Metrics 101

Pavel Pronin

HSE $\rightarrow$ NES

July 5, 2022

Outline

- This is essentially a *snapshot*. I omit many interesting topics: time series models, censored models, selection models, panel data and etc.
- However, if you decide *to dive in* it (hopefully) would be easier to grasp new models in all the trickery
- Recommended readings:
    1. (Black) Wooldridge
    2. Hansen's Econometrics (for more in depth treatment)

Econometric Concepts

Expectation, Variance, Covariance

- Let $\xi \in \Omega$ be some random variable. Then *mathematical expectation* of ξ is

$$\mathbb{E}[\xi] = \int_\Omega \xi dP(\xi) = \begin{cases} \int_{-\infty}^{+\infty} \xi f(\xi) d\xi & \text{for continuous RV} \\ \sum_{i=1}^{|\Omega|} \xi_i P(\xi_i) & \text{for discrete RV} \end{cases}$$

- Properties of $\mathbb{E}[\xi]$
  1. $\mathbb{E}[\xi + \eta] = \mathbb{E}[\xi] + \mathbb{E}[\eta]$, where η is another RV.
  2. $\mathbb{E}[a\xi + b\eta] = a\mathbb{E}[\xi] + b\mathbb{E}[\eta]$
  3. $\eta = g(\xi) \implies \mathbb{E}[\eta] = \mathbb{E}[g(\xi)]$, where $g(\cdot)$ is some *measurable* function.
- Example: suppose $\xi \sim U[a, b]$

$$f(\xi) = \begin{cases} \frac{1}{b-a} & \xi \in [a, b] \\ 0 & \xi \notin [a, b] \end{cases} \implies \mathbb{E}[\xi] = \frac{1}{b-a} \int_a^b \xi d\xi = \frac{1}{2} \frac{b^2 - a^2}{b - a} = \frac{b + a}{2}$$

$$\mathbb{E}[\xi^2] = \frac{1}{b-a} \int_a^b \xi^2 d\xi = \frac{1}{3} \frac{(b^3 - a^3)}{b - a} = \frac{b^2 + ab + a^2}{3}$$

Expectation, Variance, Covariance

- Variance is defined as

$$\mathbb{V}[\xi] = \mathbb{E}[(\xi - \mathbb{E}[\xi])(\xi - \mathbb{E}[\xi]) = \mathbb{E}[\xi^2] - \mathbb{E}[\xi]^2$$
$$\mathbb{C}[\xi, \eta] = \mathbb{E}[(\xi - \mathbb{E}[\xi])(\eta - \mathbb{E}[\eta])] = \mathbb{E}[\xi\eta] - \mathbb{E}[\xi]\mathbb{E}[\eta]$$

- Properties are
  1. $\mathbb{V}[a\xi + b] = a^2\mathbb{V}[\xi]$
  2.

$$\begin{aligned}
\mathbb{V}[\xi + \eta] &= \mathbb{E}\left\{[((\xi + \eta) - \mathbb{E}[\xi + \eta])^2\right\} = \mathbb{E}[(\xi + \eta)^2] - \mathbb{E}[\xi + \eta]^2 \\
&= \mathbb{E}[\xi^2] + \mathbb{E}[\eta^2] + 2\mathbb{E}[\eta\xi] - \mathbb{E}[\xi + \eta]^2 \\
&= \mathbb{E}[\xi^2] + \mathbb{E}[\eta^2] + 2\mathbb{E}[\eta\xi] - \mathbb{E}[\xi]^2 - \mathbb{E}[\eta]^2 - 2\mathbb{E}[\xi]\mathbb{E}[\eta] \\
&= \underbrace{\mathbb{E}[\xi^2] - \mathbb{E}[\xi]^2}_{\mathbb{V}[\xi]} + \underbrace{\mathbb{E}[\eta^2] - \mathbb{E}[\eta]^2}_{\mathbb{V}[\eta]} + 2\underbrace{\left(\mathbb{E}[\xi\eta] - \mathbb{E}[\xi]\mathbb{E}[\eta]\right)}_{\mathbb{C}[\xi,\eta]}
\end{aligned}$$

Variance

- Back to our example, $\xi \sim U[a, b]$

$$f(\xi) = \begin{cases} \frac{1}{b-a} & \xi \in [a, b] \\ 0 & \xi \notin [a, b] \end{cases}$$

$$\mathbb{V}[\xi] = \mathbb{E}[\xi^2] - \mathbb{E}[\xi]^2$$

$$\mathbb{E}[\xi] = \frac{b+a}{2}, \mathbb{E}[\xi^2] = \frac{b^2 + ab + a^2}{3}$$

$$\implies \mathbb{V}[\xi] = \frac{b^2 + ab + a^2}{3} - \frac{b^2 + a^2 + 2ab}{4} = \frac{b^2 + a^2 - 2ab}{12} = \frac{(a-b)^2}{12}$$

- For

Conditional Expectation

- Consider two RVs ξ, η with joint distribution $f_{\xi,\eta}(\xi,\eta)$.
- Recall Bayes Theorem

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \implies f_{\xi|\eta}(\xi|\eta) = \frac{f_{\xi,\eta}(\xi,\eta)}{f_\eta(\eta)}$$

- Recall Law of Total Probability

$$P(B) = P(B|A)P(A) \implies f_\eta(\eta) = \int_{\mathbb{R}} f_{\xi,\eta}(\xi,\eta)d\xi$$

Conditional Expectation example

- Let (x, y) be a random pair of variables (e.g. height and weight)
- Suppose they have joint density function $f_{x,y}(x, y)$, e.g.

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}\right)$$

- Conditional density can be found by

$$f_{x_1|x_2}(x_1|x_2) = \frac{f_{x_1,x_2}(x_1, x_2)}{f_{x_2}(x_2)} = \frac{f_{x_1,x_2}(x_1, x_2)}{\int_{-\infty}^{+\infty} f_{x_1,x_2}(x_1, x_2) dx_2}$$

$$f_{x_1}(x_1) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{1}{2}\frac{(x_1 - \mu_1)^2}{\sigma_1^2}\right)$$

$$f_{x_1,x_2}(x) = \frac{1}{\sqrt{det(2\pi\Sigma)}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)'\Sigma^{-1}(\mathbf{x} - \mu)\right)$$

Conditional Expectation

- It can be shown (just integrate and factor out) that:

$$x_1 | x_2 \sim \mathcal{N}\left(\mu, \sigma^2\right)$$
$$\mu = \mu_1 + \frac{\sigma_{12}}{\sigma_2^2}(x_2 - \mu_2)$$
$$\sigma^2 = \sigma_1^2 - \left(\frac{\sigma_{12}}{\sigma_2}\right)^2$$

- Therefore,

$$\mathbb{E}[x_1 | x_2] = \mu_1 + \frac{\sigma_{12}}{\sigma_2^2}(x_2 - \mu_2)$$

- Hmmm.... Actually, *this is regression*.

