'(P) & \frac{1}{\lambda} \end{bmatrix}$$

$$\hat{\boldsymbol{\Omega}} = \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} y_i - \bar{y} \\ \log y_i - \overline{\log y} \end{bmatrix}\begin{bmatrix} y_i - \bar{y} & \log y_i - \overline{\log y} \end{bmatrix}$$

Example 2: Nonlinear IV Least Squares

$y_i = f(\mathbf{x}_i, \beta) + \varepsilon_i$, $\mathbf{z_i}$ = the set of K instrumental variables

$\text{Var}[\varepsilon_i] = \sigma^2$

$\mathbf{m}_i = \mathbf{z}_i \varepsilon_i$

$\text{Var}[\mathbf{m}_i] = \sigma^2 \mathbf{z_i z_i'}$

With independent observations, observations are uncorrelated

$\text{Var}[\bar{\mathbf{m}}(\beta)] = (1/n^2)\sum_{i=1}^{n} \sigma^2 \mathbf{z_i z_i'} = (\sigma^2 / n^2)\mathbf{Z'Z}$

$\mathbf{G} = (1/n)\sum_{i=1}^{n} -\mathbf{z}_i \mathbf{x}_i^0{'}$ where $\mathbf{x}_i^0$ is the vector of 'pseudo-regressors,'

$\mathbf{x}_i^0 = \dfrac{\partial f(\mathbf{x}_i, \beta)}{\partial \boldsymbol{\beta}}$. In the linear model, this is just $\mathbf{x}_i$.

$\mathbf{G} = -(1/n)\mathbf{Z'X}^0$.

$$(\mathbf{G}^{-1})\mathbf{V}(\mathbf{G}^{-1})' = [-(1/n)\mathbf{Z'X}^0]^{-1}[(\sigma^2/n^2)\mathbf{Z'Z}][-(1/n)\mathbf{X}^0{'}\mathbf{Z}]^{-1}$$
$$= \sigma^2[\mathbf{Z'X}^0]^{-1}[\mathbf{Z'Z}][\mathbf{X}^0{'}\mathbf{Z}]^{-1}$$

Variance of the Moments

How to estimate $\mathbf{V} = (1/n)\boldsymbol{\Omega} = \text{Var}[\bar{\mathbf{m}}(\beta)]$

$$\text{Var}[\bar{\mathbf{m}}(\boldsymbol{\beta})]=(1/n)\text{Var}[\mathbf{m}_i(\boldsymbol{\beta})] = (1/n)\boldsymbol{\Omega}$$

Estimate $\text{Var}[\mathbf{m}_i(\boldsymbol{\beta})]$ with $\text{Est.Var}[\mathbf{m}_i(\boldsymbol{\beta})] = (1/n)\sum_{i=1}^{n}\mathbf{m}_i(\boldsymbol{\beta})\mathbf{m}_i(\boldsymbol{\beta})'$

Then,

$$\hat{\mathbf{V}} = (1/n)\times(1/n)\times\sum_{i=1}^{n}\mathbf{m}_i(\hat{\boldsymbol{\beta}})\mathbf{m}_i(\hat{\boldsymbol{\beta}})'$$

For the linear regression model,

$$\mathbf{m}_i = \mathbf{x}_i\varepsilon_i,$$

$$\hat{\mathbf{V}} = (1/n)\times(1/n)\times\sum_{i=1}^{n}\mathbf{x}_i e_i e_i \mathbf{x}_i' = (1/n)\times(1/n)\times\sum_{i=1}^{n}e_i^2\mathbf{x}_i\mathbf{x}_i'$$

$$\mathbf{G} = (1/n)\mathbf{X'X}$$

$$\text{Est.Var}[\mathbf{b}_{MOM}] = [(1/n)\mathbf{X'X}]^{-1}[(1/n)\times(1/n)\times\sum_{i=1}^{n}e_i^2\mathbf{x}_i\mathbf{x}_i'][(1/n)\mathbf{X'X}]^{-1}$$

$$= [\mathbf{X'X}]^{-1}[\sum_{i=1}^{n}e_i^2\mathbf{x}_i\mathbf{x}_i'][\mathbf{X'X}]^{-1} \qquad \text{(familiar?)}$$

