ECON 480
LECTURE 6: GMM & EL

Ivan A. Canay
Northwestern University

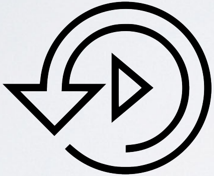
Past & Future

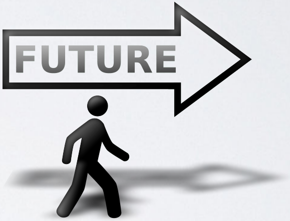
Last Class

- Efficiency of 2SLS
- Weak IV
- LATE

TODAY

- Generalized Method of Moments
- Empirical Likelihood
- Asymptotic Properties





Generalized Method of Moments

- Let (Y, X, Z, U) be a random vector where Y and U take values in $\mathbf{R}$, X takes values in $\mathbf{R}^{k+1}$, and Z takes values in $\mathbf{R}^{\ell+1}$.
  Assume $\ell \geqslant k$, 1 $E[ZU] = 0$, 2 $E[ZX'] < \infty$, and 3 $\text{rank}(E[ZX']) = k+1$.
- Let β be such that
$$Y = X'\beta + U\,.$$
- Using the fact that $U = Y - X'\beta$ and $E[ZU] = 0$, we see that β solves the system of equations
$$E[Z(Y - X'\beta)] = 0\,.$$
- Since $\ell \geqslant k$, this may be an over-determined system of equations and today we focus on the case $\ell > k$. This is called over-identified: there are $\ell - k = r$ more moment restrictions than parameters.
- We usually call r the number of over-identifying restrictions.

Generalized Method of Moments

- The above is a **special case** of a more general class of moment condition models.
- Let $m(Y, X, Z, \beta)$ be an $\ell + 1$ dimensional function of a $k + 1$ **dimensional parameter** β such that

$$E[m(Y, X, Z, \beta)] = 0 \; .$$

- **Linear model**: $m(Y, X, Z, \beta) = Z(Y - X'\beta)$.
- In econometrics, this class of models are called moment condition models. In the statistics literature, these are known as estimating equations.
- **Today**: two related ways to estimate β in moment condition models: GMM and EL.

The GMM Estimator

- Let (Y, X, Z, U) be as described and let P be the marginal distribution of (Y, X, Z).
- Let $(Y_1, X_1, Z_1), \ldots, (Y_n, X_n, Z_n)$ be an i.i.d. sequence of random variables with distribution P.
- Define the **sample analog** of $E[m(Y, X, Z, \beta)]$ by

$$\bar{m}_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} m_i(\beta) = \frac{1}{n}\sum_{i=1}^{n} Z_i(Y_i - X_i'\beta) = \frac{1}{n}\mathbb{Z}'(\mathbb{Y} - \mathbb{X}\beta)\ ,$$

  where in what follows we will use the notation $m_i(\beta) = m(Y_i, X_i, Z_i, \beta)$.
- The **method of moments** estimator for β is defined as the parameter value which sets $\bar{m}_n(\beta) = 0$.
- This is generally not possible when $\ell > k$ as there are more equations than free parameters.
- **Generalized method of moments (GMM)**: define an estimator that sets $\bar{m}_n(\beta)$ “close” to zero, given a notion of “distance”.

The GMM Estimator

- Let Λ_n be an $(\ell+1)\times(\ell+1)$ matrix such that $\Lambda_n \xrightarrow{P} \Lambda$ for a symmetric pd matrix Λ and define

$$Q_n(\beta) = n\,\bar{m}_n(\beta)'\Lambda_n\bar{m}_n(\beta)\ .$$

- $Q_n(\beta)$ is a non-negative measure of the distance between the vector $\bar{m}_n(\beta)$ and the origin.

- If $\Lambda_n = \mathbb{I}$, then $Q_n(\beta) = n\,|\bar{m}_n(\beta)|^2$, the square of the Euclidean norm, scaled by the sample size n.