Conditional Expectation

- **LIE** – Law of Iterated Expectations

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]]$$

For our example, recall that $\mathbb{E}[x_1] = \mu_1$ and apply the formula

$$\mathbb{E}[x_1] = \mathbb{E}[\mathbb{E}[x_1|x_2]] = \mathbb{E}\left[\mu_1 + \frac{\sigma_{12}}{\sigma_2^2}(x_2 - \mu_2)\right] = \mu_1$$

Conditional Variance

- Conditional Variance

$$\mathbb{V}[x_1|x_2] = \mathbb{E}[(x_1 - \mathbb{E}[x_1|x_2])^2|x_2] = \mathbb{E}[x_1^2|x_2] - \mathbb{E}[x_1|x_2]^2 = \sigma^2 = \sigma_1^2 - \left(\frac{\sigma_{12}}{\sigma_2}\right)^2$$

- **Law of total Variance**

$$\mathbb{V}[X] = \mathbb{V}[\mathbb{E}[X|Y]] + \mathbb{E}[V[X|Y]]$$

$$\mathbb{V}[x_1] = \mathbb{V}[\mu_1 + \frac{\sigma_{12}}{\sigma_2^2}(x_2 - \mu_2)] + \sigma_1^2 - \left(\frac{\sigma_{12}}{\sigma_2}\right)^2 = \frac{\sigma_{12}^2}{\sigma_2^4}\sigma_2^2 + \sigma_1 - \frac{\sigma_{12}^2}{\sigma_2^2} = \sigma_1$$

Now we are ready

Prediction Problem

- MSPE – **M**ean **S**quared **P**rediction **E**rror.

$$\arg\min_{g(x)} \mathbb{E}\left[(y - g(x))^2\right]$$

- This *convex* optimization problem.
  - A function $f : \mathbb{R} \to \mathbb{R}$ is said to be *convex* if

$$\begin{aligned} &(1)\ \forall \alpha \in (0,1) : f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y) \\ \text{or }& (2)^*\ \ f'(x)(y-x) \leq f(y) - f(x) \\ \text{or }& (3)^*\ \ f''(x) \geq 0 \forall x \end{aligned}$$

  In our case we have

$$(y - g(x))^2 \equiv f(z) = z^2, f''(z) = 2 \geq 0 \forall z \implies \text{ convex optimization}$$

- $\implies$ global minimum exists.

Prediction Problem

$$\mathbb{E}\left[(y-g(x))^2\right] = \mathbb{E}\left[(y-\mathbb{E}[y|x]+\mathbb{E}[y|x]-g(x))^2\right]$$
$$= \mathbb{E}\left[(y-\mathbb{E}[y|x])^2 + (\mathbb{E}[y|x]-g(x))^2 + \underbrace{2(y-\mathbb{E}[y|x])(\mathbb{E}[y|x]-g(x))}_{2A}\right]$$

Consider A. And apply *Jensen's Inequality:* for RV x and *convex* function ϕ (recall that $f(z)=z^2$ is convex)

$$\phi(\mathbb{E}[x]) \leq \mathbb{E}[\phi(x)]$$

Then

$$A = y\mathbb{E}[y|x] - \mathbb{E}[y|x]^2 - yg(x) + g(x)\mathbb{E}[y|x]$$
$$\mathbb{E}[A|x] = \mathbb{E}[y|x]^2 - \mathbb{E}[y|x]^2 - \mathbb{E}[y|x]g(x) + g(x)\mathbb{E}[y|x] = 0 \implies \mathbb{E}[A] = \mathbb{E}[\mathbb{E}[A|x]] = 0$$

Prediction Problem

Therefore, we shown that

$$\mathbb{E}[(y-g(x))^2] = \mathbb{E}\left[(y-\mathbb{E}[y|x])^2\right] + \underbrace{\mathbb{E}\left[(\mathbb{E}[y|x]-g(x))^2\right]}_{\geq 0}$$

Thus we proofed, that if we want to minimize MSPE, *optimal predictor* of y by x is conditional expectation $\mathbb{E}[y|x]$.
And *optimal prediction error* is just

$$e = y - \mathbb{E}[y|x]$$

we simply *can not do better*. Very important property: Apply LIE

$$\mathbb{E}[e|x] = \mathbb{E}[y|x] - \mathbb{E}[y|x] = 0$$
$$\mathbb{E}[e] = \mathbb{E}[\mathbb{E}[e|x]] = 0$$

Moreover, consider any function $h(x)$ and apply LIE

$$\mathbb{E}[eh(x)|x] = h(x)\mathbb{E}[e|x] = 0 \implies \mathbb{E}[eh(x)] = 0$$

Analogy Principle: Where do estimators come from?

- So far we were dealing with theoretical distributions, expectations, etc.
- Suppose parameter θ is some function of distribution $F(z)$. We only have a limited sample and want to estimate this parameter.
- *Analogy Principle*: given random sample $\{z_i\}_{i=1}^n$ construct analog estimator $\hat{\theta}$ as same function of *empirical* distribution $F_n(z)$, where

$$F_n(z) = \frac{\#\text{elements in the sample} \leq z}{n}$$

- Example with mean. Suppose $\theta = \mathbb{E}[z]$
  1. True value

$$\theta = \mathbb{E}[z] = \int_{-\infty}^{+\infty} z dF(z)$$

  2. Analogy Principle

$$\hat{\theta} = \int_{-\infty}^{+\infty} z dF_n(z) = \frac{1}{n} \sum_{i=1}^{n} z_i$$

Analogy Principle: Where do estimators come from?

- Another example: Suppose that

$$\theta = \frac{cov(x, y)}{var(x)} = \frac{\mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y]}{\mathbb{E}[x^2] - \mathbb{E}[x]^2}$$

- True value is

$$\theta = \frac{\int_{-\infty}^{+\infty} xydF(x, y) - \left(\int_{-\infty}^{+\infty} xdF(x)\right)\left(\int_{-\infty}^{+\infty} ydF(y)\right)}{\int_{-\infty}^{+\infty} x^2 dF(x) - \left(\int_{-\infty}^{+\infty} xdF(x)\right)^2}$$

- Sample analog is just

$$\hat{\theta} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)}{\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2}$$

Analogy Principle: Why exactly this works and how exactly we switched from Integral to Sum?

- Recall that Riemann Integral is just

$$\int_a^b f(x)dx = \lim_{n\to\infty} \sum_{i=1}^{n} \Delta x \cdot f(x_i)$$

$$\Delta x = \frac{b-a}{n}, x_i = a + \Delta x \cdot i$$

- Riemann-Stieltjes Integral is (where $f : \mathbb{R} \to \mathbb{R}, g : \mathbb{R} \to \mathbb{R}$)

$$\int_a^b f(x)dg(x) = \lim_{n\to\infty} \sum_{i=0}^{n-1} f(c_i)\left[g(x_{i+1}) - g(x_i)\right], c_i \in [x_i; x_{i+1}]$$

- Just apply the definition to Analogy principle

$$\int_{-\infty}^{+\infty} zdF_n(z) = \lim_{n\to\infty} \sum_{i=0}^{n-1} z_i\left[F_n(z_{i+1}) - F_n(z_i)\right] = \lim_{n\to\infty} \sum_{i=1}^{n} z_i \frac{1}{n}$$

- This proves by definition that $\hat{\theta}$ defined by Analogy Principle $\to \theta$ as $n \to \infty$

Back to our Optimization Problem

Suppose that

$$
\begin{aligned}
\theta &= \arg\min_q \mathbb{E}[h(z, q)] \\
\implies \hat{\theta} &= \arg\min_q \frac{1}{n} \sum_{i=1}^{n} h(z_i, q)
\end{aligned}
$$

Very Convenient!