Properties of the MOM Estimator

- Consistent?
  - The LLN implies that the moments are consistent estimators of their population counterparts (zero)
  - Use the Slutsky theorem to assert consistency of the functions of the moments
- Asymptotically normal? The moments are sample means. Invoke a central limit theorem.
- Efficient? Not necessarily
  - Sometimes yes. (Gamma example)
  - Perhaps not. Depends on the model and the available information (and how much of it is used).

Generalizing the Method of Moments Estimator

- More moments than parameters – the overidentified case
- Example: Instrumental variable case, M > K instruments

Two Stage Least Squares

How to use an "excess" of instrumental variables

(1) **X** is K variables. Some (at least one) of the K variables in **X** are correlated with **ε**.

(2) **Z** is M > K variables. Some of the variables in **Z** are also in **X**, some are not. None of the variables in **Z** are correlated with **ε.**

(3) Which K variables to use to compute **Z'X** and **Z'y?**

Choosing the Instruments

- Choose K randomly?
- Choose the included Xs and the remainder randomly?
- Use all of them? How?
- A theorem: (Brundy and Jorgenson, ca. 1972) There is a most efficient way to construct the IV estimator from this subset:
  - (1) For each column (variable) in **X**, compute the predictions of that variable using all the columns of **Z**.
  - (2) Linearly regress **y** on these K predictions.
- This is two stage least squares

2SLS Algebra

$$\hat{\mathbf{X}} = \mathbf{Z(Z'Z)^{-1}Z'X}$$

$$\mathbf{b_{2SLS}} = (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'\mathbf{y}$$

But, $\mathbf{Z(Z'Z)^{-1}Z'X} = (\mathbf{I} - \mathbf{M_Z})\mathbf{X}$ and $(\mathbf{I} - \mathbf{M_Z})$ is idempotent.

$\hat{\mathbf{X}}'\hat{\mathbf{X}} = \mathbf{X}'(\mathbf{I} - \mathbf{M_Z})(\mathbf{I} - \mathbf{M_Z})\mathbf{X} = \mathbf{X}'(\mathbf{I} - \mathbf{M_Z})\mathbf{X}$ so

$\mathbf{b_{2SLS}} = (\hat{\mathbf{X}}'\mathbf{X})^{-1}\hat{\mathbf{X}}'\mathbf{y}$ = a real IV estimator by the definition.

Note, plim($\hat{\mathbf{X}}'\varepsilon$/n) = $\mathbf{0}$ since columns of $\hat{\mathbf{X}}$ are linear combinations of the columns of $\mathbf{Z}$, all of which are uncorrelated with ε.

$$\mathbf{b_{2SLS}} = [\mathbf{X}'(\mathbf{I} - \mathbf{M_Z})\mathbf{X}]^{-1}\mathbf{X}'(\mathbf{I} - \mathbf{M_Z})\mathbf{y}$$

Method of Moments Estimation

Same Moment Equation

$$\bar{\mathbf{m}}(\boldsymbol{\beta})=\mathbf{0}$$

Now, M moment equations, K parameters. There is no unique solution. There is also no exact solution to

$$\bar{\mathbf{m}}(\boldsymbol{\beta})=\mathbf{0}.$$

We get as close as we can.

