Intermediate Econometrics
Homeworks
Máté Kormos

Compiled on October 1, 2017

This note is written to cover some of the homeworks without the aim of being comprehensive. Some comments are included as well.

1. Week 1, lecture 1 (Sep 19)

Topic: OLS basics summary

The Basic model:

$$y = X\beta + u$$

or

$$y_i = x_i'\beta + u_i$$

where $y = [y_1, \ldots, y_n]$ is the vector of observations of the dependent variable, X is an $(n \times K)$ matrix of the observations of the explanatory variables, β is the $(K \times 1)$ vector of unknown parameters, and u is the $(n \times 1)$ vector of the disturbances. Assumptions:

- 1. $u \sim (0, \sigma^2)$ i.i.d.
- 2. $\mathbb{E}[X'u] = 0$
- 3. $u \sim N(0, \sigma^2)$

The OLS estimator

$$(y - X\beta)'(y - X\beta) \to \min$$

$$\hat{\beta} = (X'X)^{-1}X'y \qquad \hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{n-k}$$

Basic Properties assuming 1. and 2.: Finite sample

$$\mathbb{E}\left[\hat{\beta}\right] = (X'X)^{-1}X'(X\beta + u) = \beta$$

$$\mathrm{Var}\,(\hat{\beta}) = \mathbb{E}\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right] = \mathbb{E}\left[(X'X)^{-1}X'uu'X(X'X)^{=1}\right] = \sigma^2(X'X)^{-1}$$

So $\hat{\beta} \sim (\beta, \sigma^2(X'X)^{-1})$ and if 3. also true then $\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$

Asymptotic properties

Assumptions:

- 4. $\operatorname{plim} X'u = 0$

- 5. $\frac{X'X}{n} = Q$ finite, positive definite matrix
- 6. $\text{plim}(uu') = \sigma^2 I$

Then

$$\text{plim}\,\hat{\beta} = \text{plim}\,\frac{(X'X)^{-1}}{n}\,\text{plim}\,\frac{X'u}{n} = 0$$

$$\text{plim}\,\text{Var}\left[\hat{\beta}\right] = \text{plim}(\hat{\beta}-\beta)(\hat{\beta}-\beta)'$$

$$\text{plim}(X'X)^{-1}X'\,\text{plim}(uu')\,\text{plim}\,X(X'X)^{-1} \quad = \quad \frac{\sigma^2}{n}\frac{(X'X)^{-1}}{n} = 0$$

The limiting distribution of the OLS (no need nor the normality assumption 3. here)

$$\hat{\beta} \xrightarrow{d} N(\beta, 0)$$

In order to have a finite covariance matrix here we need to normalize and look at plim $\sqrt{n}(\hat{\beta}-\beta)$

$$\text{plim}\,\sqrt{n}(\hat{\beta}-\beta) \quad \overset{A}{\sim} \quad N(0, \sigma^2 Q^{-1})$$

which is the asymptotic distribution of the OLS estimator.

Extensions - Non-scalar covariance matrix

When we no longer maintain assumption 1. then $\mathbb{E}\left[(uu')\right] \neq \sigma^2 I$, but $\mathbb{E}\left[(uu')\right] = \Sigma$. Then the covariance matrix of the OLS estimator is

$$\mathbb{E}\left[\underbrace{(X'X)^{-1}X'uu'X(X'X)^{-1}}_{\Omega}\right] = (X'X)^{-1}X'\Sigma X(X'X)^{-1}$$

where Σ is unknown. In the case of heteroskedasticity Σ is diagonal, with $\sigma_1^2, \ldots, \sigma_n^2$ in the diagonal elements. White has shown that the

$$(X'X)^{-1}X'\text{diag}[\hat{u}_1^2, \ldots, \hat{u}_n^2]X(X'X)^{-1}$$

is a consistent estimator of the Ω matrix, where $\hat{u}_i^2$ are the OLS residuals, so can be used to get the asymptotic standard errors of the OLS estimator in case of heteroscedasticity.

1.1. Comments

Model identification.

A structure is a function which assigns numerical values to the model parameters, formally $g : \theta \in \Theta \mapsto \mathbb{R}^{dim(\theta)}$. Then a model is identified iff there is no structure $h \neq g$ such that $f_g(\boldsymbol{y} \mid \boldsymbol{X}) = f_h(\boldsymbol{y} \mid \boldsymbol{X})$. That is to say, for the same $\boldsymbol{X}$, there are no two different structures which produce the same conditional density for $\boldsymbol{y}$.

E.g. probit model: $y_i := \mathbb{1}(y_{i1} - y_{i0} > 0) = \mathbb{1}(\boldsymbol{x_i'\beta} - u_i > 0)$, $u_i \sim \mathcal{N}(0,1)$, where we only observe y_i and $\boldsymbol{x_i}$. Altenatively consider $y_i := \mathbb{1}(z_{i1} - z_{i0} > 0) = \mathbb{1}(\sigma(\boldsymbol{x_i'\beta} - u_i) > 0) := \mathbb{1}(\boldsymbol{x_i'\delta} + v_i > 0) = \mathbb{1}(\boldsymbol{x_i'\beta} - u_i > 0)$ for $\sigma > 0$ and $v_i \sim \mathcal{N}(0, \sigma^2)$. Given the same $\boldsymbol{x_i}$, the two model $(\boldsymbol{\beta}, \mathrm{Var}\left[\text{disturbance}\right] = 1)$ and $(\boldsymbol{\delta}, \mathrm{Var}\left[\text{disturbance}\right] = \sigma^2)$ will generate the same y_i, so we cannot tell apart the two models, cannot separate the scale.

1.2. Tasks

1. Show that on violation of the homoskedasticity assumption, OLS stops being BLUE inasmuch as it won't have the smallest variance.

   There are two ways to show this. First, to show that in this case the OLS covariance matrix does not reaches the Cramer-Rao lower bound, so it is not the best. Or derive another estimator and show that it is better, has smaller covariance matrix, than the OLS.