Linear Mean Regression

Linear Mean Regression

Suppose that

$$\mathbb{E}[y|x] = x'\beta,$$

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix}, x' = \begin{bmatrix} x_1 & \dots & x_k \end{bmatrix}, \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, x'\beta = \beta_1 x_1 + \ldots + \beta_k x_k = \sum_{i=1}^{k} x_i \beta_i$$

Equivalently

$$\iff y = x'\beta + e, \mathbb{E}[e|x] = 0$$

If $\mathbb{E}[e|x]$ – this is not even regression!!!!

Why Linear?

1. (Very implausible) Statistical Properties. Recall that if x_1, x_2 are joint normal than

$$\mathbb{E}[x_1|x_2] = \mu_1 + \frac{\sigma_{12}}{\sigma_2^2}(x_2 - \mu_2)$$

2. (Must be) It comes from **theory.** For example, if the production is Cobb-Douglas and error is multiplicative we have:

$$Y = AK^{\alpha}L^{1-\alpha}\exp(u), \mathbb{E}[u|K, L] = 0$$
$$\mathbb{E}[\ln Y|K, L] = \underbrace{\ln A}_{\beta_0} + \underbrace{\alpha}_{\beta_1} \ln K + \underbrace{(1-\alpha)}_{\beta_2} \ln L + u$$

Simple regression by hand

If

$$y = \alpha + \beta x + e, \mathbb{E}[e|x] = 0 \implies \mathbb{E}[y|x] = \alpha + \beta x$$
$$\beta = \arg\min_{a,b} \mathbb{E}[(y - \mathbb{E}[y|x])^2] = \arg\min_{a,b} \mathbb{E}[(y - a - bx)^2]$$

This is unconstrained optimization problem $\implies$ apply Lagrange Theorem and take FOCs.

Caveat: Interchanging expectation and differentiation.

Suppose

$$\mathbb{E}[f(x,\xi)] \to \max_x \iff \int_{-\infty}^{+\infty} f(x,\xi)p(\xi)d\xi \to \max_x$$

Leibniz Integral Rule

$$\frac{d}{dx}\left(\int_{a(x)}^{b(x)} f(x,t)dt\right) = f(x,b(x)) \cdot \frac{d}{dx}b(x) - f(x,a(x))\frac{d}{dx}a(x) + \int_{a(x)}^{b(x)} \frac{\partial}{\partial x} f(x,t)dt$$

As Limits of integration are independent of x in our case we have that

$$\frac{d}{dx}\left(\int_{-\infty}^{+\infty} f(x,\xi)p(\xi)d\xi\right) = \int_{-\infty}^{+\infty} \frac{\partial}{\partial x} f(x,\xi)p(\xi)d\xi$$

$$\implies \frac{\partial}{\partial x}\mathbb{E}\left[f(x,\xi)\right] = \mathbb{E}\left[\frac{\partial}{\partial x} f(x,\xi)\right]$$

Back to our problem. Step 1: Solve the problem

$$\beta = \arg\min_{a,b} \mathbb{E}[(y - a - bx)^2]$$

$$V = \min_{a,b} \mathbb{E}[(y - a - bx)^2]$$

$$0 = \frac{\partial V}{\partial a} \implies -2\mathbb{E}[(y - a - bx)] = 0 \implies a = \mathbb{E}[y] - b\mathbb{E}[x]$$

$$0 = \frac{\partial V}{\partial b} \implies -2\mathbb{E}[x(y - a - bx)] = 0 \implies b = \frac{\mathbb{E}[xy] - a\mathbb{E}[x]}{\mathbb{E}[x^2]}$$

$$b = \frac{\mathbb{E}[xy] - \mathbb{E}[y]\mathbb{E}[x] + b\mathbb{E}[x]^2}{\mathbb{E}[x^2]}$$

$$\implies b = \frac{\mathbb{E}[xy] - \mathbb{E}[y]\mathbb{E}[x]}{\mathbb{E}[x^2] - \mathbb{E}[x]^2} = \frac{Cov(x, y)}{Var(x)}$$

Back to our problem. Step 2: Apply Analogy Principle

$$\beta = \frac{Cov(x,y)}{Var(x)} = \frac{\mathbb{E}[xy] - \mathbb{E}[y]\mathbb{E}[x]}{\mathbb{E}[x^2] - \mathbb{E}[x]^2}$$

$$\hat{\beta} = \frac{\left(\frac{1}{n}\sum_{i=1}^{n} x_i y_i\right) - \left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)}{\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2}$$

$$\alpha = \mathbb{E}[y] - \beta\mathbb{E}[x]$$

$$\hat{\alpha} = \left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) - \hat{\beta}\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)$$

Multiple Regression

$$y = x'\beta + e, \mathbb{E}[e|x] = 0 \implies \mathbb{E}[y|x] = x'\beta$$

$$\beta = \arg\min_b \mathbb{E}[(y - \mathbb{E}[y|x])^2] = \arg\min_b \mathbb{E}[(y - x'b)^2]$$

$$\hat{\beta} = \arg\min_b \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'b)^2 = \arg\min_b (Y - Xb)^2$$

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}_{1\times n}, X = \begin{bmatrix} x_1' \\ \vdots \\ x_n' \end{bmatrix}_{n\times k}, b = \begin{bmatrix} b_1 \\ \vdots \\ b_k \end{bmatrix}_{k\times 1}$$

$$0 = \frac{\partial V}{\partial b} \implies 2X'(Y - Xb) = X'Y - X'Xb = 0$$

$$\hat{\beta} = (X'X)^{-1}X'Y \iff \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1}\left(\sum_{i=1}^{n} x_i y_i\right)$$

This is OLS.

