Extremum Estimators

Giselle Montamat

Harvard University

Spring 2020

Identification meets inference/estimation...
...or ECON2140 meets ECON2120 :)

Model

Identifying assumptions $\Downarrow \Uparrow$ *(1) Identification*

Population distribution of observable variables

Sampling $\Downarrow \Uparrow$ *(2) Estimation, Inference*

Observations

We'll assume n iid obervations $D = (D_1, ..., D_n)$. We observe D and want to infer/estimate F or features of F.

Identification meets inference/estimation...
...or ECON2140 meets ECON2120 :)

So how does one go about estimating the identified objects from the models we've seen?

$\Rightarrow$ Most require estimating some conditional expectation.

- **Nonparametric estimation**:
  - Plug-in principle estimation:
    - the sample analog of an expectation is a mean
    - great if X_i discrete and not many support points:

$$\hat{E}[Y_i|X_i = x] = \frac{\frac{1}{n}\sum Y_i 1\{X_i = x\}}{\frac{1}{n}\sum 1\{X_i = x\}}$$

    - but issues when X_i has many support points relative to n or is continuous
  - Other nonparametric estimators:
    - basically, the plug-in estimation version for continuous variables
    - coming to theaters soon...

Identification meets inference/estimation...
...or ECON2140 meets ECON2120 :)

- **Parametric estimation**:
  - extremum estimators
    - ★ m-estimation. E.g.: MLE, OLS, NLS
    - ★ GMM. E.g.: OLS, IV (2SLS)
      $\Rightarrow$ Efficient GMM E.g.: OLS, IV (2SLS) if homoskedasticity
      $\Rightarrow$ Two-step GMM estimators
    - ★ Minimum distance. E.g.: IV

We'll look at properties of these estimators:

- Consistency
- Asymptotic distribution $\Rightarrow$ particularly interested in the **s.e.**

Notation convention I use for derivatives:

$$\underbrace{\beta}_{k\times 1} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \quad ; \quad \underbrace{m(\beta)}_{1\times 1} \quad ; \quad \underbrace{g(\beta)}_{L\times 1} = \begin{bmatrix} g_1(\beta) \\ \vdots \\ g_L(\beta) \end{bmatrix}$$

Then:

$$\underbrace{\frac{\partial m(\beta)}{\partial \beta}}_{k\times 1} = \begin{bmatrix} \frac{\partial m}{\partial \beta_1} \\ \vdots \\ \frac{\partial m}{\partial \beta_k} \end{bmatrix} \; ; \; \underbrace{\frac{\partial^2 m(\beta)}{\partial \beta \partial \beta'}}_{k\times k} = \begin{bmatrix} \frac{\partial^2 m}{\partial \beta_1 \partial \beta_1} & \cdots & \frac{\partial^2 m}{\partial \beta_1 \partial \beta_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 m}{\partial \beta_k \partial \beta_1} & \cdots & \frac{\partial^2 m}{\partial \beta_k \partial \beta_k} \end{bmatrix} \; ; \; \underbrace{\frac{\partial g(\beta)}{\partial \beta}}_{L\times k} = \begin{bmatrix} \frac{\partial g_1}{\partial \beta_1} & \cdots & \frac{\partial g_1}{\partial \beta_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_L}{\partial \beta_1} & \cdots & \frac{\partial g_L}{\partial \beta_k} \end{bmatrix}$$

These are the gradient, Hessien, Jacobian, respectively.

Extremum estimators: a summary

Identified object we wish to estimate (estimand):

$$\beta_0 = \arg\min_{\beta} Q(\beta)$$

Estimator:

$$\hat{\beta} = \arg\min_{\beta} \hat{Q}_n(\beta)$$

Properties:

- Consistency: $\hat{\beta} \xrightarrow{p} \beta_0$ as $n \to \infty$
- Assymptotic normality: $\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, V)$

(Reference: Newey & McFadden (1994) chapter in Handbook of Econometrics)

Extremum estimators: a summary

Extremum Consistency Theorem:

If there is a function $Q(\beta)$ and a vector $\beta_0 \in B$ such that:

1. $Q(\beta)$ is uniquely minimized at β_0
2. B is compact
3. $Q(\beta)$ is continuous
4. $\hat{Q}_n(\beta)$ converges uniformly in probability to $Q(\beta)$:

$$sup_\beta |\hat{Q}_n(\beta) - Q(\beta)| \xrightarrow{p} 0$$

Then, $\hat{\beta} \xrightarrow{p} \beta_0$.