- **GMM estimator**: defined as the value that minimizes $Q_n(\beta)$,

$$\hat{\beta}_n = \underset{b\in\mathbf{R}^{k+1}}{\operatorname{argmin}}\, Q_n(b)\ . \tag{1}$$

- If $k=\ell$: $\bar{m}_n(\hat{\beta}_n) = 0$ and the GMM estimator is the method of moments estimator.

GMM: First Order Condition

$$Q_n(\beta) = n\, \bar{m}_n(\beta)' \Lambda_n \bar{m}_n(\beta) \qquad \text{with} \qquad \bar{m}_n(\beta) = \frac{1}{n} \mathbb{Z}' (\mathbb{Y} - \mathbb{X}\beta) \,,$$

QUESTIONS?

GMM: consistency

$$\hat{\beta}_n = \left(\left(\mathbb{Z}'\mathbb{X}\right)' \Lambda_n \left(\mathbb{Z}'\mathbb{X}\right) \right)^{-1} \left(\mathbb{Z}'\mathbb{X}\right)' \Lambda_n \left(\mathbb{Z}'\mathbb{Y}\right) \; .$$

GMM: Asymptotic Normality

$$\sqrt{n}(\hat{\beta}_n - \beta) = \left(\left(\frac{1}{n}\sum_{i=1}^{n} Z_i X_i'\right)' \Lambda_n \left(\frac{1}{n}\sum_{i=1}^{n} Z_i X_i'\right)\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} Z_i X_i'\right)' \Lambda_n \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i U_i\right) .$$

Let $\Sigma = E[ZX']$ and assume $\Omega = E[ZZ'U^2]$ is well defined and invertible.

GMM: Efficient Weighting Matrix

$$\mathbb{V} = (\Sigma'\Lambda\Sigma)^{-1}(\Sigma'\Lambda\Omega\Lambda\Sigma)(\Sigma'\Lambda\Sigma)^{-1} .$$

- The optimal weigh matrix Λ^* is the one which minimizes $\mathbb{V}$.
- This turn out to be $\Lambda^* = \Omega^{-1}$ and yields the **efficient GMM** estimator
$$\hat{\beta}_n = \left((\mathbb{Z}'\mathbb{X})' \, \Omega^{-1} \, (\mathbb{Z}'\mathbb{X}) \right)^{-1} (\mathbb{Z}'\mathbb{X})' \, \Omega^{-1} \, (\mathbb{Z}'\mathbb{Y}) \ ,$$
which satisfies
$$\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{d} N\left(0, (\Sigma'\Omega^{-1}\Sigma)^{-1}\right) .$$
- Ω can be estimated consistently: for any $\hat{\Omega}_n \xrightarrow{P} \Omega$, we call $\hat{\beta}_n$ the efficient GMM estimator
- By "efficient", we mean that this estimator has the smallest asymptotic variance in the class of GMM estimators with this set of moment conditions. This is a **weak** concept of optimality.
- **Gary Chamberlain (1948-2020)**: showed that the GMM estimator is semiparametrically efficient.

GMM: Estimation of Weighting Matrix

- Given any weight matrix Λ_n with the properties previously discussed, the GMM estimator $\hat{\beta}_n$ is consistent yet *inefficient*.

- **Example**: set $\Lambda_n = \mathbb{I}$. In the linear model, a better choice is $\Lambda_n = (\mathbb{Z}'\mathbb{Z})^{-1}$, which leads to TSLS

$$\begin{aligned}
\hat{\beta}_n &= \left((\mathbb{Z}'\mathbb{X})'\Lambda_n(\mathbb{Z}'\mathbb{X})\right)^{-1}(\mathbb{Z}'\mathbb{X})'\Lambda_n(\mathbb{Z}'\mathbb{Y}) \\
&= \left((\mathbb{Z}'\mathbb{X})'(\mathbb{Z}'\mathbb{Z})^{-1}(\mathbb{Z}'\mathbb{X})\right)^{-1}(\mathbb{Z}'\mathbb{X})'(\mathbb{Z}'\mathbb{Z})^{-1}(\mathbb{Z}'\mathbb{Y}) \\
&= (\mathbb{X}'\mathbb{P}_Z\mathbb{X})^{-1}(\mathbb{X}'\mathbb{P}_Z\mathbb{Y}) \ ,
\end{aligned}$$

  where, as before, $\mathbb{P}_Z = \mathbb{Z}(\mathbb{Z}'\mathbb{Z})^{-1}\mathbb{Z}'$ is a projection matrix.