Elements of Asymptotic Theory

Convergence

- Convergence of sequence. Suppose that S_n is some sequence than

$$S = \lim_{n\to\infty} S_n \iff \forall e > 0 : \exists N : |S_n - S| < e$$

- Convergence of sequence of RV $Z_n \overset{as}{\to} Z$ and/or $Z_n \overset{ms}{\to} Z \implies Z_n \overset{p}{\to} Z$

  1. Almost surely $Z_n \overset{as}{\to} Z \iff P\left\{\lim_{n\to\infty} Z_n = Z\right\} = 1$
  2. In Probability $Z_n \overset{p}{\to} Z \iff \forall e \lim_{n\to\infty} P\left\{||Z_n - Z|| > e\right\} = 0$
  3. In Mean Square $Z_n \overset{ms}{\to} Z \iff \lim_{n\to\infty} ||Z_n - Z||^2 = 0$
  4. In distribution $Z_n \overset{d}{\to} Z \iff P\left\{Z_n \le z\right\} = P\left\{Z \le z\right\}$

- Result

$$Z_n \overset{as}{\to} Z \text{ and/or } Z_n \overset{ms}{\to} Z \implies Z_n \overset{p}{\to} Z$$

Asymptotics

- Consistency: $\hat{\theta} \xrightarrow{p} \theta$
- Asymptotic Normality: $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, V_\theta)$
- Law of Large Numbers (Kolmogorov, IID):

$$\frac{1}{n}\sum_{i=1}^{n} Z_i \xrightarrow{as} \mathbb{E}[Z_i] \text{ as } n \to \infty$$

- Central Limit Theorem

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i - \mu\right) \xrightarrow{d} N(0, \sigma^2)$$
$$\mu = \mathbb{E}[Z_i], \sigma^2 = \mathbb{V}[Z_i]$$

Mann-Wald Theorem and Slutsky Theorem

- **Mann-Wald Theorem** Suppose that $g(z)$ is continuous $R \to \mathbb{R}$ function. Then

$$Z_n \xrightarrow{*} Z \implies g(Z_n) \xrightarrow{*} g(Z),$$

  where $* \in \{as, p, d, ms\}$ for ms $g(z)$ should be linear. Moreover, if Z is constant, than continuity only at Z suffices.

- **Slutsky Theorem**

$$\begin{cases} U_n \xrightarrow{p} U = const \\ V_n \xrightarrow{d} V \end{cases} \implies \begin{cases} U_n + V_n \xrightarrow{d} U + V \\ U_n V_n \xrightarrow{d} UV, \; V_n U_n \xrightarrow{d} VU \\ V_n / U_n \xrightarrow{d} V/U \end{cases}$$

Asymptotics of OLS: Unbiasedness and Variance

Recall that

$$Y = X\beta + e$$
$$\hat{\beta} = (X'X)^{-1}X'Y = (X'X)^{-1}X'(X\beta + e) = \beta + (X'X)^{-1}X'e$$
$$\mathbb{E}[\hat{\beta}|X] = \beta + (X'X)^{-1}X'\mathbb{E}[e|X] = \beta \implies \mathbb{E}[\hat{\beta}] = \beta$$

Therefore, coefficients are unbiased

$$\mathbb{V}[\hat{\beta}|X] = (X'X)^{-1}X'\underbrace{\mathbb{V}[e|X]}_{\Omega}X(X'X)^{-1}$$

Asymptotics of OLS: Consistency

- Consistency

$$\hat{\beta} = \beta + (X'X)^{-1}X'e = \beta + \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} x_i e_i$$

Recall LLN

$$\frac{1}{n}\sum_{i=1}^{n} x_i x_i' \xrightarrow{as} \mathbb{E}[x_i x_i'] = \mathbb{E}[xx']$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i e_i \xrightarrow{as} \mathbb{E}[x_i e_i] = \mathbb{E}[xe]$$

$$\implies \hat{\beta} \xrightarrow{p} \beta + \mathbb{E}[xx']^{-1}\mathbb{E}[xe]$$

$$\mathbb{E}[xe] = \mathbb{E}[x\mathbb{E}[e|x]] = 0 \implies \hat{\beta} \xrightarrow{p} \beta$$

Asymptotics of OLS: Asymptotic Normality

- Asymptotic Normality

$$\sqrt{n}(\hat{\beta} - \beta) = \left(\frac{1}{n} \sum_{i=1}^{n} x_i x_i' \right)^{-1} \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^{n} x_i e_i \right)$$

$$LLN : \frac{1}{n} \sum_{i=1}^{n} x_i x_i' \xrightarrow{p} \mathbb{E}[xx'] = Q_{xx}$$

$$CLT : \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^{n} x_i e_i \right) \xrightarrow{d} N(0, \sigma^2)$$

$$\text{Tot. Var.} : \mathbb{V}[x_i e_i] = \mathbb{E}[x_i x_i' \mathbb{V}[e_i|x]]$$

$$\mathbb{V}[e_i|x] = \mathbb{E}[(e_i - \mathbb{E}[e_i|x])(e_i - \mathbb{E}[e_i|x])'] = \mathbb{E}[e_i^2]$$

$$\implies \sigma^2 = \mathbb{E}[x_i x_i' e_i^2] = V_{xe}$$

$$\text{Slutsky:} \begin{cases} U_n \xrightarrow{p} U = const \\ V_n \xrightarrow{d} V \end{cases} \implies U_n V_n \xrightarrow{d} UV$$

Asymptotics of OLS: Asymptotic Normality

- Asymptotic Normality

$$\text{Slutsky: } \begin{cases} U_n \xrightarrow{p} U = const \\ V_n \xrightarrow{d} V \end{cases} \implies U_n V_n \xrightarrow{d} UV$$

$$U = Q_{xx}^{-1}, V = N(0, V_{xe})$$

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} Q_{xx}^{-1} N(0, V_{xe})$$

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, V_\beta)$$

$$V_\beta = Q_{xx}^{-1} V_{xe} Q_{xx}^{-1}$$

Asymptotics of OLS: Pivotization

- Apply Analogy Principle to V_β to get $\hat{V}_\beta$:

$$V_\beta = Q_{xx}^{-1} V_{xe} Q_{xx}^{-1}$$

$$\hat{Q}_{xx}^{-1} = \frac{1}{n}\sum_{i=1}^{n} x_i x_i' \xrightarrow{p} Q_{xx}$$

$$\hat{V}_{xe} = \frac{1}{n}\sum_{i=1}^{n} x_i x_i' \hat{e}_i^2 \xrightarrow{p} V_{xe}$$

$$\hat{e}_i^2 = (y_i - x_i'\hat{\beta})^2$$

$$\implies \hat{V}_\beta = \hat{Q}_{xx}^{-1} \hat{V}_{xe} \hat{Q}_{xx}^{-1} \xrightarrow{p} V_\beta$$

Asymptotics of OLS: T and CI

- Now we can formulate

$$se(\hat{\beta}) = \sqrt{\frac{1}{n}\left[\hat{V}_{\beta}\right]_{jj}}$$

$$t = \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \xrightarrow{d} N(0,1)$$

$$CI_{\alpha}(\mu) = \left[\hat{\beta}_j \mp se(\hat{\beta}_j) q_{1-\frac{\alpha}{2}}^{N(0,1)}\right]$$

Can we be more efficient: Gauss-Markov Theorem

- Consider

$$Y = X\beta + e, \mathbb{E}[e|X] = 0, \Omega = diag\{\sigma^2(x_i)\}_{i=1}^n$$

- **Theorem:** *Class of unbiased linear estimators* of β contains estimators $\mathcal{A}Y$, where $\mathcal{A}$ is $k \times n$ matrix depending only on X and having property that $\mathcal{A}X = I_k$.
  1. $\mathcal{A} = (X'X)^{-1}X' \implies$
  2. $\mathcal{A} = (X'WX)^{-1}X'W \implies$ WLS, where W is some $n \times n$ symmetric, positive definite matrix.
  3. $\mathcal{A} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1} \implies$ GLS. **Gauss-Markov Theorem** GLS is the efficient estimator in the class of unbiased linear estimators.