   Se here we present another estimator, GLS, which is linear, unbiased, and has a samller covariance matrix than OLS. The GLS first transforms the model from $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}$ to $\boldsymbol{Wy} = \boldsymbol{WX\beta} + \boldsymbol{Wu}$, or with another notation $\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{u}}$ (note that $\boldsymbol{W}$ is invertible, hence the transformation is valid). The weight matrix $\boldsymbol{W}$ is constructed to eliminate the heteroskedasticity from the model:

$$\boldsymbol{W} = \begin{bmatrix} \frac{1}{\sigma_1} & & & \\ & \frac{1}{\sigma_2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_n} \end{bmatrix}, \tag{1}$$

$$\sigma_i = \mathbb{E}\left[u_i^2\right]^{1/2} = \mathrm{Var}\left[u_i\right]^{1/2}. \tag{2}$$

Consequently, the new error term will exhibit unit variance:

$$
\begin{aligned}
\operatorname{Var}[\tilde{\boldsymbol{u}}] &= \mathbb{E}\left[\tilde{\boldsymbol{u}}\tilde{\boldsymbol{u}}'\right] && (3)\\
&= \mathbb{E}\left[\boldsymbol{W}\boldsymbol{u}\boldsymbol{u}'\boldsymbol{W}'\right] && (4)\\
&= \boldsymbol{W}\,\mathbb{E}\left[\boldsymbol{u}\boldsymbol{u}'\right]\boldsymbol{W}' && (5)\\
&= \boldsymbol{W}\boldsymbol{\Sigma}\boldsymbol{W}' && (6)\\
&= \begin{bmatrix} \frac{1}{\sigma_1} & & & \\ & \frac{1}{\sigma_2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_n} \end{bmatrix}
\begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_n^2 \end{bmatrix}
\begin{bmatrix} \frac{1}{\sigma_1} & & & \\ & \frac{1}{\sigma_2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_n} \end{bmatrix} && (7)\\
&= \boldsymbol{I} && (8)\\
\boldsymbol{W} &= \boldsymbol{\Sigma}^{-1/2} = \boldsymbol{W}', \qquad \boldsymbol{W}\boldsymbol{W}' = \boldsymbol{\Sigma}^{-1}. && (9)
\end{aligned}
$$

The GLS estimates the tranformed model with OLS, which makes it a linear estimator. The unbiasedness is not proven here but the proof is pretty much the same as for the OLS, but this with $\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{u}}$. Now, let's compare the covariance matrices. First, the OLS estimator has

$$
\begin{aligned}
\operatorname{Var}\left[\hat{\boldsymbol{\beta}}^{OLS}\right] &= \mathbb{E}\left[\left(\hat{\boldsymbol{\beta}}^{OLS} - \mathbb{E}\left[\hat{\boldsymbol{\beta}}^{OLS}\right]\right)\left(\hat{\boldsymbol{\beta}}^{OLS} - \mathbb{E}\left[\hat{\boldsymbol{\beta}}^{OLS}\right]\right)'\right] && (10)\\
&= \mathbb{E}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}\boldsymbol{u}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\right] \quad \text{OLS still unbaised} && (11)\\
&\overset{L.I.E.}{=} \mathbb{E}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\mathbb{E}\left[\boldsymbol{u}\boldsymbol{u}' \mid \boldsymbol{X}\right]\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\right] && (12)\\
&= \mathbb{E}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\right]. && (13)
\end{aligned}
$$

Second, the GLS, the estimate of which is $\hat{\boldsymbol{\beta}}^{GLS} = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{W}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}'\boldsymbol{W}\boldsymbol{u}$, has

$$
\begin{aligned}
\operatorname{Var}\left[\hat{\boldsymbol{\beta}}^{GLS}\right] &= \mathbb{E}\left[\left(\hat{\boldsymbol{\beta}}^{GLS} - \mathbb{E}\left[\hat{\boldsymbol{\beta}}^{GLS}\right]\right)\left(\hat{\boldsymbol{\beta}}^{GLS} - \mathbb{E}\left[\hat{\boldsymbol{\beta}}^{GLS}\right]\right)'\right] && (14)\\
&= \mathbb{E}\left[(\boldsymbol{X}'\boldsymbol{W}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}'\boldsymbol{W}\boldsymbol{u}\boldsymbol{u}'\boldsymbol{W}'\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{W}'\boldsymbol{W}\boldsymbol{X}')^{-1}\right] && (15)\\
&= \mathbb{E}\left[(\boldsymbol{X}'\boldsymbol{W}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}'\,\mathbb{E}\left[W'uu'W \mid \boldsymbol{X}\right]'\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{W}'\boldsymbol{W}\boldsymbol{X}')^{-1}\right] && (16)\\
&= \mathbb{E}\left[(\boldsymbol{X}'\boldsymbol{W}'\boldsymbol{W}\boldsymbol{X})^{-1}\right] && (17)\\
&= \mathbb{E}\left[(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\right]. && (18)
\end{aligned}
$$

Then we compare the covariance matrices: if Var $\left[\hat{\boldsymbol{\beta}}^{OLS}\right]$ − Var $\left[\hat{\boldsymbol{\beta}}^{GLS}\right]$ is positive definite, we showed GLS to be more efficient.

$$\text{Var}\left[\hat{\boldsymbol{\beta}}^{OLS}\right] - \text{Var}\left[\hat{\boldsymbol{\beta}}^{GLS}\right] = \mathbb{E}\left[(\boldsymbol{X'X})^{-1}\boldsymbol{X'\Sigma X}(\boldsymbol{X'X})^{-1}\right] - \mathbb{E}\left[(\boldsymbol{X'\Sigma}^{-1}\boldsymbol{X})^{-1}\right] \tag{19}$$

$$\mathbb{E}\left[(\boldsymbol{X'X})^{-1}\boldsymbol{X'\Sigma X}(\boldsymbol{X'X})^{-1} - (\boldsymbol{X'\Sigma}^{-1}\boldsymbol{X})^{-1}\right]. \tag{20}$$