How to choose the estimator? Least squares is an obvious choice.

$$\text{Minimize wrt } \boldsymbol{\beta} : \bar{\mathbf{m}}(\boldsymbol{\beta})'\bar{\mathbf{m}}(\boldsymbol{\beta})$$

E.g., $\text{Minimize wrt } \boldsymbol{\beta} : [(1/n)\varepsilon(\boldsymbol{\beta})'\mathbf{Z}][(1/n)\mathbf{Z}'\varepsilon(\boldsymbol{\beta})]=(1/n^2)\varepsilon(\boldsymbol{\beta})'\mathbf{ZZ}'\varepsilon(\boldsymbol{\beta})$

FOC for MOM

First order conditions

(1) General

$$\partial \bar{\mathbf{m}}(\boldsymbol{\beta})'\bar{\mathbf{m}}(\boldsymbol{\beta})/\partial\boldsymbol{\beta} = 2\mathbf{G}(\boldsymbol{\beta})'\bar{\mathbf{m}}(\boldsymbol{\beta}) = 0$$

(2) The Instrumental Variables Problem

$$\partial(1/n^2)\varepsilon(\boldsymbol{\beta})'\mathbf{ZZ}'\varepsilon(\boldsymbol{\beta})/\partial\boldsymbol{\beta} = -(2/n^2)(\mathbf{X}'\mathbf{Z})[\mathbf{Z}'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})]$$

$$= \mathbf{0}$$

Or, $(\mathbf{X}'\mathbf{Z})[\mathbf{Z}'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})] = \mathbf{0}$

$$(K\times M)\ (M\times n)(n\times 1) = \mathbf{0}$$

At the solution, $(\mathbf{X}'\mathbf{Z})[\mathbf{Z}'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})] = \mathbf{0}$

But, $[\mathbf{Z}'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})] \neq \mathbf{0}$ as it was before.

Computing the Estimator

- Programming Program
- No all purpose solution
- Nonlinear optimization problem – solution varies from setting to setting.

Asymptotic Covariance Matrix

General Result for Method of Moments when $M \geq K$

$$\text{Moment Equations:} E[\bar{\mathbf{m}}(\boldsymbol{\beta})] = \mathbf{0}$$

$$\text{Solution - FOC: } \mathbf{G}(\boldsymbol{\beta})'\bar{\mathbf{m}}(\boldsymbol{\beta}) = \mathbf{0},\ \mathbf{G}(\boldsymbol{\beta})' \text{ is } K \times M$$

Asymptotic Covariance Matrix

$$\text{Asy.Var}[\hat{\boldsymbol{\beta}}] = [\mathbf{G}(\boldsymbol{\beta})'\mathbf{V}^{-1}\mathbf{G}(\boldsymbol{\beta})]^{-1},\ \mathbf{V} = \text{Asy.Var}[\bar{\mathbf{m}}(\boldsymbol{\beta})]$$

Special Case - Exactly Identified: $M = K$ and

$\mathbf{G}(\boldsymbol{\beta})$ is nonsingular. Then $[\mathbf{G}(\boldsymbol{\beta})]^{-1}$ exists and

$$\text{Asy.Var}[\hat{\boldsymbol{\beta}}] = [\mathbf{G}(\boldsymbol{\beta})]^{-1}\ \mathbf{V}\ [\mathbf{G}(\boldsymbol{\beta})']^{-1}$$

More Efficient Estimation

We have used least squares,

Minimize wrt $\boldsymbol{\beta}$: $\bar{\mathbf{m}}(\boldsymbol{\beta})'\bar{\mathbf{m}}(\boldsymbol{\beta})$

to find the estimator of $\boldsymbol{\beta}$. Is this the most efficient way to proceed?

Generally not: We consider a more general approach

Minimum Distance Estimation

Let $\mathbf{A}$ be any positive definite matrix:

Let $\hat{\boldsymbol{\beta}}_{MD}$ = the solution to Minimize wrt $\boldsymbol{\beta}$ **:**

$$\mathbf{q} = \bar{\mathbf{m}}(\boldsymbol{\beta})'\,\mathbf{A}\,\bar{\mathbf{m}}(\boldsymbol{\beta})$$

This is a minimum distance (in the metric of **A**) estimator.