- To estimate the efficient weighting matrix, we conduct two steps.
  1. Get a consistent (preliminary) estimator of β
  2. Use that preliminary estimator to form residuals and estimate Λ^*

GMM: Estimation of Weighting Matrix

- Let $\hat{\beta}_n$ be a first-step estimator of β. Let $\hat{U}_i = Y_i - X_i'\hat{\beta}_n$ and $\hat{m}_i = Z_i\hat{U}_i$. Construct,

$$\hat{m}_i^* = \hat{m}_i - \frac{1}{n}\sum_{i=1}^{n}\hat{m}_i \, .$$

Define

$$\Lambda_n^* = \left(\frac{1}{n}\sum_{i=1}^{n}\hat{m}_i^*\hat{m}_i^{*\prime}\right)^{-1} \quad .$$

- It can be shown that $\Lambda_n^* \xrightarrow{P} \Omega^{-1}$ and GMM using the weighting matrix above is asymptotically efficient. This is typically referred to as the **efficient two-step GMM estimator**.

- A less preferred alternative choice is to use

$$\Lambda_n = \left(\frac{1}{n}\sum_{i=1}^{n}\hat{m}_i\hat{m}_i'\right)^{-1} \quad ,$$

which uses the uncentered moment conditions. Performs poorly for testing hypotheses.

GMM: Overidentification Test

- **Test Statistic**:

$$Q_n^*(\hat{\beta}_n) = n\,\bar{m}_n(\hat{\beta}_n)'\Lambda_n^*\bar{m}_n(\hat{\beta}_n)\ .$$

- Under correct model specification, it can be shown that

$$Q_n^*(\hat{\beta}_n) \xrightarrow{d} \chi^2_{\ell-k}\ ,$$

as $n \to \infty$, where $\hat{\beta}_n$ is the efficient two-step GMM estimator.

- In addition, $Q_n^*(\hat{\beta}_n) \to \infty$ if $E[m(Y,X,Z,\beta)] \neq 0$ for all $\beta \in \mathbf{R}^{k+1}$.

- The degrees of freedom of the asymptotic distribution are the number of overidentifying restrictions.

- **Overidentification test**: reject the null hypothesis "there exist $\beta \in \mathbf{R}^{k+1}$ such that the model holds" when $Q_n^*(\hat{\beta}_n)$ exceeds the $1-\alpha$ quantile of $\chi^2_{\ell-k}$.

QUESTIONS?

Empirical Likelihood

- **Empirical Likelihood (EL)**: a data-driven nonparametric method of estimation and inference for moment restriction models. It does not require weight matrix estimation like GMM and is invariant to nonsingular linear transformations of the moment conditions. Developed by Art Owen.

- It is basically a non-parametric analog of Maximum Likelihood for the model

$$E[m(Y, X, Z, \beta)] = 0 .$$

- EL may be viewed as parametric inference in moment condition models, using a data-determined parametric family of distributions. The parametric family is a multinomial distribution on the observed values $(Y_1, X_1, Z_1), \ldots, (Y_n, X_n, Z_n)$.

- The parametric family has $n-1$ parameters. Having the number of parameters grow as quickly as the sample size makes empirical likelihood *very different* than parametric likelihood.

Empirical Likelihood

- The multinomial distribution with probability p_i at each observation satisfies the moment condition **iff**

$$\sum_{i=1}^{n} p_i m_i(\beta) = 0 \, .$$

- **EL estimator**: value of β maximizing the multinomial log-likelihood subject to this restriction.