To simplify our proof, let's assume that there is only one regressor. This we can do without loss of generality: the estimator's properties do not depend on the number of regressors in this setting. So with $\boldsymbol{X} = \boldsymbol{x} \in \mathcal{M}^{n\times 1}$

$$(\boldsymbol{X'X})^{-1}\boldsymbol{X'\Sigma X}(\boldsymbol{X'X})^{-1} = \frac{\sum_i x_i^2\sigma_i^2}{(\sum_i x_i^2)(\sum_i x_i^2)} \tag{21}$$

$$(\boldsymbol{X'\Sigma}^{-1}\boldsymbol{X})^{-1} = \left(\sum_i \frac{x_i^2}{\sigma_i^2}\right)^{-1} = \frac{1}{\sum_i \frac{x_i^2}{\sigma_i^2}}, \quad \text{so the difference} \tag{22}$$

$$\frac{\sum_i x_i^2\sigma_i^2}{(\sum_i x_i^2)(\sum_i x_i^2)} - \frac{1}{\sum_i \frac{x_i^2}{\sigma_i^2}} = \frac{\left(\sum_i \frac{x_i^2}{\sigma_i^2}\right)(\sum_i x_i^2\sigma_i^2) - (\sum_i x_i^2)(\sum_i x_i^2)}{(\sum_i x_i^2)(\sum_i x_i^2)\left(\sum_i \frac{x_i^2}{\sigma_i^2}\right)} \tag{23}$$

$$= \frac{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} x_i^2 x_j^2 \left(\frac{\sigma_j^2}{\sigma_i^2} + \frac{\sigma_i^2}{\sigma_j^2} - 2\right)}{(\sum_i x_i^2)(\sum_i x_i^2)\left(\sum_i \frac{x_i^2}{\sigma_i^2}\right)}, \tag{24}$$

where the last result is obtained by some rather soul-destroying arithmetic. We can be reasonably happy now as $(\frac{\sigma_j^2}{\sigma_i^2} + \frac{\sigma_i^2}{\sigma_j^2}$ is bounded from below, with a global minimum of 2, reached at $\sigma_i = \sigma_j$. Therefore, as the denominator is always positive, the numerator will be larger than or equal to 0, reaching 0 only in the homoskedastic case ($\sigma_i = \sigma_j\ \forall(i,j)$), so the expectation will also be positive in the heteroskedastic case $\iff$ OLS exhibits larger variance than GLS, our claim is proved.

2. Derive the finite and the limiting distribution of $\hat{\boldsymbol{\beta}}^{OLS}$. By definition of OLS we have

$$\hat{\boldsymbol{\beta}}^{OLS} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'y} \tag{25}$$

$$= \boldsymbol{\beta} + (\boldsymbol{X'X})^{-1}\boldsymbol{X'u} \tag{26}$$

$$= \boldsymbol{\beta} + \left(\sum_{i=1}^{n} \boldsymbol{x_i x_i'}\right)^{-1} \sum_{i=1}^{n} \boldsymbol{x_i} u_i \tag{27}$$

$$= \boldsymbol{\beta} + \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x_i x_i'}\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x_i} u_i. \tag{28}$$

Limiting distribution.

Use convergence properties (i) $\operatorname{plim} A = a,\ B \xrightarrow{d} b \implies AB \xrightarrow{d} ab$ (ii) $\operatorname{plim} \bar{c} = \mathbb{E}[c]$ (iii) continuous mapping of plim, and apply the central limit theorem:

$$\begin{aligned}
\operatorname{plim}\left(\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x_i}\boldsymbol{x}_i'\right)^{-1}\right) &= \mathbb{E}\left[\boldsymbol{x}\boldsymbol{x'}\right]^{-1} && (29)\\
\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x_i}u_i &\xrightarrow{d}\mathcal{N}\left(\mathbb{E}\left[\boldsymbol{x}u\right], \frac{\operatorname{Var}\left[\boldsymbol{x}u\right]}{n}\right), && (30)\\
\operatorname{Var}\left[\boldsymbol{x}u\right] &= \mathbb{E}\left[(\boldsymbol{x}u-\mathbb{E}\left[\boldsymbol{x}u\right])(\boldsymbol{x}u-\mathbb{E}\left[\boldsymbol{x}u\right])'\right] && (31)\\
&= \mathbb{E}\left[\boldsymbol{x'}uu'\boldsymbol{x'}\right] && (32)\\
&= \sigma^2\,\mathbb{E}\left[\boldsymbol{x}\boldsymbol{x'}\right] \implies && (33)
\end{aligned}$$

$$\begin{aligned}
\hat{\boldsymbol{\beta}}^{OLS} &\xrightarrow{d}\boldsymbol{\beta}+\mathbb{E}\left[\boldsymbol{x}\boldsymbol{x'}\right]^{-1}\mathcal{N}\left(\boldsymbol{0}, \frac{\mathbb{E}\left[\boldsymbol{x}\boldsymbol{x'}\right]}{n}\right) && (34)\\
&\xrightarrow{d}\mathcal{N}\left(\boldsymbol{\beta}, \frac{\sigma^2}{n}\,\mathbb{E}\left[\boldsymbol{x}\boldsymbol{x'}\right]^{-1}\mathbb{E}\left[\boldsymbol{x}\boldsymbol{x'}\right]\mathbb{E}\left[\boldsymbol{x}\boldsymbol{x'}\right]^{-1}\right) && (35)\\
&\xrightarrow{d}\mathcal{N}\left(\boldsymbol{\beta}, \frac{\sigma^2}{n}\,\mathbb{E}\left[\boldsymbol{x}\boldsymbol{x'}\right]^{-1}\right) && (36)\\
&\xrightarrow{d}\mathcal{N}\left(\boldsymbol{\beta}, \frac{\sigma^2}{n}\boldsymbol{Q}^{-1}\right) = \mathcal{N}(\boldsymbol{\beta},\boldsymbol{0}) \quad \text{because } n\to\infty. && (37)
\end{aligned}$$

It is important to distinguish between limiting distribution and asymptotic distribution. As we see above, the limiting distribution exhibits zero variance, with which no inference can be made. Hence, we define the asymptotic distribution, where the variance does not collapse, as follows.

Asymptotic distribution.

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \overset{A}{\sim} \mathcal{N}(\mathbf{0}, \sigma^2\, \mathbb{E}\left[\boldsymbol{x}\boldsymbol{x'}^{-1}\right]). \tag{38}$$

Finite distribution.