Extremum estimators: a summary

Uniform convergence in probability is a key condition. It says that for each $\epsilon > 0$, the entire function $\hat{Q}_n(\beta)$ is inside the sleeve $|Q(\beta) - \epsilon, Q(\beta) + \epsilon|$ with probability approaching 1.

Lemma:

Let $\hat{Q}_n(\beta) = \frac{1}{n}\sum m(D_i, \beta)$ and $Q(\beta) = E[m(D_i, \beta)]$.
If:

1. D_i are iid
2. B is compact
3. $m(d, \beta)$ is continuous at each β with prob one
4. $\exists\ M(d)$ such that $|m(d, \beta)| \leq M(d)$ for all β and $E[M(D_i)] < \infty$

Then $Q(\beta)$ is continous and $\hat{Q}_n(\beta)$ satisfies uniform convergence.

Extremum estimators: a summary

Asymptotic Normality Theorem:

If $\hat{\beta} \xrightarrow{p} \beta_0$:

1. β_0 is in the interior of B
2. $\hat{Q}_n(\beta)$ is twice constinuously differentiable in a neighborhood N if θ_0
3. $\sqrt{n}\frac{\partial}{\partial\beta}\hat{Q}_n(\beta_0) \xrightarrow{d} N(0,\Omega)$
4. $\frac{\partial^2}{\partial\beta\partial\beta'}\hat{Q}_n(\beta)$ converges uniformly in probability to $\frac{\partial^2}{\partial\beta\partial\beta'}Q(\beta)$, which is continuous at β_0:

$$sup_{\beta\in\mathbb{N}}\left|\frac{\partial^2}{\partial\beta\partial\beta'}\hat{Q}_n(\beta) - \frac{\partial^2}{\partial\beta\partial\beta'}Q(\beta)\right| \xrightarrow{p} 0$$

5. $\frac{\partial^2}{\partial\beta\partial\beta'}Q(\beta_0)$ is invertible

$$\sqrt{n}(\hat{\beta}-\beta_0) \xrightarrow{d} N\left(0, \left[\frac{\partial^2}{\partial\beta\partial\beta'}Q(\beta_0)\right]^{-1}\Omega\left[\frac{\partial^2}{\partial\beta\partial\beta'}Q(\beta_0)\right]^{-1}\right)$$

Extremum estimators: m-estimation

$$\beta_0 = arg\min_{\beta} E[m(D_i, \beta)]$$

$$\text{FOC:} \frac{\partial}{\partial \beta} E[m(D_i, \beta_0)] = E\left[\frac{\partial}{\partial \beta} m(D_i, \beta_0)\right] = 0$$

Estimator:

$$\hat{\beta}_m = arg\min_{\beta} \frac{1}{n} \sum m(D_i, \beta)$$

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N\left(0, \left[\frac{\partial^2}{\partial \beta \partial \beta'} E[m(D_i, \beta_0)]\right]^{-1} \Omega \left[\frac{\partial^2}{\partial \beta \partial \beta'} E[m(D_i, \beta_0)]\right]^{-1}\right)$$

Extremum estimators: m-estimation

Key observations from where to derive this:

- $\frac{\partial}{\partial\beta}\left(\frac{1}{n}\sum m(D_i,\hat{\beta})\right) = \frac{\partial}{\partial\beta}\left(\frac{1}{n}\sum m(D_i,\beta_0)\right) + \frac{\partial^2}{\partial\beta\partial\beta'}\left(\frac{1}{n}\sum m(D_i,\beta^*)\right)(\hat{\beta}-\beta_0) = 0$
- $\sqrt{n}\frac{\partial}{\partial\beta}\left(\frac{1}{n}\sum m(D_i,\beta_0)\right) = \frac{1}{\sqrt{n}}\sum\left(\frac{\partial}{\partial\beta}m(D_i,\beta_0)\right) \xrightarrow{d} N\left(\underbrace{E\left[\frac{\partial}{\partial\beta}m(D_i,\beta_0)\right]}_{=0},\Omega\right)$