- So

$$\beta^{OLS} = (X'X)^{-1}X'Y, \beta^{WLS} = (X'WX)^{-1}X'WY, \beta^{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$$

Asymptotics of GLS

- Apply Analogy Principle!

$$\hat{\beta} = \left(\frac{1}{n} \sum_{i=1}^{n} \frac{x_i x_i'}{\sigma^2(x_i)} \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} \frac{x_i y_i}{\sigma^2(x_i)} = \beta + \left(\frac{1}{n} \sum_{i=1}^{n} \frac{x_i x_i'}{\sigma^2(x_i)} \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} \frac{x_i e_i}{\sigma^2(x_i)}$$

$$LLN : \begin{cases} \frac{1}{n} \sum_{i=1}^{n} \frac{x_i x_i'}{\sigma^2(x_i)} \xrightarrow{as} \mathbb{E}\left[\frac{xx'}{\sigma^2}\right] = Q_{xx/\sigma^2} \\ \frac{1}{n} \sum_{i=1}^{n} \frac{x_i e_i}{\sigma^2(x_i)} \xrightarrow{p} \mathbb{E}\left[\frac{xe}{\sigma^2}\right] = 0 \end{cases} \implies \hat{\beta} \xrightarrow{p} \beta$$

$$\mathbb{V}\left[\frac{xe}{\sigma^2(x)}\right] = \mathbb{E}\left[\left(\frac{xe}{\sigma^2(x)}\right)\left(\frac{xe}{\sigma^2(x)}\right)'\right] = \mathbb{E}\left[\frac{xx'\mathbb{E}[e^2|X]}{\sigma^4(x)}\right]$$

$$\mathbb{E}[e^2|X] \equiv \sigma^2(x) \implies \mathbb{V}\left[\frac{xe}{\sigma^2(x)}\right] = Q_{xx/\sigma^2}$$

$$CLT : \sqrt{n}\left(\frac{1}{n} \sum_{i=1}^{n} \frac{x_i e_i}{\sigma^2(x_i)}\right) \xrightarrow{d} N(0, Q_{xx/\sigma^2})$$

Obtaining FGLS

- Finally, Apply Slutsky's theorem

$$\implies \sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} Q_{xx/\sigma^2} N(0, Q_{xx/\sigma^2}) = N(0, Q^{-1}_{xx/\sigma^2})$$

- And here is FGLS.

$$\mathcal{A} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1},$$

where $\hat{\Omega} = diag\{\hat{\sigma}^2(x_i)\}_{i=1}^n$. And we just shown that

$$\hat{\beta}^{GLS} \sim N\left(\beta, \frac{1}{n}Q^{-1}_{xx/\sigma^2}\right), \hat{Q}^{-1}_{xx/\sigma^2} = X'\hat{\Omega}^{-1}X$$

Non Linear Least Squares

Consider the following models

$$\mathbb{E}[y|x] = g(x,\beta), \sigma^2(x) = \mathbb{V}(y|x)$$

1. Power regression $g(x,\beta) = \beta_1 + \beta_2 x^{\beta_3}$
2. Exponential regression $g(x,\beta) = \beta_1 + \beta_2 \exp(\beta_3 x)$
3. Probit regression $g(x,\beta) = \Phi(x'\beta)$
4. Threshold regression $g(x,\beta) = (\beta_1 + \beta_2 x_1)\mathbb{I}[x_2 \leq \beta_5] + (\beta_3 + \beta_4 x_1)\mathbb{I}[x_2 > \beta_5]$
5. Smooth threshold regression
   $g(x,\beta) = (\beta_1 + \beta_2 x_1)\Phi\left(\frac{x_2 - \beta_5}{\beta_6}\right) + (\beta_3 + \beta_4 x_1)\left[1 - \Phi\left(\frac{x_2 - \beta_5}{\beta_6}\right)\right]$

Quasi-Regressors

Quasi-regressor Assume that $g(x, b)$ is differentiable in b. Let $g_\beta(x, b) = \frac{\partial g(x,b)}{\partial b}$.
True quasi-regressors is $g_\beta(x, \beta)$.
Examples:

1. Linear regression $g_\beta(x, b) = x$
2. Power Regression

$$g_\beta(x, b) = \begin{pmatrix} 1 \\ x^{b_3} \\ b_2 x^{b_3} \ln x \end{pmatrix}$$

3. Probit regression

$$g_\beta(x, b) = \phi(x'b)x$$

Optimization Problem

$$\beta \in \arg\min_{b} \mathbb{E}[(y - g(x, b))^2]$$

What can go wrong?

- Does solution exists? Yes, it's convex minimization!
- Is solution unique? Not necessarily!
  1. *Global ID condition:* $Pr\{g(x, b_1) \neq g(x, b_2)\} > 0$ for any $\forall b_1 \neq b_2$
  2. *Local ID condition:* Q_{gg} is non-singular (or positive definite/invertible/full rank)

$$Q_{gg} = \mathbb{E}[g_\beta(x, \beta) g_\beta(x, \beta)']$$

Obtaining NLLS

Suppose ID holds and take FOCs

$$V = \min_b \mathbb{E}[(y - g(x, b))^2]$$
$$0 = \frac{\partial V}{\partial b} \implies \mathbb{E}[(y - g(x, b))g_\beta(x, b)] = 0$$

Apply Analogy principle

$$0 = \frac{\partial V}{\partial b} \implies \frac{1}{n}\sum_{i=1}^{n}(y_i - g(x_i, \hat{\beta}))g_\beta(x_i, \hat{\beta}) = 0$$

No Analytical Solution in general! Use `scipy.optimize.minimize` or smth.

Some Properties of NLLS

- NLLS is biased $\mathbb{E}[\hat{\beta}|X] \neq \beta$ (but actually this is not important)
- NLLS is consistent $\hat{\beta} \xrightarrow{p} \beta$ (can be shown by apply Taylor expansion up to 2nd order around true β)
- NLLS is asymptotically normal!

$$\sqrt{n}(\hat{\beta} - \beta) \sim N(0, V_\beta)$$

$$V_\beta \equiv Q_{gg}^{-1} V_{ge} Q_{gg}^{-1}$$

$$Q_{gg} = \mathbb{E}[g_\beta(x, \beta) g_\beta(x, \beta)'] \implies \hat{Q}_{gg} = \frac{1}{n} \sum_{i=1}^{n} g_\beta(x_i, \hat{\beta}) g_\beta(x_i, \hat{\beta})'$$

$$V_{ge} = \mathbb{E}[g_\beta(x, \beta) g_\beta(x, \beta)' e^2] \implies \hat{V}_{ge} = \frac{1}{n} \sum_{i=1}^{n} g_\beta(x_i, \hat{\beta}) g_\beta(x_i, \hat{\beta})' \hat{e}_i^2$$

$$\hat{e}_i = y_i - g(x_i, \hat{\beta})$$

Can we do better than NLLS?

- Yes. WNLLS is more efficient – read in spare time if interested.
- Also: Probit, Logit and etc. are all NLLS. I do not know why I put it here...