Minimum Distance Estimation

Let $\mathbf{A}$ be any positive definite matrix:

Let $\hat{\boldsymbol{\beta}}_{MD}$ = the solution to Minimize wrt $\boldsymbol{\beta}$:

$$\mathbf{q} = \bar{\mathbf{m}}(\boldsymbol{\beta})'\,\mathbf{A}\,\bar{\mathbf{m}}(\boldsymbol{\beta})$$

where $E[\bar{\mathbf{m}}(\boldsymbol{\beta})] = 0$ (the usual moment conditions).

This is a minimum distance (in the metric of **A**) estimator.

$\hat{\boldsymbol{\beta}}_{MD}$ is consistent

$\hat{\boldsymbol{\beta}}_{MD}$ is asymptotically normally distributed.

Same arguments as for the GMM estimator. Efficiency of the estimator depends on the choice of **A**.

MDE Estimation: Application

N units, T observations per unit, T > K

$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_\mathbf{i},\ E[\boldsymbol{\varepsilon}_\mathbf{i} \mid \mathbf{X}_i] = 0$

Consider the following estimation strategy:

(1) OLS country by country, $\mathbf{b}_i$ produces N estimators of $\boldsymbol{\beta}$

(2) How to combine the estimators?

We have 'moment' equation: $E\begin{bmatrix} \mathbf{b}_1 - \boldsymbol{\beta} \\ \mathbf{b}_2 - \boldsymbol{\beta} \\ \mathbf{...} \\ \mathbf{b}_N - \boldsymbol{\beta} \end{bmatrix} = \mathbf{0}$

How can I combine the N estimators of **β?**

Least Squares

$$E\begin{bmatrix} \mathbf{b}_1 - \boldsymbol{\beta} \\ \mathbf{b}_2 - \boldsymbol{\beta} \\ \cdots \\ \mathbf{b}_N - \boldsymbol{\beta} \end{bmatrix} = \mathbf{0}. \quad \mathbf{m}(\boldsymbol{\beta}) = \begin{bmatrix} \mathbf{b}_1 - \boldsymbol{\beta} \\ \mathbf{b}_2 - \boldsymbol{\beta} \\ \cdots \\ \mathbf{b}_N - \boldsymbol{\beta} \end{bmatrix}$$

To minimize $\mathbf{m}(\boldsymbol{\beta})'\mathbf{m}(\boldsymbol{\beta}) = \sum_{i=1}^{N} (\mathbf{b}_i - \boldsymbol{\beta})'(\mathbf{b}_i - \boldsymbol{\beta})$

$$\frac{\partial \mathbf{m}(\boldsymbol{\beta})'\mathbf{m}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2[\mathbf{I}, \mathbf{I}, \ldots, \mathbf{I}] \begin{bmatrix} \mathbf{b}_1 - \boldsymbol{\beta} \\ \mathbf{b}_2 - \boldsymbol{\beta} \\ \cdots \\ \mathbf{b}_N - \boldsymbol{\beta} \end{bmatrix} = -2\sum_{i=1}^{N} (\mathbf{b}_i - \boldsymbol{\beta}) = \mathbf{0}.$$

The solution is $\sum_{i=1}^{N} (\mathbf{b}_i - \boldsymbol{\beta}) = \mathbf{0}$ or $\boldsymbol{\beta} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{b}_i = \bar{\mathbf{b}}$

Generalized Least Squares

The preceding used OLS - simple unweighted least squares.

Also, it uses $\mathbf{A} = \begin{bmatrix} \mathrm{I} & 0 & \ldots & 0 \\ 0 & \mathrm{I} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \mathrm{I} \end{bmatrix}$. Suppose we use weighted, GLS.

