

GMM is a fundamental concept in graduate-level econometrics
Very popular because it nests many common estimators:

Very popular because it nests many common estimators:

OLS

Very popular because it nests many common estimators:

- OLS
- IV and 2SLS

Very popular because it nests many common estimators:

- OLS
- IV and 2SLS
- Nonlinear least squares (NLLS)

Very popular because it nests many common estimators:

- OLS
- IV and 2SLS
- Nonlinear least squares (NLLS)

Can also (asymptotically) replicate MLE when using score-based moment conditions

Very popular because it nests many common estimators:

- OLS
- IV and 2SLS
- Nonlinear least squares (NLLS)

Can also (asymptotically) replicate MLE when using score-based moment conditions

Unlike MLE, GMM need not assume exact distribution of errors

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

Assume 
$$\mathbb{E}[\varepsilon \mid \mathbf{x}] = 0$$
 (conditional independence)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

Assume  $\mathbb{E}[\varepsilon \mid \mathbf{x}] = 0$  (conditional independence)

Then we can form a system of 3 equations and 3 unknowns

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

Assume  $\mathbb{E}[\varepsilon \mid \mathbf{x}] = 0$  (conditional independence)

Then we can form a system of 3 equations and 3 unknowns

Note: If we were doing MLE, we would need to assume, e.g.  $\varepsilon \stackrel{iid}{\sim} N(0, \sigma^2)$ 

## OLS population moment conditions:

$$\mathbb{E}[arepsilon]=0$$

$$\mathbb{E}[\varepsilon x_1] = 0$$

$$\mathbb{E}[\varepsilon x_2] = 0$$

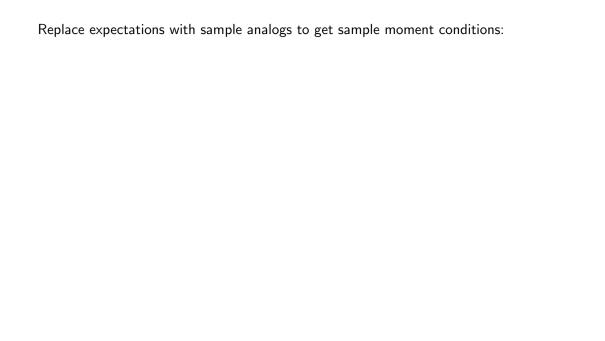
Rewriting in terms of parameters  $(\beta_0, \beta_1, \beta_2)$ :

$$2[(y - \beta_0 - \beta_1 x_1 - \beta_0 x_0)] -$$

$$\mathbb{E}[(y-\beta_0-\beta_1x_1-\beta_2x_2)]=0$$

$$\mathbb{E}[(y - \beta_0 - \beta_1 x_1 - \beta_2 x_2)x_1] = 0$$

$$\mathbb{E}[(y - \beta_0 - \beta_1 x_1 - \beta_2 x_2) x_2] = 0$$



Replace expectations with sample analogs to get sample moment conditions:

$$g(\beta) = \begin{cases} \frac{1}{N} \sum_{i=1}^{N} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2}) &= 0\\ \frac{1}{N} \sum_{i=1}^{N} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2}) x_{i1} &= 0\\ \frac{1}{N} \sum_{i=1}^{N} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2}) x_{i2} &= 0 \end{cases}$$

Replace expectations with sample analogs to get sample moment conditions:

$$g(\beta) = \begin{cases} \frac{1}{N} \sum_{i=1}^{N} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2}) &= 0\\ \frac{1}{N} \sum_{i=1}^{N} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2}) x_{i1} &= 0\\ \frac{1}{N} \sum_{i=1}^{N} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2}) x_{i2} &= 0 \end{cases}$$

