GMM is a fundamental concept in graduate-level econometrics

GMM is a fundamental concept in graduate-level econometrics

Very popular because it nests many common estimators:

GMM is a fundamental concept in graduate-level econometrics

Very popular because it nests many common estimators:

- OLS

GMM is a fundamental concept in graduate-level econometrics

Very popular because it nests many common estimators:

- OLS
- IV and 2SLS

GMM is a fundamental concept in graduate-level econometrics

Very popular because it nests many common estimators:

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

GMM is a fundamental concept in graduate-level econometrics

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

GMM is a fundamental concept in graduate-level econometrics

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

Consider a simple regression model:

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

Consider a simple regression model:

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

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

Consider a simple regression model:

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

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

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

Consider a simple regression model:

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

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. $\varepsilon \overset{iid}{\sim} N(0, \sigma^2)$

OLS population moment conditions:

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

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

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

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

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

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

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

Replace expectations with sample analogs to get sample moment conditions:

Replace expectations with sample analogs to get sample moment conditions:

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

Replace expectations with sample analogs to get sample moment conditions:

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

Estimate by exactly-identified GMM:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} J\left(\boldsymbol{\beta}\right)$$

where

$$J\left(\boldsymbol{\beta}\right) = Ng\left(\boldsymbol{\beta}\right)' g\left(\boldsymbol{\beta}\right)$$

Replace expectations with sample analogs to get sample moment conditions:

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

Estimate by exactly-identified GMM:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} J(\boldsymbol{\beta})$$

where

$$J(\boldsymbol{\beta}) = Ng(\boldsymbol{\beta})' g(\boldsymbol{\beta})$$

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:

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

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} J\left(\boldsymbol{\beta}, \hat{\mathbf{W}}\right)$$

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

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} J\left(\boldsymbol{\beta}, \hat{\mathbf{W}}\right)$$

where

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

There is extensive econometric theory about:

There is extensive econometric theory about:

- The optimal weighting matrix $\hat{\mathbf{W}}$

There is extensive econometric theory about:

- The optimal weighting matrix $\hat{\mathbf{W}}$ (weight each moment by inverse of its variance)

There is extensive econometric theory about:

- The optimal weighting matrix $\hat{\mathbf{W}}$ (weight each moment by inverse of its variance)
- Asymptotic properties of the GMM estimator (they're good)

Another approach to OLS:

Another approach to OLS:

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

Another approach to OLS:

Previously: solved K equations of $\mathbb{E}[\varepsilon X_k] = 0$ and $\mathbb{E}[\varepsilon] = 0$

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

Another approach to OLS:

Previously: solved K equations of $\mathbb{E}[\varepsilon X_k] = 0$ and $\mathbb{E}[\varepsilon] = 0$

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

- Set $g = y - X\beta$

Another approach to OLS:

Previously: solved K equations of $\mathbb{E}[\varepsilon X_k] = 0$ and $\mathbb{E}[\varepsilon] = 0$

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

Another approach to 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 $N \times N$ Identity matrix for $\mathbf{W}$

Another approach to OLS:

Previously: solved K equations of $\mathbb{E}[\varepsilon X_k] = 0$ and $\mathbb{E}[\varepsilon] = 0$

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 $N \times N$ Identity matrix for $\mathbf{W}$
- This is precisely OLS

Another approach to OLS:

Previously: solved K equations of $\mathbb{E}[\varepsilon X_k] = 0$ and $\mathbb{E}[\varepsilon] = 0$

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 $N \times N$ Identity matrix for $\mathbf{W}$
- This is precisely OLS

In practice, this approach often has better computational properties

Classical approach for binary logit:

Classical approach for binary logit:

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

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

Classical approach for binary logit:

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

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

Same formula for J as in the OLS case

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 $P \in (0, 1)$

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

This is known as Nonlinear Least Squares (NLLS)