Then, $\mathbf{A} = \begin{bmatrix} [\sigma_1^2(\mathbf{X}_1'\mathbf{X}_1)^{-1}]^{-1} & 0 & \ldots & 0 \\ 0 & [\sigma_2^2(\mathbf{X}_2'\mathbf{X}_2)^{-1}]^{-1} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & [\sigma_N^2(\mathbf{X}_N'\mathbf{X}_N)^{-1}]^{-1} \end{bmatrix}$

The first order condition for minimizing **m(β)'Am(β)** is

$$\sum_{i=1}^{N} \{[\sigma_i^2(\mathbf{X}_i'\mathbf{X}_i)^{-1}]^{-1}\}(\mathbf{b}_i - \boldsymbol{\beta}) = 0$$

$$\text{or} \quad \boldsymbol{\beta} = \left(\sum_{i=1}^{N} \{[\sigma_i^2(\mathbf{X}_i'\mathbf{X}_i)^{-1}]^{-1}\}\right)^{-1} \sum_{i=1}^{N} \{[\sigma_i^2(\mathbf{X}_i'\mathbf{X}_i)^{-1}]^{-1}\}\mathbf{b}_i$$

$$= \sum_{i=1}^{N} \mathbf{W}_i\mathbf{b}_i \quad = \quad \text{a matrix weighted average.}$$

Minimum Distance Estimation

The minimum distance estimator minimizes

$$\mathbf{q} = \bar{\mathbf{m}}(\boldsymbol{\beta})'\mathbf{A}\,\bar{\mathbf{m}}(\boldsymbol{\beta})$$

The estimator is

(1) Consistent

(2) Asymptotically normally distributed

(3) Has asymptotic covariance matrix

$$\text{Asy.Var}[\hat{\boldsymbol{\beta}}_{\mathbf{MD}}] = [\mathbf{G}(\boldsymbol{\beta})'\mathbf{A}\mathbf{G}(\boldsymbol{\beta})]^{-1}[\mathbf{G}(\boldsymbol{\beta})'\mathbf{A}\mathbf{V}\mathbf{A}\mathbf{G}(\boldsymbol{\beta})][\mathbf{G}(\boldsymbol{\beta})'\mathbf{A}\mathbf{G}(\boldsymbol{\beta})]^{-1}$$

Optimal Weighting Matrix

A is the Weighting Matrix of the minimum distance estimator.

Are some **A**'s better than others? (Yes)

Is there a best choice for **A**? Yes

The variance of the MDE is minimized when

$$\mathbf{A} = \{\text{Asy.Var}[\bar{\mathbf{m}}(\boldsymbol{\beta})]\}^{-1}$$

This defines the **generalized method of moments estimator**.

$$\mathbf{A} = \mathbf{V}^{-1}$$

GMM Estimation

$$\bar{\mathbf{m}}(\boldsymbol{\beta}) = \frac{1}{n}\Sigma_{i=1}^{n}\mathbf{m}_i(y_i, \mathbf{x_i}, \boldsymbol{\beta})$$

$$\text{Asy.Var}[\bar{\mathbf{m}}(\boldsymbol{\beta})] \text{ estimated with } \mathbf{W} = \frac{1}{n}\left(\frac{1}{n}\Sigma_{i=1}^{n}\mathbf{m}_i(y_i, \mathbf{x_i}, \boldsymbol{\beta})\mathbf{m}_i(y_i, \mathbf{x_i}, \boldsymbol{\beta})'\right)$$

The GMM estimator of $\boldsymbol{\beta}$ then minimizes

$$q = \left(\frac{1}{n}\Sigma_{i=1}^{n}\mathbf{m}_i(y_i, \mathbf{x_i}, \boldsymbol{\beta})\right)'\mathbf{W}^{-1}\left(\frac{1}{n}\Sigma_{i=1}^{n}\mathbf{m}_i(y_i, \mathbf{x_i}, \boldsymbol{\beta})\right).$$

$$\text{Est.Asy.Var}[\hat{\boldsymbol{\beta}}_{\mathbf{GMM}}] = [\mathbf{G'W^{-1}G}]^{-1}, \ \mathbf{G} = \frac{\partial\bar{\mathbf{m}}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}'}$$

GMM Estimation

Exactly identified GMM problems

When $\bar{\mathbf{m}}(\boldsymbol{\beta}) = \frac{1}{n}\Sigma_{i=1}^{n}\mathbf{m}_i(y_i, \mathbf{x_i}, \boldsymbol{\beta}) = \mathbf{0}$ is K equations in