Where, with iid data:

$$\Omega = Var\left(\frac{\partial}{\partial\beta}m(D_i,\beta_0)\right) =$$

$$E\left[\left(\frac{\partial}{\partial\beta}m(D_i,\beta_0) - \underbrace{E\left[\frac{\partial}{\partial\beta}m(D_i,\beta_0)\right]}_{=0}\right)\left(\frac{\partial}{\partial\beta}m(D_i,\beta_0) - \underbrace{E\left[\frac{\partial}{\partial\beta}m(D_i,\beta_0)\right]}_{=0}\right)'\right]$$

Example: NLS

$D_i = (Y_i, X_i)$ has CEF: $E[Y_i|X_i]$

Remember Problem Set 4, Exercise 2:

$$E[(Y_i - h(X_i,\theta))^2] = E[(Y_i - E[Y_i|X_i])^2] + E[(E[Y_i|X_i] - h(X_i,\theta))^2]$$

Our model (family of conditional expectations): $h(x,\theta)$

If our model is well specified, there exists a θ_0 such that $E[Y_i|X_i] = h(X_i,\theta_0)$, so:

$$\theta_0 = arg\min_{\theta} E[(Y_i - h(X_i,\theta))^2]$$

θ_0 is the unique minimum (and thus θ_0 is identified) if $P(h(X_i,\theta) \neq h(X_i,\theta_0)) > 0$ ($\Leftrightarrow P(h(X_i,\theta) = h(X_i,\theta_0)) < 1$) $\forall \theta \neq \theta_0$.
(And, on the contrary, there is no unique min if there are two distinct values θ_0 and θ_0' for which $P(h(X_i,\theta_0') = h(X_i,\theta_0)) = 1$)

$$\hat{\theta}_{NLS} = arg\min_{\theta} \frac{1}{n} \sum (Y_i - h(X_i,\theta))^2$$

Example: NLS

If our model is miss-specified, there is no θ_0 such that $E[Y_i|X_i] = h(X_i, \theta_0)$ (we're looking at the wrong parametric family of CEFs), so:

$$\theta_0^* = arg\min_{\theta} E[(Y_i - h(X_i, \theta))^2] = arg\min_{\theta} E[(E[Y_i|X_i] - h(X_i, \theta))^2]$$

Then $\hat{\theta}_{NLS}$ is estimating the θ_0^* for which $h(X_i, \theta_0^*)$, which is within our family of CEFs, is the best approximation to the true conditional expectation $E[Y_i|X_i]$ ("best" in the sense that it minimizes the mean squared error).

Example: NLS

Exercise: Problem Set 4, Exercise 2 walked you through an example of an extremum estimator and asked you to prove consistency.

- Model:

$$Y_i = T_i'\beta_0 + \epsilon_i \; ; \; T_i \perp \epsilon_i \; ; \; \epsilon_i \sim N(0, \sigma_0^2)$$

$$D_i = (T_i, Y_i^*) \; ; \; Y_i^* = min\{Y_i, c\}$$

- Find what the CEF is:

$$E[Y_i^*|T_i] = \Phi\left(\frac{c - T_i'\beta_0}{\sigma_0}\right)\left(T_i'\beta_0 - \frac{\phi\left(\frac{c-T_i'\beta_0}{\sigma_0}\right)}{\Phi\left(\frac{c-T_i'\beta_0}{\sigma_0}\right)}\sigma_0\right) + \left[1 - \Phi\left(\frac{c - T_i'\beta_0}{\sigma_0}\right)\right]c$$

  Notice that it is a function of β_0: $E[Y_i^*|T_i] \equiv h(T_i, \beta_0)$.

- Determine that OLS of Y_i^* on T_i converges to the BLP of Y_i^* given T_i, but this doesn't identify β_0:

$$\hat{\beta}_{OLS} \xrightarrow{p} E[T_i T_i']^{-1} E[T_i Y_i^*] \equiv \tilde{\beta}$$

$$\tilde{\beta} \neq \beta_0 = E[T_i T_i']^{-1} E[T_i Y_i]$$

Example: NLS

- Recall the properties of a CEF:

$$E[f(T_i)(Y_i^* - E[Y_i^*|T_i])] = 0 \Rightarrow E[f(T_i)(Y_i^* - h(T_i, \beta_0))] = 0$$

 For any function $f(.)$. This result helps establish that β_0 minimizes the objective function set up in the following step.