- **Empirical likelihood function**:

$$\mathcal{R}_n(b) \equiv \max_{p_1,\dots,p_n} \left\{ \prod_{i=1}^{n} np_i \,\middle|\, p_i > 0; \sum_{i=1}^{n} p_i = 1; \sum_{i=1}^{n} p_i m_i(b) = 0 \right\} \, .$$

- **Lagrangian for the empirical log-likelihood**:

$$\mathcal{L}(b, p_1, \dots, p_n, \lambda, \kappa) = \sum_{i=1}^{n} \log(np_i) - \kappa \left(\sum_{i=1}^{n} p_i - 1 \right) - n\lambda' \sum_{i=1}^{n} p_i m_i(b)$$

where κ and λ are Lagrange multipliers.

EL: First Order Conditions

$$\mathcal{L}(b, p_1, \ldots, p_n, \lambda, \kappa) = \sum_{i=1}^{n} \log(np_i) - \kappa\left(\sum_{i=1}^{n} p_i - 1\right) - n\lambda' \sum_{i=1}^{n} p_i m_i(b) \; .$$

EL: Dual Representation

log-likelihood $\sum_{i=1}^{n} \log(np_i)$ with $p_i(b) = \frac{1}{n}\frac{1}{1+\lambda(b)' m_i(b)}$.

QUESTIONS?

Asymptotic Properties of EL

- It turns out that the limit distribution of the EL estimator is the same as that of efficient GMM,

$$\sqrt{n}(\tilde{\beta}_n - \beta) \xrightarrow{d} N(0, (\Sigma'\Omega^{-1}\Sigma)^{-1}) .$$

- **First order conditions**: interesting to compare the FOCs of EL and two-step GMM.

- To do so, let's re-write the first order condition of the EL estimator, paying specific attention to the linear model where $m_i(\beta) = Z_i(Y_i - X_i'\beta)$. Define

$$M_i(\beta) = -\frac{\partial}{\partial \beta'} m_i(\beta) = Z_i X_i' ,$$

  and let

$$\Sigma(\beta) = E\left[M_i(\beta)\right] = E[Z_i X_i']$$
$$\Omega(\beta) = E\left[m_i(\beta) m_i(\beta)'\right] = E\left[Z_i Z_i' U_i^2\right] .$$

EL vs GMM: First order conditions

The EL estimators $(\tilde{\beta}_n, \tilde{\lambda}_n)$ jointly solve

$$\frac{1}{n}\sum_{i=1}^{n}\frac{m_i(\tilde{\beta}_n)}{1+\tilde{\lambda}_n' m_i(\tilde{\beta}_n)} = 0 \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n}\frac{M_i(\tilde{\beta}_n)'\tilde{\lambda}_n}{1+\tilde{\lambda}_n' m_i(\tilde{\beta}_n)} = 0$$

EL vs GMM: First order conditions

EL first order condition

$$\tilde{\Sigma}_n(\tilde{\beta}_n)'\tilde{\Omega}_n^{-1}(\tilde{\beta}_n)\bar{m}_n(\tilde{\beta}_n) = 0$$

EL vs GMM: First order conditions

$$\tilde{\Sigma}_n(\tilde{\beta}_n)'\tilde{\Omega}_n^{-1}(\tilde{\beta}_n)\bar{m}_n(\tilde{\beta}_n) = 0 \quad \text{versus} \quad \hat{\Sigma}_n(\hat{\beta}_n)'\hat{\Omega}_n^{-1}\bar{m}_n(\hat{\beta}_n) = 0\,.$$

- We can now see that EL and GMM have very similar first order conditions.
- GMM uses $\hat{\Omega}_n$ as a consistent estimator of Ω based on a preliminary estimator of β.
- It is not surprising then that these two estimators are first order equivalent.
- However, using a different estimator of the Jacobian matrix Σ and the absence of a preliminary estimator for Ω gives EL some favorable second order properties relative to GMM.
- The cost of this is some additional computational complexity.
- This is a topic we cover in Econ 481.

THE END!