K unknown parameters (the exactly identified case),

the weighting matrix in

$$q = \left(\frac{1}{n}\Sigma_{i=1}^{n}\mathbf{m}_i(y_i, \mathbf{x_i}, \boldsymbol{\beta})\right)'\mathbf{W}^{-1}\left(\frac{1}{n}\Sigma_{i=1}^{n}\mathbf{m}_i(y_i, \mathbf{x_i}, \boldsymbol{\beta})\right)$$

is irrelevant to the solution, since we can set exactly

$\bar{\mathbf{m}}(\boldsymbol{\beta}) = \mathbf{0}$ so q = 0. And, the asymptotic covariance matrix

(estimator) is the product of 3 square matrices and becomes

$[\mathbf{G}'\mathbf{W}^{-1}\mathbf{G}]^{-1} = \mathbf{G}^{-1}\mathbf{W}\mathbf{G}'^{-1}$

A Practical Problem

$\text{Asy.Var}[\bar{\mathbf{m}}(\boldsymbol{\beta})]$ estimated with

$$\mathbf{W} = \frac{1}{n}\left(\frac{1}{n}\Sigma_{i=1}^{n}\mathbf{m}_i(y_i, \mathbf{x_i}, \boldsymbol{\beta})\mathbf{m}_i(y_i, \mathbf{x_i}, \boldsymbol{\beta})'\right)$$

The GMM estimator of $\boldsymbol{\beta}$ then minimizes

$$q = \left(\frac{1}{n}\Sigma_{i=1}^{n}\mathbf{m}_i(y_i, \mathbf{x_i}, \boldsymbol{\beta})\right)'\mathbf{W}^{-1}\left(\frac{1}{n}\Sigma_{i=1}^{n}\mathbf{m}_i(y_i, \mathbf{x_i}, \boldsymbol{\beta})\right).$$

In order to compute $\mathbf{W}$, you need to know $\boldsymbol{\beta}$, and you are trying to estimate $\boldsymbol{\beta}$. How to proceed?

Typically two steps:

(1) Use $\mathbf{A} = \mathbf{I}$. Simple least squares, to get a preliminary estimator of $\boldsymbol{\beta}$. This is consistent, though not efficient.

(2) Compute the weighting matrix, then use GMM.

Inference

Testing hypotheses about the parameters:

- Wald test
- A counterpart to the likelihood ratio test

Testing the overidentifying restrictions

Testing Hypotheses

(1) Wald Tests in the usual fashion.

(2) A counterpart to likelihood ratio tests

GMM criterion is $q = \bar{\mathbf{m}}(\boldsymbol{\beta})'\mathbf{W}\,\bar{\mathbf{m}}(\boldsymbol{\beta})$

when restrictions are imposed on $\boldsymbol{\beta}$

q increases.

$$q_{restricted} - q_{unrestricted} \xrightarrow{d} chi-squared[J]$$

(The weighting matrix must be the same for both.)

(3) Testing the overidentifying restrictions: q would be 0 if exactly identified. q - 0 > 0 results from the overidentifying restrictions.

Application: Innovation

Bertschek and Lechner applied the GMM estimator to an analysis of the product innovation activity of 1,270 German manufacturing firms observed in five years, 1984 - 1988, in response to imports and foreign direct investment. [See Bertschek (1995).] The basic model to be estimated is a probit model based on the latent regression

$$y_{it}^* = \beta_1 + \sum_{k=2}^{8} x_{k,it}\beta_k + \varepsilon_{it}, \;\; y_{it} = \mathbf{1}\left(y_{it}^* > 0\right), i = 1,...,1270, t = 1984,...,1988.$$

where

y_{it} = 1 if a product innovation was realized by firm i in year t, 0 otherwise,
$x_{2,it}$ = Log of industry sales in DM,
$x_{3,it}$ = Import share = ratio of industry imports to (industry sales plus imports),
$x_{4,it}$ = Relative firm size = ratio of employment in business unit to employment in the industry (times 30),
$x_{5,it}$ = FDI share = Ratio of industry foreign direct investment to (industry sales, plus imports),
$x_{6,it}$ = Productivity = Ratio of industry value added to industry employment,
$x_{7,it}$ = Raw materials sector = 1 if the firm is in this sector,
$x_{8,it}$ = Investment goods sector = 1 if the firm is in this sector,