GMM

Idea of GMM: It's all about moments

- Suppose we have some distribution with some CDF $F(z; \theta_1, \dots, \theta_k)$
- We have a sample $\{z_i\}_{i=1}^n$ and we want to estimate $\theta_1, \dots, \theta_k$
- We can compute moments

$$\mathbb{E}[z] = \mu_1(\theta_1, \dots, \theta_k) = \int_{-\infty}^{+\infty} z f(z; \theta_1, \dots, \theta_k) dz$$

$$\mathbb{E}[z^k] = \mu_k(\theta_1, \dots, \theta_k) = \int_{-\infty}^{+\infty} z^k f(z; \theta_1, \dots, \theta_k) dz$$

- Apply Analogy Principle

$$m_k = \hat{\mu}_k(\theta_1, \dots, \theta_k) = \frac{1}{n} \sum_{i=1}^{n} z^k$$

and solve the system of equations

Example: Normal distribution

- Suppose that $z_i \sim$ i.i.d. $N(\mu, \sigma^2)$. And we want to estimate $\theta_1 = \mu$ and $\theta_2 = \sigma^2$.
- Moments are

$$\mu_1(\theta_1, \theta_2) = \theta_1$$
$$\mu_2(\theta_1, \theta_2) = \theta_2 + \theta_1^2$$
$$\Longrightarrow \begin{cases} \theta_1 = \mu_1 \\ \theta_2 = \mu_2 - \mu_1^2 \end{cases}$$

- Apply Analogy Principle

$$\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} z_i$$
$$\hat{\theta}_2 = \frac{1}{n}\sum_{i=1}^{n} z_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} z_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}(z_i^2 - \overline{z})$$

More general formulation

- Moment conditions

$$\mathbb{E}[m(z,\theta)] = 0,$$

  where θ is $k \times 1$ and $m(z,\theta)$ is $l \times 1$
- If $l = k$ we have *Just Identification*
- If $l > k$ we have *Overidentification*

Just Identitication: CMM

Apply Analogy Principle

$$\frac{1}{n}\sum_{i=1}^{n} m(z_i, \hat{\theta}) = 0$$

this produces a sysmem with l equations and $l = k$ unknowns $\implies$ unique solution.

Overidentification: Transformed GMM

Again we have

$$\frac{1}{n}\sum_{i=1}^{n} m(z_i, \hat{\theta}) = 0,$$

but know we have l equations and $l > k$ unknowns $\implies$ no solution. But we can transform the problem as

$$\theta = \arg\min_{q\in\Theta} \underbrace{\mathbb{E}[m(z,q)]'}_{1\times l} \underbrace{W}_{l\times l} \underbrace{\mathbb{E}[m(z,q)]}_{l\times 1},$$

where W is some $l \times l$ *weight matrix* (what?).
Know we can apply analogy principle and solve

$$\hat{\theta} = \arg\min_{q\in\Theta} \left(\frac{1}{n}\sum_{i=1}^{n} m(z_i, q)\right)' W_n \left(\frac{1}{n}\sum_{i=1}^{n} m(z_i, q)\right), W_n \xrightarrow{p} W$$

GMM: Identification Conditions

- *Global Identification condition* $\mathbb{E}[m(z,q)] = 0 \iff q = \theta$
- *Local Identification condition* $l \times k$ matrix of expected derivatives and $l \times l$ moment variance matrix have full ranks k and l

$$Q_{\partial m} = \mathbb{E}\left[\frac{\partial m(z,\theta)}{\partial q'}\right]$$

$$V_m = \mathbb{E}\left[m(z,\theta)m(z,\theta)'\right]$$

Asymptotics of GMM

Under ID conditions

1. GMM is consistent: $\hat{\beta} \xrightarrow{p} \beta$
2. GMM is Asymptotically Normal:

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, V_\theta),$$

where

$$V_\theta = (Q'_{\partial m} W Q_{\partial m})^{-1} Q'_{\partial m} W V_m W Q_{\partial m} (Q'_{\partial m} W Q_{\partial m})^{-1}$$

with exact identification it collapses to

$$V_\theta^{exact} = Q_{\partial m}^{-1} V_m Q_{\partial m}'^{-1}$$

Efficient GMM

- Can we select W such that V_θ is minimized?
- Yes! It turns out that

$$W^{opt} = V_m^{-1}$$
$$V_\theta^{opt} = \left(Q'_{\partial m} V_m^{-1} Q_{\partial m}\right)^{-1}$$

- But! GMM estimator that uses $W^{opt} = V_m^{-1}$ is *efficient GMM estimator*, but V_m^{-1} is unknown :(But we can use the *estimated weight matrix* W_n, sample analog of V_m^{-1} such that

$$W_n \xrightarrow{p} V_m^{-1}$$

This leads to 2-step GMM

1. Compute preliminary consistent estimate of $\hat{\theta}_0$ of θ (inefficient GMM with weight matrix I_n). Using it, construct consistent estimate $\hat{V}_m$
2. Compute feasible efficient GMM estimate using weight matrix $W_n = \hat{V}_m^{-1}$

$$\hat{\theta} = \arg\min_{q\in\Theta} \left(\frac{1}{n}\sum_{i=1}^{n} m(z_i, q)\right)' \hat{V}_m^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} m(z_i, q)\right)$$

and compute estimate of asymptotic variance

$$\hat{V}_\theta = \left(\hat{Q}'_{\partial m} \hat{V}_m^{-1} \hat{Q}_{\partial m}\right)^{-1}$$

Example of GMM

Problem: z is non-skewed and we want to estimate mean of z. True parameter $\theta = \mathbb{E}[z]$. How to apply GMM?

Step 1: Formulat Moment Restrictions

1. z is non-skewed $\implies \mathbb{E}[(z-\theta)^3] = 0$
2. True parameter is $\theta = \mathbb{E}[z] \implies \mathbb{E}[z-\theta] = 0$

This leads to

$$\mathbb{E}[m(z,q)] = \begin{bmatrix} z-q \\ (z-q)^3 \end{bmatrix} = 0$$

Therefore, we have $l = 2$ conditions and $k = 1$ parameters (q) $\implies l > k \implies$ overidentification!

Step (?): Estimate CMM

Just for fun ignore 2nd condition and apply CMM

$$m(z, q) = z - q$$

$$\frac{1}{n}\sum_{i=1}^{n} m(z_i, \hat{\theta}) = 0 \implies \hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} z_i$$

Just for fun ignore 1st condition and apply CMM

$$m(z, q) = (z - q)^3 \implies \frac{1}{n}\sum_{i=1}^{n}(z_i - \hat{\theta}_1)^3 = 0$$

$\implies$ two estimates!!! Can we do better?

Step 2: Formulate GMM optimization problem

If we apply GMM

$$\mathbb{E}\begin{bmatrix} z-q \\ (z-q)^3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\hat{\theta} = \arg\min_q \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n z_i - q & \frac{1}{n}\sum_{i=1}^n (z_i-q)^3 \end{bmatrix} W_n \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n z_i - q \\ \frac{1}{n}\sum_{i=1}^n (z_i-q)^3 \end{bmatrix}$$

What is W_n???