- Establish an objective function that is uniquely minimized at β_0 (and prove this):

$$\beta_0 = \underset{\beta}{arg\,min}\ E[(Y_i^* - h(T_i, \beta))^2]$$

 The associated NLS estimator is:

$$\hat{\beta}_{NLS} = \underset{\beta}{arg\,min}\ \frac{1}{n} \sum (Y_i^* - h(T_i, \beta))^2$$

- Show that $\hat{\beta}_{NLS} \xrightarrow{p} \beta_0$. Based on the Extremum Consistency Theorem, need to check that conditions 1-4 are satisfied. The tricky one is to show 4. Since our case corresponds to the type $\hat{Q}_n(\beta) = \frac{1}{n} \sum m(D_i, \beta)$ and $Q(\beta) = E[m(D_i, \beta)]$, where $m(D_i, \beta) \equiv m(T_i, Y_i^*, \beta) \equiv Y_i^* - h(T_i, \beta)$, we can use Lemma.

Extremum estimators: GMM

We know/assume that β_0 is identified by moment conditions: $E[g(D_i, \beta_0)] = 0$.

$$\beta_0 = \arg\min_{\beta} E[g(D_i, \beta)]' \; W \; E[g(D_i, \beta)]$$

FOC:

$$\underbrace{E\left[\frac{\partial g(D_i, \beta)}{\partial \beta'}\right]'}_{k \times L} \underbrace{W}_{L \times L} \underbrace{E[g(D_i, \beta)]}_{L \times 1} = 0$$

Estimator:

$$\hat{\beta}_{GMM} = \arg\min_{\beta} \left[\frac{1}{n}\sum g(D_i, \beta)\right]' \hat{W} \left[\frac{1}{n}\sum g(D_i, \beta)\right]$$

FOC:

$$\underbrace{\left[\frac{1}{n}\sum \frac{\partial g(D_i, \hat{\beta}_{GMM})}{\partial \beta'}\right]'}_{k \times L} \underbrace{\hat{W}}_{L \times L} \underbrace{\left[\frac{1}{n}\sum g(D_i, \hat{\beta}_{GMM})\right]}_{L \times 1} = 0$$

Note 1: linear GMM estimators are available in closed form, eg, OLS and 2SLS.

Note 2: if system is just-identified, choice of $\hat{W}$ is inconsequential.

Extremum estimators: GMM

$$\sqrt{n}\left(\hat{\beta}_{GMM} - \beta_0\right) \xrightarrow{p} N\left(0, (G'WG)^{-1}G'W\Omega WG(G'WG)^{-1}\right)$$

$$G \equiv E\left[\frac{\partial g(D_i, \beta_0)}{\partial \beta'}\right]'$$

$$\Omega \equiv E[g(D_i, \beta_0)g(D_i, \beta_0)']$$

Efficient GMM: considers $W = \Omega^{-1}$

$$\sqrt{n}\left(\hat{\beta}_{GMM} - \beta_0\right) \xrightarrow{p} N\left(0, (G'\Omega^{-1}G)^{-1}\right)$$

Extremum estimators: GMM

Notice that efficient $\hat{\beta}_{GMM}$ requires one to have a consistent estimate $\hat{W}$ of $W = \Omega^{-1}$, where $\Omega \equiv E[g(D_i, \beta_0)g(D_i, \beta_0)']$. But note that Ω depends on β_0! So we need a consistent estimator of β_0 to get a consistent esttimator for W, to get the efficient GMM esimtator...

Efficient GMM estimator derived in two steps:

1. Obtain an initial consistent estimator $\hat{\beta}^1_{GMM}$ of β_0 using some known positive definite $\hat{W}$ (eg, I). Use $\hat{\beta}^1_{GMM}$ to obtain a consistent estimate of Ω:

$$\hat{\Omega} = \frac{1}{n} \sum g(D_i, \hat{\beta}^1_{GMM}) g(D_i, \hat{\beta}^1_{GMM})'$$

2. Obtain the efficient GMM estimator $\hat{\beta}^2_{GMM}$ using $\hat{\Omega}$ from previous step.

Extremum estimators: minimum distance

$$\beta_{MD} = arg\,\min_{\beta} r(\delta_0, \beta)' \; W \; r(\delta_0, \beta)'$$

Say that we have a consistent estimator for δ_0 and W: $\hat{\delta}$ and $\hat{W}$.

$$\hat{\beta}_{MD} = arg\,\min_{\beta} r(\hat{\delta}, \beta)' \; \hat{W} \; r(\hat{\delta}, \beta)'$$

An example is moment-matching, in which $\delta_0 = E[Z]$, where functional form of expectation is given by a model $E[Z] = h(\beta_0)$.
So define $r(\delta_0, \beta) = \delta_0 - h(\beta_0)$ and use a consistent estimator of δ_0, which leads to:

$$\hat{\beta}_{MD} = arg\,\min_{\beta} (\hat{\delta} - h(\beta)) \; \hat{W} \; (\hat{\delta} - h(\beta))'$$

Example: 2SLS

Exercise: Problem Set 3, Exercise 2 asks you to show that the IV estimator is an example of a GMM estimator and a minimum distance estimator.