Application: Innovation

$$y_{it}^* = \beta_1 + \sum_{k=2}^{8} x_{k,it}\beta_k + \varepsilon_{it}\ ,\ y_{it} = \mathbf{1}\left(y_{it}^* > 0\right),$$

$i = 1,...,1270,\ t = 1984,...,1988.$

y_{it} = 1 if a product innovation was realized by German manufacturing firm i in year t, 0 otherwise,

$x_{2,it}$ = Log of industry sales in DM,

$x_{3,it}$ = Import share = ratio of industry imports to (industry sales plus imports),

$x_{4,it}$ = Relative firm size = ratio of employment in business unit to employment in the industry (times 30),

$x_{5,it}$ = FDI share = Ratio of industry foreign direct investment to (industry sales, plus imports),

$x_{6,it}$ = Productivity = Ratio of industry value added to industry employment,

$x_{7,it}$ = Raw materials sector = 1 if the firm is in this sector,

$x_{8,it}$ = Investment goods sector = 1 if the firm is in this sector

Application: Multivariate Probit Model

5 - variate Probit Model

$$y_{it}^* = \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}, \quad y_{it} = 1[y_{it}^* > 0]$$

$$\log L_i = \int_{-\infty}^{\boldsymbol{\beta}'\mathbf{x}_{i5}} \int_{-\infty}^{\boldsymbol{\beta}'\mathbf{x}_{i4}} \int_{-\infty}^{\boldsymbol{\beta}'\mathbf{x}_{i3}} \int_{-\infty}^{\boldsymbol{\beta}'\mathbf{x}_{i2}} \int_{-\infty}^{\boldsymbol{\beta}'\mathbf{x}_{i1}} \phi_5[\{(2y_{it}-1)s_{it}, t=1,\ldots,5\}, \boldsymbol{\Sigma}]ds_{i1}ds_{i2}ds_{i3}ds_{i4}ds_{i5}$$

Requires 5 dimensional integration of the joint normal density. Very hard!

But, $E[y_{it} \mid \mathbf{x}_{it}] = \Phi(\boldsymbol{\beta}'\mathbf{x}_{it})$.

Orthogonality Conditions: $E[\{y_{it} - \Phi(\boldsymbol{\beta}'\mathbf{x}_{it})\}\mathbf{x}_{it} = \mathbf{0}$

Moment Equations: $$\frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} \{y_{i1} - \Phi(\boldsymbol{\beta}'\mathbf{x}_{i1})\}\mathbf{x}_{i1} \\ \{y_{i2} - \Phi(\boldsymbol{\beta}'\mathbf{x}_{i2})\}\mathbf{x}_{i2} \\ \{y_{i3} - \Phi(\boldsymbol{\beta}'\mathbf{x}_{i3})\}\mathbf{x}_{i3} \\ \{y_{i4} - \Phi(\boldsymbol{\beta}'\mathbf{x}_{i4})\}\mathbf{x}_{i4} \\ \{y_{i5} - \Phi(\boldsymbol{\beta}'\mathbf{x}_{i5})\}\mathbf{x}_{i5} \end{bmatrix} = \mathbf{0}$$ 40 equations in 8 parameters.