Here we are restricted to the conditional distribution $\hat{\boldsymbol{\beta}}^{OLS} \mid \boldsymbol{X}$. As $\boldsymbol{u} \sim \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{I})$ is assumed, from $\hat{\boldsymbol{\beta}}^{OLS} = \boldsymbol{\beta} + (\boldsymbol{X'X})^{-1}\boldsymbol{X'u}$ we have

$$\hat{\boldsymbol{\beta}}^{OLS} \mid \boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\beta}, (\boldsymbol{X'X})^{-1}\boldsymbol{X'}\sigma^2\boldsymbol{I}\boldsymbol{X}(\boldsymbol{X'X})^{-1}) \tag{39}$$

$$\sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X'X})^{-1}). \tag{40}$$

2. Week 1, lecture 2 (Sep 21)

Topic: maximum likelihood and least squares estimators

1. Maximum likelihood estimation of multivariate normal distribution parameters under homoskedasticity.

   Given $\boldsymbol{x} \in \mathcal{M}^{M\times 1} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}$ and i.i.d. sample $\{\boldsymbol{x_i}\}_{i=1}^n$, we have

$$f(\boldsymbol{x_i}, \boldsymbol{\mu}, \sigma^2) = (2\pi)^{-\frac{M}{2}}|\sigma^2\boldsymbol{I}|^{-\frac{1}{2}}e^{-\frac{1}{2}(\boldsymbol{x_i}-\boldsymbol{\mu})'(\sigma^2\boldsymbol{I})^{-1}(\boldsymbol{x_i}-\boldsymbol{\mu})} := \mathscr{L}(\boldsymbol{x_i}, \boldsymbol{\mu}, \sigma^2) \tag{41}$$

$$\log \mathscr{L}(\boldsymbol{x_i}, \boldsymbol{\mu}, \sigma^2) = -\frac{M}{2}\log(2\pi) - \frac{1}{2}\log(|\sigma^2\boldsymbol{I}|) - \frac{1}{2}(\boldsymbol{x_i}-\boldsymbol{\mu})'(\sigma^2\boldsymbol{I})^{-1}(\boldsymbol{x_i}-\boldsymbol{\mu}) \tag{42}$$

$$= -\frac{M}{2}\log(2\pi) - \frac{M}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(\boldsymbol{x_i}-\boldsymbol{\mu})'(\boldsymbol{x_i}-\boldsymbol{\mu}) \overset{\text{i.i.id.}}{\Longrightarrow} \tag{43}$$

$$\log \mathscr{L}(\{\boldsymbol{x_i}\}_{i=1}^n, \boldsymbol{\mu}, \sigma^2) = \sum_{i=1}^n \log \mathscr{L}(\boldsymbol{x_i}, \boldsymbol{\mu}, \sigma^2) \tag{44}$$

$$= -\frac{nM}{2}\log(2\pi) - \frac{nM}{2}\log(\sigma^2) - \frac{1}{\sigma^2}\sum_{i=1}^n(\boldsymbol{x_i}-\boldsymbol{\mu})'(\boldsymbol{x_i}-\boldsymbol{\mu}) \tag{45}$$

$$= -\frac{nM}{2}\log(2\pi) - \frac{nM}{2}\log(\sigma^2) - \frac{1}{\sigma^2}\sum_{i=1}^n(\boldsymbol{x_i'x_i} - \boldsymbol{x_i'\mu} - \boldsymbol{\mu'x_i} + \boldsymbol{\mu'\mu}) \tag{46}$$

$$:= \ell(.), \tag{47}$$

then the FOCs and the estimates

$$\frac{\partial \ell(.)}{\partial \boldsymbol{\mu}} = \mathbf{0} \tag{48}$$

$$-\frac{1}{2\sigma^2}\sum_{i=1}^{n} = (-2\boldsymbol{x_i} + 2\boldsymbol{\mu}) = \mathbf{0} \implies \hat{\boldsymbol{\mu}}^{ML} = \sum_i \boldsymbol{x_i}/n = \bar{\boldsymbol{x}} \tag{49}$$

$$\frac{\partial \ell(.)}{\partial (\sigma^2)} = 0 \tag{50}$$

$$-\frac{nM}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_i (\boldsymbol{x_i} - \boldsymbol{\mu})'(\boldsymbol{x_i} - \boldsymbol{\mu}) = 0 \implies \widehat{\sigma^2}^{ML} = \sum_i (\boldsymbol{x_i} - \bar{\boldsymbol{x}})'(\boldsymbol{x_i} - \bar{\boldsymbol{x}})/(nM), \tag{51}$$

where we substituted in $\hat{\boldsymbol{\mu}}^{ML}$ while solving for $\widehat{\sigma^2}^{ML}$. Hence the Hessian,

$$\boldsymbol{H} = \frac{\partial^2 \ell(.)}{\partial \begin{bmatrix} \boldsymbol{\mu}' & \sigma^2 \end{bmatrix}' \partial \begin{bmatrix} \boldsymbol{\mu}' & \sigma^2 \end{bmatrix}} = \begin{bmatrix} -\frac{n}{\sigma^2}\boldsymbol{I}_M & -\frac{1}{\sigma^4}\sum_i (\boldsymbol{x_i} - \boldsymbol{\mu}) \\ -\frac{1}{\sigma^4}\sum_i (\boldsymbol{x_i} - \boldsymbol{\mu}) & \frac{nM}{2\sigma^4} - \frac{\sum_i (\boldsymbol{x_i}-\boldsymbol{\mu})'(\boldsymbol{x_i}-\boldsymbol{\mu})}{6} \end{bmatrix} \tag{52}$$

$$\mathbb{E}\left[\boldsymbol{H}\right] = \begin{bmatrix} -\frac{n}{\sigma^2}\boldsymbol{I}_M & \mathbf{0}_M \\ \mathbf{0}_M' & \frac{nM}{2\sigma^4} - \frac{nM\sigma^2}{\sigma^6} \end{bmatrix} \tag{53}$$

$$= \begin{bmatrix} -\frac{n}{\sigma^2}\boldsymbol{I}_M & \mathbf{0}_M \\ \mathbf{0}_M' & -\frac{nM}{2\sigma^4} \end{bmatrix} \tag{54}$$

$$(\mathbb{E}\left[\boldsymbol{H}\right])^{-1} = \begin{bmatrix} -\frac{\sigma^2}{n}\boldsymbol{I}_M & \mathbf{0}_M \\ \mathbf{0}_M' & -\frac{2\sigma^4}{nM} \end{bmatrix} \tag{55}$$

$$\mathrm{Var}\left[\begin{bmatrix} \hat{\boldsymbol{\mu}}' & \widehat{\sigma^2}^{ML} \end{bmatrix}'\right] = -(\mathbb{E}\left[\boldsymbol{H}\right])^{-1} \tag{56}$$

$$= \begin{bmatrix} \frac{\sigma^2}{n}\boldsymbol{I}_M & \mathbf{0}_M \\ \mathbf{0}_M' & \frac{2\sigma^4}{nM} \end{bmatrix} \tag{57}$$

$$= -(\boldsymbol{I}(\Theta))^{-1}, \quad \text{the inverse information matrix} \tag{58}$$

$$= \text{Cramer-Rao lower bound (CRLB)} \tag{59}$$

2. Least squares estimation of univariate normal parameters. I.i.d. sample: $\{x_i\}_{i=1}^n$, $x \sim \mathcal{N}(\mu, \sigma^2)$.