Step 3: Obtain W_n

We know that optimal

$$W^{opt} = V_m^{-1}$$

Denote $\mu_r = \mathbb{E}[(z-\theta)^r]$. Then

$$Q_{\partial m} = \mathbb{E}\left[\frac{\partial m(z,\theta)}{\partial q'}\right] = \mathbb{E}\begin{bmatrix} -1 \\ -3(z-q)^2 \end{bmatrix}_{q=\theta} = -\begin{bmatrix} 1 \\ 3\mu_2 \end{bmatrix}$$

$$V_m = \mathbb{E}\left[m(z,\theta)m(z,\theta)'\right] = \begin{bmatrix} (z-q)^2 & (z-q)^4 \\ (z-q)^4 & (z-q)^6 \end{bmatrix}_{q=\theta} = \begin{bmatrix} \mu_2 & \mu_4 \\ \mu_4 & \mu_6 \end{bmatrix}$$

And apply Analogy principle

$$W^{opt} = \hat{V}_m^{-1} = \begin{bmatrix} \hat{\mu}_2 & \hat{\mu}_4 \\ \hat{\mu}_4 & \hat{\mu}_6 \end{bmatrix}^{-1}$$

Step 4: Formulate efficient GMM optimization problem

$$\hat{\theta} = \arg\min_q \left[\frac{1}{n}\sum_{i=1}^n z_i - q \quad \frac{1}{n}\sum_{i=1}^n (z_i - q)^3\right] \begin{bmatrix} \hat{\mu}_2 & \hat{\mu}_4 \\ \hat{\mu}_2 & \hat{\mu}_6 \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n z_i - q \\ \frac{1}{n}\sum_{i=1}^n (z_i - q)^3 \end{bmatrix}$$

Solve to obtain $\hat{\theta}$. For variance recall the formula

$$V_\theta = \left(Q'_{\partial m} V_m^{-1} Q_{\partial m}\right)^{-1} = \left(\begin{bmatrix} -1 & -3\mu_2 \end{bmatrix} \begin{bmatrix} \mu_2 & \mu_4 \\ \mu_4 & \mu_6 \end{bmatrix}^{-1} \begin{bmatrix} -1 \\ -3\mu_2 \end{bmatrix}\right)^{-1}$$

This is much more efficient than CMM.

Instrumental Variables

Problem

Suppose we have

$$y = x'\beta + e$$

But $\mathbb{E}[e|x] \neq 0$ or $\mathbb{E}[y|x] \neq x'\beta \implies$ not regression!
$\mathbb{E}[ex] \neq 0 \implies$ not even BLP!.
x is endogenous in these cases.

- Simultaneity $\implies$ endogeneity
- Unaccounted common factors $\implies$ endogeneity
- Errors in variables $\implies$ endogeneity

Example: Demand and Supply

Suppose we have

$$\text{Demand: } P = \gamma_D Q + e_D$$
$$\text{Supply: } P = \gamma_S Q + e_S,$$

where e_D, e_S are independent and $N(0, 1)$. and we want to estimate OLS. What can go wrong? What will coefficients in OLS regression show us? Will they be consistent? Asymptotically normal?

Let's calculate regression by hand

$$\beta = \mathbb{E}[xx']^{-1}\mathbb{E}[xy] = \frac{\mathbb{E}[xy]}{\mathbb{E}[xx']}$$

Therefore, if we apply this to our 2 equations

$$\beta_P = \frac{\mathbb{E}[PQ]}{\mathbb{E}[P^2]}$$
$$\beta_Q = \frac{\mathbb{E}[PQ]}{\mathbb{E}[Q^2]}$$

Let's calculate regression by hand

Solve the system of equations to obtain

$$\gamma_D Q + e_D = \gamma_S Q + e_S$$
$$Q = \frac{e_S - e_D}{\gamma_D - \gamma_S}$$
$$P = \frac{e_S \gamma_D - \gamma_S e_D}{\gamma_D - \gamma_S}$$

Therefore,

$$PQ = \frac{(e_S - e_D)(e_S \gamma_D - \gamma_S e_D)}{(\gamma_D - \gamma_S)^2} = \frac{e_S^2 \gamma_D - e_D e_S \gamma_D - \gamma_S e_D e_S + \gamma_S e_D^2}{(\gamma_D - \gamma_S)^2}$$

Take expectation (recall that e_D, e_S are independent and $N(0,1)$)

$$\mathbb{E}[PQ] = \frac{\gamma_D + \gamma_S}{(\gamma_D - \gamma_S)^2}$$
$$\mathbb{E}[Q^2] = \frac{2}{(\gamma_D + \gamma_S)^2}, \mathbb{E}[P^2] = \frac{\gamma_D^2 + \gamma_S^2}{(\gamma_D - \gamma_S)^2}$$

Let's calculate regression by hand

Thus, we have

$$\begin{cases} \beta_P = \frac{\mathbb{E}[PQ]}{\mathbb{E}[P^2]}, \beta_Q = \frac{\mathbb{E}[PQ]}{\mathbb{E}[Q^2]} \\ \mathbb{E}[Q^2] = \frac{2}{(\gamma_D+\gamma_S)^2}, \mathbb{E}[P^2] = \frac{\gamma_D^2+\gamma_S^2}{(\gamma_D-\gamma_S)^2} \\ \mathbb{E}[PQ] = \frac{\gamma_D+\gamma_S}{(\gamma_D-\gamma_S)^2} \end{cases}$$

$$\implies \beta_P = \frac{\gamma_D + \gamma_S}{\gamma_D^2 + \gamma_S^2}$$

$$\implies \beta_Q = \frac{\gamma_D + \gamma_S}{2}$$

Not quite the thing that we expected.... How to even enterpret it? Possible we can

$$\gamma_D = 2\beta_Q - \gamma_S$$

$$2\gamma_S^2 - 4\beta_Q\gamma_S = 2\frac{\beta_Q}{\beta_P} - 4\beta_Q^2$$

But there will be two roots.....

Instruments

$$y = x'\beta + e$$

We have $l \times 1$ instrument z. We need

- Instrument *validity*

$$\mathbb{E}[e|z] = 0$$

- Instrument *relevance*

$$\underset{k\times l}{Q_{xz}} = \mathbb{E}[xz'] \text{ is full rank } k \implies l \geq k$$

- Instrument *non-collinearity*

$$\underset{k\times l}{Q_{zz}} = \mathbb{E}[zz'] \text{ is full rank } l$$

IV: Exact Identification

$l = k$. Instrument validity implies moment condition

$$\mathbb{E}[z(y - x'\beta)] = 0$$

Use **CMM** (yeaaaah) to yield instrumental variable estimator

$$\hat{\beta}_{IV} = \left(\sum_{i=1}^{n} z_i x_i'\right)^{-1} \sum_{i=1}^{n} z_i y_i$$