Estimate by exactly-identified GMM:

$$\hat{oldsymbol{eta}} = rg \min_{oldsymbol{eta}} J(oldsymbol{eta})$$

where

$$J(\beta) = Ng(\beta)'g(\beta)$$

Replace expectations with sample analogs to get sample moment conditions:

$$g(\beta) = \begin{cases} \frac{1}{N} \sum_{i=1}^{N} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2}) &= 0\\ \frac{1}{N} \sum_{i=1}^{N} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2}) x_{i1} &= 0\\ \frac{1}{N} \sum_{i=1}^{N} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2}) x_{i2} &= 0 \end{cases}$$

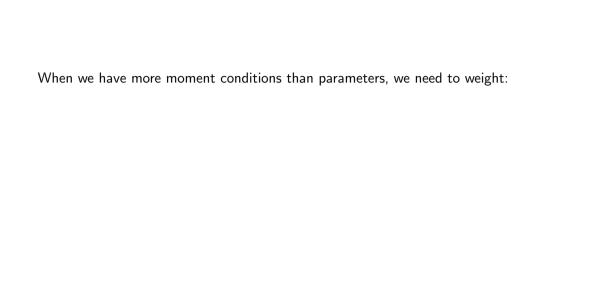
Estimate by exactly-identified GMM:

$$\hat{oldsymbol{eta}} = rg \min_{oldsymbol{eta}} J(oldsymbol{eta})$$

where

$$J(\beta) = Ng(\beta)'g(\beta)$$

For OLS, this objective function has a closed form solution:  $(X'X)^{-1}X'y$ 



When we have more moment conditions than parameters, we need to weight:

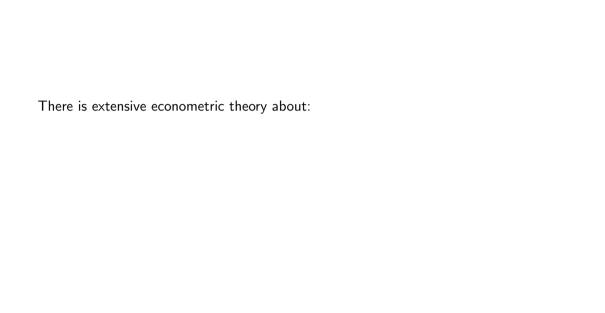
 $\hat{oldsymbol{eta}} = \mathop{\sf arg\,min}\limits_{oldsymbol{eta}} J\left(oldsymbol{eta}, \hat{oldsymbol{\mathsf{W}}}
ight)$ 

When we have more moment conditions than parameters, we need to weight:

$$\hat{oldsymbol{eta}} = rg \min_{oldsymbol{eta}} J\left(oldsymbol{eta}, \hat{oldsymbol{W}}
ight)$$

where

$$J(eta) = Ng(eta)'\hat{\mathbf{W}}(eta)g(eta)$$



There is extensive econometric theory about:

ullet The optimal weighting matrix  $\hat{f W}$ 

There is extensive econometric theory about:
$ullet$ The optimal weighting matrix $\hat{f W}$ (weight each moment by inverse of its variance)

There is extensive econometric theory about:
$ullet$ The optimal weighting matrix $\hat{f W}$ (weight each moment by inverse of its variance)
• Asymptotic properties of the GMM estimator (they're good)

 $\label{eq:continuous} \textbf{Another approach to OLS:}$ 

Previously: solved K equations of  $\mathbb{E}\left[\varepsilon X_{k}\right]=0$  and  $\mathbb{E}\left[\varepsilon\right]=0$ 

Previously: solved K equations of  $\mathbb{E}\left[\varepsilon X_{k}\right]=0$  and  $\mathbb{E}\left[\varepsilon\right]=0$ 

Previously: solved K equations of  $\mathbb{E}\left[\varepsilon X_{k}\right]=0$  and  $\mathbb{E}\left[\varepsilon\right]=0$ 

Alternative: attempt to match y to  $X\beta$  for every observation (as closely as possible)