First, μ.

$$\hat{\mu}^{LS} = \arg\min_{\mu} \sum_{i=1}^{n} (x_i - \mu)^2 \implies \text{FOC}: \tag{60}$$

$$= -2\sum_i (x_i - \mu) = 0 \implies \hat{\mu}^{LS} = \bar{x}. \tag{61}$$

Second, σ^2. Let $\hat{\mu}$ denote $\hat{\mu}^{LS}$ from now on. Then

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{n}(x_i-\hat{\mu})^2\right] &= \mathbb{E}\left[\sum_i (x_i^2 - 2x_i\hat{\mu} + \hat{\mu}^2)\right] && (62)\\
&= \mathbb{E}\left[\sum_i x_i^2 - 2\hat{\mu}\sum_i x_i + n\hat{\mu}^2\right] && (63)\\
&= \mathbb{E}\left[\sum_i x_i^2\right] - 2\,\mathbb{E}\left[\hat{\mu}\sum_i x_i\right] + n\,\mathbb{E}\left[\hat{\mu}^2\right] && (64)\\
&\overset{\text{iid}}{=} n\,\mathbb{E}\left[x_i^2\right] - 2\,\mathbb{E}\left[\frac{(\sum_i x_i)}{n}\left(\sum_i x_i\right)\right] + n\,\mathbb{E}\left[\frac{(\sum_i x_i)(\sum_i x_i)}{n^2}\right] && (65)\\
&= n\,\mathbb{E}\left[x_i^2\right] - \frac{1}{n}\,\mathbb{E}\left[\left(\sum_i x_i\right)\left(\sum_i x_i\right)\right] && (66)\\
&= n\,\mathbb{E}\left[x_i^2\right] - \frac{1}{n}\left(n\,\mathbb{E}\left[x_i^2\right] + n(n-1)\,\mathbb{E}\left[x_i x_j\right]\right) \quad i \neq j && (67)\\
&= n\,\mathbb{E}\left[x_i^2\right] - \frac{1}{n}\left(n\,\mathbb{E}\left[x_i^2\right] + n(n-1)\,\mathbb{E}\left[x_i\right]\mathbb{E}\left[x_j\right]\right) \quad \text{iid sample:} x_i \perp\!\!\!\perp x_j && (68)\\
&= n\,\mathbb{E}\left[x_i^2\right] - \frac{1}{n}\left(n\,\mathbb{E}\left[x_i^2\right] + n(n-1)\mu^2\right) && (69)\\
&= n\,\mathbb{E}\left[x_i^2\right] - \mathbb{E}\left[x_i^2\right] - (n-1)\mu^2 && (70)\\
&= (n-1)\,\mathbb{E}\left[x_i^2\right] - (n-1)\mu^2 && (71)\\
&= (n-1)(\mathbb{E}\left[x_i^2\right] - \mu^2) \quad (\operatorname{Var} W = \mathbb{E}\left[W^2\right] - \mathbb{E}\left[W\right]^2) && (72)\\
&= (n-1)\sigma^2 \implies && (73)\\
\widehat{\sigma^2}^{LS} &= \frac{\sum_{i=1}^{n}(x_i-\hat{\mu})^2}{n-1}. && (74)
\end{aligned}$$

3. Least squares estimation of the classical regression model $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}$, $\mathbb{E}\left[\boldsymbol{y} \mid \boldsymbol{X}\right] = \boldsymbol{X\beta}$ under homoskedasticity: $\boldsymbol{u} \sim \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{I})$.

   First, $\boldsymbol{\beta}$.

$$\hat{\boldsymbol{\beta}}^{LS} = \arg\min_{\boldsymbol{\beta}}(\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta}) \iff \tag{75}$$

$$\hat{\boldsymbol{\beta}}^{LS} = \arg\min_{\boldsymbol{\beta}}(\boldsymbol{y'y} - \boldsymbol{y'X\beta} - \boldsymbol{\beta'X'y} + \boldsymbol{\beta'X'X\beta}) \implies \text{FOC:} \tag{76}$$

$$-\boldsymbol{X'y} - \boldsymbol{X'y} + 2\boldsymbol{X'X\beta} = \mathbf{0}_K \implies \hat{\boldsymbol{\beta}}^{LS} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'y}. \tag{77}$$

Second, σ^2. Entirely based on Greene's treatment (Econometric Analysis, p. 51 in 6th edition).

$$\begin{aligned}
\mathbb{E}\left[\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}} \mid \boldsymbol{X}\right] &= \mathbb{E}\left[\boldsymbol{u}'\boldsymbol{M}'\boldsymbol{M}\boldsymbol{u} \mid \boldsymbol{X}\right] \quad \boldsymbol{M} \text{ is symm. and idempotent } \Longrightarrow && (78)\\
&= \mathbb{E}\left[\boldsymbol{u}'\boldsymbol{M}\boldsymbol{u} \mid \boldsymbol{X}\right] && (79)\\
&= \mathbb{E}\left[\operatorname{tr}\left(\boldsymbol{u}'\boldsymbol{M}\boldsymbol{u}\right) \mid \boldsymbol{X}\right] \quad \text{trace of a scalar is the scalar itself} && (80)\\
&= \mathbb{E}\left[\operatorname{tr}\left(\boldsymbol{M}\boldsymbol{u}\boldsymbol{u}'\right) \mid X\right] \quad \text{trace property} && (81)\\
&= \operatorname{tr}\left(\mathbb{E}\left[\boldsymbol{M}\boldsymbol{u}\boldsymbol{u}' \mid \boldsymbol{X}\right]\right) \quad \text{linearity of expectation} && (82)\\
&= \operatorname{tr}\left(\boldsymbol{M}\,\mathbb{E}\left[\boldsymbol{u}\boldsymbol{u}' \mid \boldsymbol{X}\right]\right) && (83)\\
&= \operatorname{tr}\left(\boldsymbol{M}\sigma^2\boldsymbol{I}\right) && (84)\\
&= \sigma^2\operatorname{tr}(M). && (85)
\end{aligned}$$