Pooled Probit – Ignoring Correlation

Table 1. Estimated Pooled Probit Model

		Estimated Standard Errors				Marginal Effects		
Variable	Estimate[a]	se(1)[b]	se(2)[c]	se(3)[d]	se(4)[e]	Partial[f]	Std. Err.	
Constant	-1.960**	0.21	0.230	0.377	0.373	—	—	
log Sales	0.177**	0.025	0.0222	0.0375	0.0358	0.0683	0.0138**	
Rel Size	1.072**	0.21	0.142	0.306	0.269	0.413	0.103**	
Imports	1.134**	0.15	0.151	0.246	0.243	0.437	0.0938**	
FDI	2.853**	0.47	0.402	0.679	0.642	1.099	0.247**	
Prod.	-2.341**	1.10	0.715	1.300	1.115	-0.902	0.429*	
Raw Mtl	-0.279**	0.097	0.0807	0.133	0.126	-0.110[g]	0.0503*	
Inv Good	0.188**	0.040	0.0392	0.0630	0.0628	0.0723[g]	0.0241**	

[a] Recomputed. Only two digits were reported in the earlier paper.

[b] Obtained from results in Bertschek and Lechner, Table 10.

[c] Square roots of the diagonals of the negative inverse of the Hessian

[d] Based on the Avery et al. GMM estimator

[e] Based on the cluster estimator.

[f] Coefficient scaled by the density evaluated at the sample means

[g] Computed as the difference in the fitted probability with the dummy variable equal to one then zero.

* Indicates significant at 95% level, ** indicates significant at 99% level based on a two tailed test. Significance tests based on se(4).

Random Effects: $\Sigma=(1-\rho)I+\rho ii'$

Table 2. Estimated Random Effects Models

	Random Effects			
	Quadrature Estimator		Simulation Estimator	
Variable	Estimate	Std.Error	Estimate	Std.Error
Constant	-2.839**	0.533	-2.884**	0.543
log Sales	0.244**	0.0522	0.249**	0.0510
Rel Size	1.522**	0.257	1.452**	0.281
Imports	1.779**	0.360	1.796**	0.360
FDI	3.652**	0.870	3.724**	0.831
Prod.	-2.307	1.911	-2.321**	0.151
Raw Mtl	-0.477*	0.202	-0.469*	0.186
Inv Good	0.331**	0.0952	0.331**	0.0915
ρ	0.578**	0.0189	0.578**[a]	0.0231

[a]Based on estimated standard deviation of the random constant of 1.1707 with estimated standard error of 0.01865.

* Indicates significant at 95% level, ** indicates significant at 99% level based on a two tailed test.

Unrestricted Correlation Matrix

Table 3. Estimated Constrained Multivariate Probit Model

Coefficients	β	Std. Error	BL GMM[a]	Std. Error
Constant	-1.797**	0.341	-1.74**	0.37
log Sales	0.154**	0.0334	0.15**	0.034
Relative size	0.953**	0.160	0.95**	0.20
Imports	1.155**	0.228	1.14**	0.24
FDI	2.426**	0.573	2.59**	0.59
Productivity	-1.578	1.216	-1.91*	0.82
Raw Material	-0.292**	0.130	-0.28*	0.12
Investment Goods	0.224**	0.0605	0.21**	0.063
Estimated Correlations				
1984,1985	0.460**	0.0301		
1984,1986	0.599**	0.0323		
1985,1986	0.643**	0.0308		
1984,1987	0.540**	0.0308		
1985,1987	0.546**	0.0348		
1986,1987	0.610**	0.0322		
1984,1988	0.483**	0.0364		
1985,1988	0.446**	0.0380		
1986,1988	0.524**	0.0355		
1987,1988	0.605**	0.0325		

Estimated Correlation Matrix

	1984	**1985**	**1986**	**1987**	**1988**
1984	1.000				
1985	0.460	1.000			
1986	0.599	0.643	1.000		
1987	0.540	0.546	0.610	1.000	
1988	0.483	0.446	0.524	0.605	1.000

[a]Estimates are BL's WNP-joint uniform estimates with $k = 880$. Estimates are from their Table 9, standard errors from their Table 10.

* Indicates significant at 95% level, ** indicates significant at 99% level based on a two tailed test.