• Set  $g = y - X\beta$ 

Previously: solved K equations of  $\mathbb{E}\left[\varepsilon X_{k}\right]=0$  and  $\mathbb{E}\left[\varepsilon\right]=0$ 

- Set  $g = y X\beta$
- $\bullet$  N "moment conditions" (residuals) and  ${\cal K}+1$  parameters

Previously: solved K equations of  $\mathbb{E}\left[\varepsilon X_{k}\right]=0$  and  $\mathbb{E}\left[\varepsilon\right]=0$ 

- Set  $g = y X\beta$
- ullet N "moment conditions" (residuals) and  ${\cal K}+1$  parameters
- ullet Use  ${\it N} imes {\it N}$  Identity matrix for  ${\it W}$

Previously: solved K equations of  $\mathbb{E}\left[\varepsilon X_{k}\right]=0$  and  $\mathbb{E}\left[\varepsilon\right]=0$ 

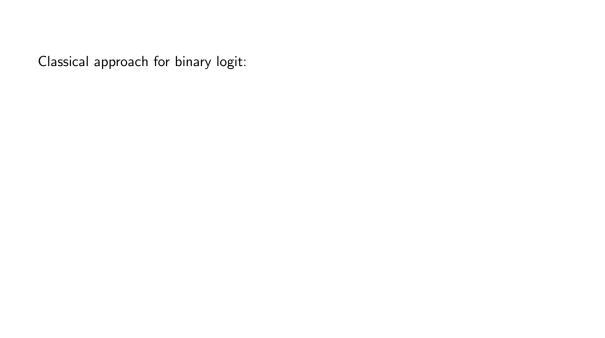
- Set  $g = y X\beta$
- ullet N "moment conditions" (residuals) and K+1 parameters
- Use  $N \times N$  Identity matrix for W
- This is precisely OLS

Previously: solved K equations of  $\mathbb{E}\left[\varepsilon X_{k}\right]=0$  and  $\mathbb{E}\left[\varepsilon\right]=0$ 

Alternative: attempt to match y to  $X\beta$  for every observation (as closely as possible)

- Set  $g = y X\beta$
- N "moment conditions" (residuals) and K+1 parameters
- Use N × N Identity matrix for W
- This is precisely OLS

In practice, this approach often has better computational properties



Classical approach for binary logit:

$$g(\beta) = \begin{cases} \frac{1}{N} \sum_{i=1}^{N} \left[ y_i - \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})} \right] &= 0\\ \frac{1}{N} \sum_{i=1}^{N} \left[ y_i - \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})} \right] x_{i1} &= 0\\ \frac{1}{N} \sum_{i=1}^{N} \left[ y_i - \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})} \right] x_{i2} &= 0 \end{cases}$$

$$\int \frac{1}{N}$$

where  $y_i \in \{0, 1\}$ 

Classical approach for binary logit:

$$g(\beta) = \begin{cases} \frac{1}{N} \sum_{i=1}^{N} \left[ y_i - \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})} \right] &= 0\\ \frac{1}{N} \sum_{i=1}^{N} \left[ y_i - \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})} \right] x_{i1} &= 0\\ \frac{1}{N} \sum_{i=1}^{N} \left[ y_i - \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})} \right] x_{i2} &= 0 \end{cases}$$

Same formula for J as in the OLS case

where  $v_i \in \{0, 1\}$ 

Alternative approach: use g=y-P where  $y\in\{0,1\}$  and  $P\in(0,1)$ 

Alternative approach: use g = y - P where  $y \in \{0,1\}$  and  $P \in (0,1)$ 

Again: N "moment conditions" (residuals) and K+1 parameters to be estimated

Alternative approach: use g = y - P where  $y \in \{0, 1\}$  and  $P \in (0, 1)$ 

This is known as Nonlinear Least Squares (NLLS)

Again: N "moment conditions" (residuals) and K+1 parameters to be estimated

Again. N moment conditions (residuais) and N + 1 parameters to be estimated