Stop here for a minute and see where $\boldsymbol{M}$ came from. It is the so called residual-maker (as $\boldsymbol{M}\boldsymbol{y} = \hat{\boldsymbol{u}}$), is defined as $\boldsymbol{M} = \boldsymbol{I_n} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ and has the nice property:

$$\begin{aligned}
\hat{\boldsymbol{u}} &= \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} && (86)\\
&= \boldsymbol{y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{y} && (87)\\
&= (\boldsymbol{I_n} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{y} && (88)\\
&= (\boldsymbol{I_n} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}) && (89)\\
&= (\boldsymbol{I_n} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{u} && (90)\\
&= \boldsymbol{M}\boldsymbol{u}. && (91)
\end{aligned}$$

Therefore to pick up where we left

$$\begin{aligned}
\mathbb{E}\left[\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}} \mid \boldsymbol{X}\right] &= \sigma^2\operatorname{tr}(M) && (92)\\
&= \sigma^2(\operatorname{tr}\left((\boldsymbol{I_n} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\right)) && (93)\\
&= \sigma^2\left(\operatorname{tr}(I_n) - \operatorname{tr}\left(\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}\right)\right) && (94)\\
&= \sigma^2\left(n - \operatorname{tr}\left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\right)\right) \quad \text{trace property} && (95)\\
&= \sigma^2\left(n - \operatorname{tr}(\boldsymbol{I_K})\right) && (96)\\
&= \sigma^2(n-K) \implies \widehat{\sigma^2} = \frac{\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}}{n-K} && (97)
\end{aligned}$$

3. Week 2, lecture 3 (Sep 26)

Topic: hypothesis testing; endogeneity problems with OLS, introduced by lagged dependent variable, measurement error is x

3.1. Comments

3.2. Tasks

1. Consider the lagged dependent variable model

$$y_t = \boldsymbol{x_t'\beta} + \alpha y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma^2) \text{ (white noise)}, \tag{98}$$

and $(y_t, \boldsymbol{z_t})$, $\boldsymbol{z_t} = (\boldsymbol{x_t'}, y_{t-1})'$ exhibits joint weak stationarity (the covariance matrix of $(y_t, \boldsymbol{z_t})$ is time-indepdendent, and so are that of $\boldsymbol{z_t}$ and y_{t-1} each.), and $\boldsymbol{x_t}$ is stictly exogenous: $\mathbb{E}\left[\boldsymbol{x_s}\varepsilon_t\right] = \boldsymbol{0}\forall(t,s)$.

Show that OLS is biased yet consistent estimator of the above model.

Bias Let $\boldsymbol{y} = [y_1, \ldots, y_T]'$, $\boldsymbol{y_L} = [y_0, \ldots, y_{T-1}]'$, then

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{X} & \boldsymbol{y_L} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \alpha \end{bmatrix} + \boldsymbol{\varepsilon} \tag{99}$$

$$:= \boldsymbol{Z\delta} + \boldsymbol{\varepsilon}, \tag{100}$$

$$\hat{\boldsymbol{\delta}} = (\boldsymbol{Z'Z})^{-1}\boldsymbol{Z'y} = \boldsymbol{\delta} + (\boldsymbol{Z'Z})^{-1}\boldsymbol{Z'\varepsilon} \tag{101}$$

so OLS would be unbiased iff $\mathbb{E}\left[(\boldsymbol{Z'Z})^{-1}\boldsymbol{Z'\varepsilon}\right] = \mathbb{E}\left[(\boldsymbol{Z'Z})^{-1}\boldsymbol{Z'}\,\mathbb{E}\left[\boldsymbol{\varepsilon} \mid \boldsymbol{Z}\right]\right] = \boldsymbol{0}$, that is $\mathbb{E}\left[\varepsilon \mid \boldsymbol{Z}\right] = \boldsymbol{0}$, or equivalently, $\mathbb{E}\left[\varepsilon_t \mid \boldsymbol{z_s}\right] = 0\ \forall(t,s)$. Note that this condition is met by contemporaneous exogeneity in the cross-sectional i.i.d. sampling case, however this is not the case with time-series. Let's see in detail why. $\mathbb{E}\left[\varepsilon_t \mid \boldsymbol{z_s}\right] = 0 \iff \mathbb{E}\left[\mathbb{E}\left[\boldsymbol{z_s}\varepsilon_t \mid z_s\right]\right] = \boldsymbol{0} \iff \mathbb{E}\left[z_S\varepsilon_t\right] = 0$. As $\mathbf{cov}(\boldsymbol{z_s}, \epsilon_t) = \mathbb{E}\left[\boldsymbol{z_s}\varepsilon_t\right] - \mathbb{E}\left[\boldsymbol{z_s}\right]\mathbb{E}\left[\varepsilon_t\right]$, where the substracted term is zero ($\mathbb{E}\left[\varepsilon_t\right] = 0$), we can write $\mathbb{E}\left[\boldsymbol{z_s}\varepsilon_t\right] = \mathbf{cov}(\boldsymbol{z_s}, \varepsilon_t)$. Therefore, the required condition for exogeneity is $\mathbf{cov}(\boldsymbol{z_s}, \varepsilon_t) = \boldsymbol{0}\ \forall(t,s)$.