- Z_i is $L \times 1$ and T_i is $k \times 1$
- 2SLS (IV) estimand from where we derive the 2SLS (IV) estimator is:

$$\beta_{2SLS} = \left(E[T_i Z_i']E[Z_i Z_i']^{-1}E[Z_i T_i']\right)^{-1} E[T_i Z_i']E[Z_i Z_i']^{-1}E[Z_i Y_i]$$

Example: 2SLS

- Can be seen as a special case of GMM:
  Moment condition: $E[Z_i u_i] = 0$ where $Y_i = T_i'\beta_0 + u_i$

$$\beta_{2SLS} = \arg\min_{\beta} \underbrace{\overbrace{E[Z_i(Y_i - T_i'\beta)]'}^{1\times L}}_{E[g(D_i,\beta)]'} \underbrace{\overbrace{E[Z_i Z_i']^{-1}}^{L\times L}}_{W} \underbrace{\overbrace{E[Z_i(Y_i - T_i'\beta)]}^{L\times 1}}_{E[g(D_i,\beta)]}$$

FOC:

$$\overbrace{\underbrace{E[T_i Z_i']}_{E\left[\frac{\partial g(D_i,\beta)}{\partial \beta}\right]'} \; \underbrace{E[Z_i Z_i']}_{W}^{-1}}^{k\times L} \underbrace{\overbrace{E[Z_i(Y_i - T_i'\beta_{2SLS})]}^{L\times 1}}_{E[g(D_i,\beta)]} = 0$$

From this FOC, it is clear that we get the formula for β_{2SLS} from previous slide.

Remember GMM objective function:

$$\beta_{GMM} = \arg\min_{\beta} E[g(D_i,\beta)]' \; W \; E[g(D_i,\beta)]$$

FOC:

$$E\left[\frac{\partial g(D_i,\beta_{GMM})}{\partial \beta}\right]' W E[g(D_i,\beta_{GMM})] = 0$$

Example: 2SLS

So 2SLS is the GMM estimator that considers weight matrix $W = E[Z_i Z_i']^{-1}$.

$$\sqrt{n}\left(\hat{\beta}_{2SLS} - \beta_0\right) \xrightarrow{p} N\left(0, (G'WG)^{-1} G'W\Omega WG(G'WG)^{-1}\right)$$

$$G \equiv E\left[T_i Z_i'\right]$$

$$\Omega \equiv E[Z_i u_i u_i' Z_i'] = E[u_i^2 Z_i Z_i']$$

$$W = E[Z_i Z_i']^{-1}$$

We can further distinguish two cases: if the error is homoskedastic (in which case 2SLS is the GMM estimator with smaller variance) and if isn't.

Example: 2SLS

1) Error is homoskedastic: $E[u_i^2|Z_i] = \sigma^2$

$$\Omega = E[E[u_i^2 Z_i Z_i' | Z_i]] = \sigma^2 E[Z_i Z_i']$$

Efficient GMM takes $W = \sigma^2 \Omega^{-1} = E[Z_i Z_i']^{-1}$
(or $W = \Omega^{-1} = (\sigma^2 E[Z_i Z_i'])^{-1}$ since multiplying by a constant is inoquous).

This is exactly what 2SLS does! So 2SLS is the efficient GMM estimator under homoskedasticity and the variance simplifies:

$$\sqrt{n}\left(\hat{\beta}_{2SLS} - \beta_0\right) \xrightarrow{p} N\left(0, (G'\Omega^{-1}G)^{-1}\right)$$

$$G = E[T_i Z_i']$$

$$\Omega = \sigma^2 E[Z_i Z_i']$$

Example: 2SLS

1) Error is **not** homoskedastic:

$$\Omega = E[u_i^2 Z_i Z_i']$$

Efficient GMM takes $W = \Omega^{-1}$ where now Ω is NOT $\sigma^2 E[Z_i Z_i']$.