IV: Asymptotics

- IV estimator is consistent $\hat{\beta}_{IV} \xrightarrow{p} \beta$
- IV estimator is asymptotically normal

$$\sqrt{n}(\hat{\beta}_{IV} - \beta) \xrightarrow{d} N(0, V_\beta)$$

- IV asymptotic variance is

$$V_\beta = (Q'_{xz})^{-1} V_{ze} Q_{zx}^{-1}$$
$$Q_{xz} = \mathbb{E}[xz'], V_{ze} = \mathbb{E}[zz'e^2]$$

IV: Overidentification – 2SLS

$l > k$. Instrument validity implies moment condition

$$\begin{aligned} \mathbb{E}[z(y - x'\beta)] &= 0 \\ \implies m(x, y, z, \beta) &= z(y - x'\beta) \\ \implies Q_{\partial m} &= -Q'_{xz} \\ \implies V_m = \mathbb{E}[zz'(y - x'\beta)^2] &= \mathbb{E}[zz'e^2] = V_{ze} \end{aligned}$$

Let's use symmetric positive definite but *non-optimal* weight matrix

$$W_n^{2sls} = \left(\frac{1}{n} \sum_{i=1}^{n} z_i z_i' \right)^{-1} \xrightarrow{p} Q_{zz}^{-1} \not\propto V_m^{-1}$$

GMM yields 2SLS

$$\hat{\beta}_{2SLS} = \arg\min_b \left(\frac{1}{n}\sum_{i=1}^n z_i(y_i - x_i'b)\right)' \left(\frac{1}{n}\sum_{i=1}^n z_i z_i'\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^n z_i(y_i - x_i'b)\right)$$

$$= \left(\sum_{i=1}^n x_i z_i' \left(\sum_{i=1}^n z_i z_i'\right)^{-1} \sum_{i=1}^n z_i x_i'\right)^{-1} \sum_{i=1}^n x_i z_i' \left(\sum_{i=1}^n z_i z_i'\right)^{-1} \sum_{i=1}^n z_i y_i$$

$$= (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'Y$$

Why 2???? SLS

Observe that

$$\hat{\beta}_{2SLS} = (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'Y$$
$$= (\hat{X}'X)^{-1}\hat{X}'Y$$
$$\hat{X} = Z(Z'Z)^{-1}Z'X = \mathcal{BLP}(X|Z)$$

Therefore, we have 2 stages

1. Project x on z and obtain fitted values $\hat{x}$
2. Use IV estimation with $\hat{x}$ as exactly identifying instrument

But *2SLS is inefficient GMM*

Efficient GMM

$$W_n = \hat{V}_m^{-1} = \left(\frac{1}{n} \sum_{i=1}^{n} z_i z_i' \hat{e}_{i,2sls}^2 \right)^{-1}$$

$\hat{e}_i$ are consistent (for example, 2SLS) residuals

$$\begin{aligned}
\hat{\beta}_{GMM} &= \arg\min_b \left(\frac{1}{n} \sum_{i=1}^{n} z_i (y_i - x_i' b) \right)' \left(\frac{1}{n} \sum_{i=1}^{n} z_i z_i' \hat{e}_i^2 \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} z_i (y_i - x_i' b) \right) \\
&= \left(\sum_{i=1}^{n} x_i z_i' \left(\sum_{i=1}^{n} z_i z_i' \hat{e}_i^2 \right)^{-1} \sum_{i=1}^{n} z_i x_i' \right)^{-1} \sum_{i=1}^{n} x_i z_i' \left(\sum_{i=1}^{n} z_i z_i' \hat{e}_i^2 \right)^{-1} \sum_{i=1}^{n} z_i y_i \\
&= (X'Z((Z \cdot \hat{e})'(Z \cdot \hat{e}))^{-1} Z'X)^{-1} X'Z((Z \cdot \hat{e})'(Z \cdot \hat{e}))^{-1} Z'Y
\end{aligned}$$

Asymptotics of GMM

- GMM is consistent $\hat{\beta}_{GMM} \xrightarrow{p} \beta$
- GMM is asymptotically normal

$$\sqrt{n}(\hat{\beta}_{GMM} - \beta) \xrightarrow{d} N(0, V_\beta)$$

- Asymptotic variance is

$$V_\beta = (Q_{xz} V_{ze}^{-1} Q_{xz}')^{-1}$$
$$Q_{xz} = \mathbb{E}[xz'], V_{ze} = \mathbb{E}[zz'e^2]$$

What if we have Underidentification?

Instrument is relevant if

$$\underset{k\times l_{xz}}{Q} = \mathbb{E}[xz'] \text{ is ful rank } k$$

If no relevance. Example $l = k = 1$ and $\mathbb{E}[xz] = 0$, that is Q_{zk} has rank 0. By CLT

$$\frac{1}{n}\sum_{i=1}^{n} z_i \begin{bmatrix} e_i \\ x_i \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} V_{ze} & C_{zxe} \\ C_{zxe} & V_{zx} \end{bmatrix}\right)$$

Then for IV estimator

$$\hat{\beta}_{IV} = \frac{\sum z_i y_i}{\sum z_i x_i} = \beta + \frac{\frac{1}{\sqrt{n}}\sum_i z_i e_i}{\frac{1}{\sqrt{n}}\sum_i z_i x_i} \xrightarrow{d} \beta + \mathcal{D} \overset{d}{\neq} \beta$$

Therefore, pre-test for instrument irrelevance $\mathbb{E}[xz] = 0$ and run IV only if rejected.

The most fun part

GMM goes brrrrrrr

1. $m(z,\theta) = x(y - x'\beta) \implies$ OLS
2. $m(z,\theta) = \frac{x(y-x'\beta)}{\sigma^2(x)} \implies$ GLS
3. $m(z,\theta) = g_\beta(x,\beta)(y - g(x,\beta) \implies$ NLLS
4. $m(z,\theta) = \frac{g_\beta(x,\beta)(y-g(x,\beta)}{\sigma^2(x)} \implies$ WNLLS
5. $m(z,\theta) = \frac{\partial \ln f(y|x,\theta)}{\partial q} \implies$ MLE, QMLE
6. $m(z,\theta) = z(y - x'\beta) \implies$ IV

$\implies$ All parametric estimators are GMM estimators! Just specify the moment function correctly and obtain efficient weight matrix.

What you actually can do now

Suppose we have the following model

$$Y = AK^{\alpha}L^{\gamma} + e, \mathbb{E}[e|K, L] = 0$$
$$K|L \sim Exp(\lambda), \lambda = \gamma L^{\eta}$$

1. Estimate α, γ, η and derive their asymptotic properties
2. Suppose know that K is not observed – can we still estimate α, γ, η?
3. Suppose we made mistake and $K|L$ is not exponentially distributed. What goes wrong?

Hints: My approach

1. Step 1. Formulate Conditional MLE estimator for γ, η from $K|L \sim Exp(\gamma L^{\eta})$
2. Step 2. Formulate NLLS estimator for $\mathbb{E}[Y|K, L] = AK^{\alpha}L^{\gamma}$
3. Step 3. Recall that both NLLS and MLE are examples of GMM and formulate the system of moment conditions. Hint: you will have $l = 4, k = 3 \implies$ overidentification
4. Step 4. Estimate efficient GMM. How would you obtain W^{opt}?