Set $s = t+1$. Now

$$\text{cov}(\boldsymbol{z_{t+1}}, \varepsilon_t) = \mathbf{cov}(\begin{bmatrix} \boldsymbol{x_{t+1}'} & y_t \end{bmatrix}', \varepsilon_t) \tag{102}$$

$$= \begin{bmatrix} \boldsymbol{0_K} \\ \text{cov}(y_t, \varepsilon_t) \end{bmatrix} \qquad (\mathbb{E}\left[\boldsymbol{x_s}\varepsilon_t\right] = \boldsymbol{0}\forall(t,s)) \tag{103}$$

$$= \begin{bmatrix} \boldsymbol{0_K} \\ \text{Var}(\varepsilon_t) \end{bmatrix} = \begin{bmatrix} \boldsymbol{0_K} \\ \sigma^2 \end{bmatrix} \neq \boldsymbol{0}_{K+1}, \tag{104}$$

hence our estimate, $\hat{\boldsymbol{\delta}}$ will be biased.

Consistency

We can check the consistency with probability limit, even though due to time-seres data, we need to make additional assumptions about data, which are not detailed here.

$$\operatorname{plim}\hat{\boldsymbol{\delta}} = \boldsymbol{\delta} + \operatorname{plim}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{\varepsilon} \tag{105}$$
$$= \boldsymbol{\delta} + \operatorname{plim}\left(\frac{\boldsymbol{Z}'\boldsymbol{Z}}{T}\right)^{-1} \operatorname{plim}\left(\frac{\boldsymbol{Z}'\boldsymbol{\varepsilon}}{T}\right) \tag{106}$$
$$= \boldsymbol{\delta} + \operatorname{plim}\left(\frac{\boldsymbol{Z}'\boldsymbol{Z}}{T}\right)^{-1} \operatorname{plim}\frac{\sum_{t=1}^{T}\boldsymbol{z_t}\varepsilon_t}{T} \tag{107}$$
$$= \boldsymbol{\delta} + \boldsymbol{Q}_{\boldsymbol{Z}}^{-1}\,\mathbb{E}\left[\boldsymbol{z_t}\varepsilon_t\right] \tag{108}$$
$$\mathbb{E}\left[\boldsymbol{z_t}\varepsilon_t\right] = \mathbf{cov}(\boldsymbol{z_t}, \varepsilon_t) \tag{109}$$
$$= \mathbf{0}_{K+1} \tag{110}$$

as ε_t is white noise, drawn independently from $\boldsymbol{z_t} = \begin{bmatrix}\boldsymbol{x}_t' & y_{t-1}\end{bmatrix}'$.. Hence, for well-behaving $\boldsymbol{Q_Z}$, $\operatorname{plim}\hat{\boldsymbol{\delta}} = \boldsymbol{\delta}$, OLS is consistent.

2. Again, consider the previous model with a modification: AR(1) distrbance process.

$$y_t = \boldsymbol{x_t'\beta} + \alpha y_{t-1} + \varepsilon_t, \qquad \varepsilon_t = \rho\varepsilon_{t-1} + v_t, \quad v_t \sim \mathcal{N}(0, \sigma_v^2) \text{ (white noise)}, \tag{111}$$

and $(y_t, \boldsymbol{z_t})$, $\boldsymbol{z_t} = (\boldsymbol{x}_t', y_{t-1})'$ exhibits joint weak stationarity (the covariance matrix of $(y_t, \boldsymbol{z_t})$ is time-indepdendent, and so are that of $\boldsymbol{z_t}$ and y_{t-1} each), and the disturbance process is also stationary: $|\rho| < 1$, and $\boldsymbol{x_t}$ is strictly exogenous: $\mathbb{E}\left[\boldsymbol{x_s}\varepsilon_t\right] = \mathbf{0}\forall(t, s)$

Bias The OLS estimator is the same as in the previous case, and is biased. Analogously, $\mathbf{cov}(\boldsymbol{z_s}, \varepsilon_t) = \mathbf{0}$ does not hold $\forall(t, s)$. For example, set $s = t$ which will also come in handy for the consistency check. Then in similar fashion

$$\mathbb{E}\left[\boldsymbol{z_t}\varepsilon_t\right] = \mathbf{cov}(\boldsymbol{z_t}, \varepsilon_t) \tag{112}$$
$$= \begin{bmatrix}\mathbf{0}_{\boldsymbol{K}} \\ \operatorname{cov}(y_{t-1}, \varepsilon_t)\end{bmatrix} \qquad \text{with} \tag{113}$$
$$\operatorname{cov}(y_{t-1}, \varepsilon_t) = \operatorname{cov}(\boldsymbol{x_t'} + \alpha y_{t-2}' + \varepsilon_{t-1}, \varepsilon_t) \tag{114}$$
$$= \rho\alpha\operatorname{cov}(y_{t-2}, \varepsilon_{t-1}) + \rho\operatorname{Var}(\varepsilon_t) \tag{115}$$

here we use our weak stationarity assumptions; $\operatorname{cov}(y_{t-1}, \varepsilon_t) = \operatorname{cov}(y_{t-2}, \varepsilon_{t-1})$ and obtain

$$\operatorname{cov}(y_{t-1}, \varepsilon_t) = \frac{\rho \operatorname{Var}(\varepsilon_t)}{1-\alpha\rho} \qquad \text{where} \tag{116}$$

$$\operatorname{Var}(\varepsilon_t) = \rho^2 \operatorname{Var}(\varepsilon_{t-1}) + \sigma_v^2 \qquad \text{again, weak stat., so} \tag{117}$$

$$\operatorname{cov}(y_{t-1}, \varepsilon_t) = \frac{\rho\sigma_v^2}{(1-\rho^2)(1-\alpha\rho)}, \tag{118}$$

which is not zero, hence we conclude that OLS is biased.

Consistency

In general, as we saw

$$\begin{aligned}
\operatorname{plim} \hat{\boldsymbol{\delta}} &= \boldsymbol{\delta} + \operatorname{plim}(\boldsymbol{Z'Z})^{-1}\boldsymbol{Z'\varepsilon} && (119)\\
&= \boldsymbol{\delta} + \operatorname{plim}\left(\frac{\boldsymbol{Z'Z}}{T}\right)^{-1} \operatorname{plim}\left(\frac{\boldsymbol{Z'\varepsilon}}{T}\right) && (120)\\
&= \boldsymbol{\delta} + \operatorname{plim}\left(\frac{\boldsymbol{Z'Z}}{T}\right)^{-1} \operatorname{plim} \frac{\sum_{t=1}^{T} \boldsymbol{z_t}\varepsilon_t}{T} && (121)\\
&= \boldsymbol{\delta} + \boldsymbol{Q_Z}^{-1}\,\mathbb{E}\left[\boldsymbol{z_t}\varepsilon_t\right] && (122)\\
\mathbb{E}\left[\boldsymbol{z_t}\varepsilon_t\right] &= \mathbf{cov}(\boldsymbol{z_t}, \varepsilon_t) && (123)\\
&= \begin{bmatrix} \mathbf{0}_{\boldsymbol{K}} \\ \frac{\rho\sigma_v^2}{(1-\rho^2)(1-\alpha\rho)} \end{bmatrix}, && (124)
\end{aligned}$$

which is equal to the zero vector only for $\rho = 0$, thus this time with AR(1) distubances, OLS is inconsistent.