So 2SLS in NOT the efficient GMM estimator under heteroskedasticity.
The big chunky formula for the variance can't be simplified:

$$\sqrt{n}\left(\hat{\beta}_{2SLS} - \beta_0\right) \xrightarrow{p} N\left(0, (G'WG)^{-1} G'W\Omega WG(G'WG)^{-1}\right)$$

$$G \equiv E\left[T_i Z_i'\right]$$

$$\Omega \equiv E[u_i^2 Z_i Z_i']$$

$$W = E[Z_i Z_i']^{-1}$$

Example: 2SLS

Even under heteroskedasticity, many prefer to use 2SLS, which is NOT the efficient GMM. If we wanted an efficient GMM estimator, remember the two steps:

1. Obtain a consistent estimator $\hat{\beta}$ of β_0 using some known positive definite $\hat{W}$. For example, could choose $\hat{W} = \left(\frac{1}{n}\sum Z_i Z_i'\right)^{-1}$, which gives the 2SLS estimator.
   Obtain the residuals $\hat{u}_i = Y_i - X_i'\hat{\beta}$. A consistent estimate of Ω is:
   $$\hat{\Omega} = \left(\frac{1}{n}\sum \hat{u}_i^2 Z_i Z_i'\right)^{-1}$$
2. Obtain the efficient GMM estimator using $\hat{W} = \hat{\Omega}^{-1}$.

Example: 2SLS

- Can be seen as a special case of minimum distance:

$$\beta_{2SLS} = \arg\min_{\beta} \underbrace{(\delta_{RF} - \delta_{FS}\beta)'}_{r(\delta_0,\beta)'} \underbrace{E[Z_i Z_i']}_{W} \underbrace{(\delta_{RF} - \delta_{FS}\beta)}_{r(\delta_0,\beta)}$$

Where:

$$\delta_{RF} = E[Z_i Z_i']^{-1} E[Z_i Y_i]$$
$$\delta_{FS} = E[Z_i Z_i']^{-1} E[Z_i T_i']$$

Note: $\delta_{RF} - \delta_{FS}\beta$ is like "effect of Z on Y" - "effect of Z on T" $\times$ "effect of T on Y". β is the minimizer of this triangulation.

Remember minimum distance objective function:

$$\beta_{MD} = \arg\min_{\beta} r(\delta_0, \beta)' \; W \; r(\delta_0, \beta)'$$

Example: 2SLS

Not super obvious that solving this min problem leads to the β_{2SLS} formula:

$$\begin{aligned}\delta_{RF} - \delta_{FS}\beta &= E[Z_iZ_i']^{-1}E[Z_iY_i] - E[Z_iZ_i']^{-1}E[Z_iT_i']\beta \\ &= E[Z_iZ_i']^{-1}(E[Z_iY_i] - E[Z_iT_i']\beta) \\ &= E[Z_iZ_i']^{-1}E[Z_i(Y_i - T_i'\beta)]\end{aligned}$$

So:

$$\beta_{2SLS} = arg\min_{\beta} \underbrace{(\delta_{RF} - \delta_{FS}\beta)'}_{r(\delta_0,\beta)'} \underbrace{E[Z_iZ_i']}_{W} \underbrace{(\delta_{RF} - \delta_{FS}\beta)}_{r(\delta_0,\beta)}$$

$$\beta_{2SLS} = arg\min_{\beta} E[Z_i(Y_i - T_i'\beta)]'E[Z_iZ_i']^{-1}E[Z_iZ_i']E[Z_iZ_i']^{-1}E[Z_i(Y_i - T_i'\beta)]$$

$$\beta_{2SLS} = arg\min_{\beta} E[Z_i(Y_i - T_i'\beta)]'E[Z_iZ_i']^{-1}E[Z_i(Y_i - T_i'\beta)]$$

m-estimation v GMM

Exercise: Problem Set 3, Exercise 2 also asks you to think in which cases can an m-estimator be seen as a GMM estimator.

Remember m-estimator objective function:

$$\beta_m = arg\min_{\beta} E[m(D_i, \beta)]$$

FOC:

$$\frac{\partial E[m(D_i, \beta)]}{\partial \beta} = 0$$

Assuming we can interchange derivation and integration (expectation):

$$E\left[\frac{\partial m(D_i, \beta)}{\partial \beta}\right] = 0$$

From this moment conditions, we can define a GMM objective function and derive a GMM estimator.

But! These are moment conditions that hold at β_m (the global min of the minimization problem) but potentially also at local mins or saddle points. If there are local min or saddle points, then the FOC will be picking those up and thus the β_{GMM} can deliver different answers than m-estimation.

A disgression on: quantile regressions

τ-quantile of the distribution of Y_i:

$$y_\tau \equiv inf\{y : P(Y_i \leq y) \geq \tau\} = F^{-1}(\tau)$$

Moment condition that quantile y_τ satisfies:

$$\begin{aligned} \tau &= F(y_\tau) \\ \tau - F(y_\tau) &= 0 \\ \tau - P(Y_i \leq y_\tau) &= 0 \\ \tau - E[1\{Y_i - y_\tau \leq 0\}] &= 0 \\ E[\tau - 1\{Y_i - y_\tau \leq 0\}] &= 0 \end{aligned}$$

A disgression on: quantile regressions

The following function (written in three different ways)...

$$\rho_\tau(u) = (\tau - 1\{u \leq 0\})u$$
$$= \begin{cases} (1-\tau)|u| \ \ if \ \ u \leq 0 \\ \tau|u| \ \ if \ \ u > 0 \end{cases}$$
$$= \tau max\{u, 0\} + (1-\tau)min\{u, 0\}$$

...has derivative:

$$\frac{\partial \rho_\tau(u)}{\partial u} = \tau - 1\{u \leq 0\}$$

So we can write the moment condition as:

$$E\left[\frac{\partial \rho_\tau(Y_i - y_\tau)}{\partial y_\tau}\right] = -E[\tau - 1\{Y_i - y_\tau \leq 0\}] = 0$$

A disgression on: quantile regressions

Based on this moment condition, M-estimation problem:

$$y_\tau = arg \min_{y_\tau} \; E[\rho_\tau(Y_i - y_\tau)]$$

$$FOC: \quad \frac{\partial}{\partial y_\tau} E[\rho_\tau(Y_i - y_\tau)] = E\left[\frac{\partial \rho_\tau(Y_i - y_\tau)}{\partial y_\tau}\right] = 0$$

That is:

$$y_\tau = arg \min_{y_\tau} \; E[(\tau - 1\{(Y_i - y_\tau) \leq 0\})(Y_i - y_\tau)]$$

$$FOC: \quad E[\tau - 1\{Y_i - y_\tau \leq 0\}] = 0$$

Note: remember how we concluded that under loss $L(\delta(X), \theta) = (\tau - 1\{\theta - \delta(X) \leq 0\})(\theta - \delta(X))$, the solution to $\min_{\delta(X)} \; E_{f(\theta|X)}[L(\delta(X), \theta)]$ is the posterior τ quantile of θ, ie, $\delta(X) = \theta_\tau$?

Conditional quantile regressions and quantile IV

Suppose that, conditional on X_i, the τ quantile of Y_i is linear:

$$y_\tau = X_i'\beta_\tau$$

$$\begin{aligned}\beta_\tau &= arg\min_{\beta_\tau} \; E[\rho_\tau(Y_i - X_i'\beta_\tau)] \\ &= arg\min_{\beta_\tau} \; E[(\tau - 1(Y_i - X_i'\beta_\tau \le 0))(Y_i - X_i'\beta_\tau)]\end{aligned}$$

FOC:

$$E[\tau - 1(Y_i - X_i'\beta_\tau \le 0)X_i] = 0$$

Conditional quantitle rgeressions and quantile IV

Conditional quantile regression estimator (Koenker and Bassett 1978):

$$\hat{\beta}_\tau = arg\,min_b \frac{1}{N} \sum_i (\tau - 1(Y_i - X_i' b \leq 0))(Y_i - X_i' b)$$

Note: need numerical method to solve that is not gradient-based.

Extend the framework of linear IV to quantile IV (Chernozhukov and Hansen 2004, 2005):

- Linear IV: goes from moment condition $E[(Y_i - X_i'\beta)X_i] = 0$ to $E[(Y_i - X_i'\beta)Z_i] = 0$
- Quantile IV: goes from moment condition $E[\tau - 1(Y_i - X_i'\beta_\tau \leq 0)X_i] = 0$ to $E[\tau - 1(Y_i - X_i'\beta_\tau \leq 0)Z_i] = 0$