To examine a more specific exampe, let's assume $\boldsymbol{z_t} = y_{t-1}$, so we only have the lagged dependent variable, $y_t = \alpha y_{t-1} + \varepsilon_t$. This makes for $\boldsymbol{Q_Z}^{-1} = \frac{1}{\operatorname{Var}(y_{t-1})} = \frac{1}{\operatorname{Var}(y_t)}$ due to weak stat. and that $\mathbb{E}\left[y_t\right] = 0$ as $\mathbb{E}\left[v_t\right] = \mathbb{E}\left[\varepsilon_t\right] = 0 \forall t$. Hence

$$\begin{aligned}
\operatorname{plim} \hat{\alpha} &= \alpha + \frac{\frac{\rho\sigma_v^2}{(1-\rho^2)(1-\alpha\rho)}}{\operatorname{Var} y_t} && (125)\\
&= \alpha + \frac{\rho\sigma_v^2}{(1-\rho^2)(1-\alpha\rho)\operatorname{Var} y_t}. && (126)
\end{aligned}$$

To get a clearer picture, examine $\mathrm{Var}(y_t)$.

$$\begin{aligned}\mathrm{Var}(y_t) &= \alpha^2 \mathrm{Var}(y_{t-1}) + \mathrm{Var}(\varepsilon_t) + 2\alpha \operatorname{cov}(y_{t-1}, \varepsilon_t) && (127)\\ &= \alpha^2 \mathrm{Var}(y_t) + \mathrm{Var}(\varepsilon_t) + 2\alpha \frac{\rho\sigma_v^2}{(1-\rho^2)(1-\alpha\rho)} && (128)\\ &= \alpha^2 \mathrm{Var}(y_t) + \frac{\sigma_v^2}{1-\rho^2} + 2\alpha \frac{\rho\sigma_v^2}{(1-\rho^2)(1-\alpha\rho)} \implies && (129)\\ \mathrm{Var}\, y &= \frac{(1+\alpha\rho)\sigma_v^2}{(1-\rho^2)(1-\alpha\rho)(1-\alpha^2)}. && (130)\end{aligned}$$

Finally,

$$\begin{aligned}\operatorname{plim} \hat{\alpha} &= \alpha + \frac{\rho\sigma_v^2}{(1-\alpha\rho)(1-\rho^2)\frac{(1+\alpha\rho)\sigma_v^2}{(1-\rho^2)(1-\alpha\rho)(1-\alpha^2)}} && (131)\\ &= \alpha + \frac{\rho(1-\alpha^2)}{1+\alpha\rho}. && (132)\end{aligned}$$

3.3. Concluding remarks

On white noise and AR(1) disturbance processes when lagged dependent variable is present on RHS.

In finite samples the observed RHS variables carry information about the (expected value) of disturbances, which however is not utillised by the OLS, leading to biased stimates.

As asymptotics kicks in (i) in the white noise case, the white noise nature of the disturbances becomes apparent and OLS is consistent (ii) in the AR(1) case the estimates still cannot converge in probability to their true value due to the repercussive effects: OLS is inconsistent.

4. Week 2, lecture 4 (Sep 28)

1. Present the asymptotic properties of the IV estimator. We see the general case when the number of instruments is larger then the number of endogenous variables ($K^* > K$), and see how the results simplify for our $K^* = K$.

The estimator with $\boldsymbol{P_Z} = \boldsymbol{Z}(\boldsymbol{Z'Z})^{-1}\boldsymbol{Z'}$ is computed as

$$\begin{aligned}
\hat{\boldsymbol{\beta}}^{IV} &= \arg\min_{\boldsymbol{\beta}} (\boldsymbol{y} - \boldsymbol{X\beta})'\boldsymbol{P_Z}(\boldsymbol{y} - \boldsymbol{X\beta}) && (133)\\
&= \arg\min_{\boldsymbol{\beta}} \boldsymbol{y'P_Zy} - \boldsymbol{y'P_ZX\beta} - \boldsymbol{\beta X'P_Zy} + \boldsymbol{\beta X'P_ZX\beta} && (134)\\
\text{FOC:} \quad & -\boldsymbol{X'P_Zy} - \boldsymbol{X'P_Zy} + 2\boldsymbol{X'P_ZX\beta} = \boldsymbol{0_K} \implies && (135)\\
\hat{\boldsymbol{\beta}}^{IV} &= (\boldsymbol{X'P_ZX})^{-1}\boldsymbol{X'P_Zy} && (136)\\
&= \boldsymbol{\beta} + (\boldsymbol{X'P_ZX})^{-1}\boldsymbol{X'P_Z\varepsilon} && (137)\\
&= \boldsymbol{\beta} + (\boldsymbol{X'Z}(\boldsymbol{Z'Z})^{-1}\boldsymbol{Z'X})^{-1}\boldsymbol{X'Z}(\boldsymbol{Z'Z})^{-1}\boldsymbol{Z'\varepsilon} && (138)\\
&= \boldsymbol{\beta} + \left(\frac{\boldsymbol{X'Z}}{n}\left(\frac{\boldsymbol{Z'Z}}{n}\right)^{-1}\frac{\boldsymbol{Z'X}}{n}\right)^{-1}\frac{\boldsymbol{X'Z}}{n}\left(\frac{\boldsymbol{Z'Z}}{n}\right)^{-1}\frac{\sum_{i=1}^{n}\boldsymbol{z_i}\varepsilon_i}{n} && (139)
\end{aligned}$$

so by assuming $\text{plim}(1/n)\boldsymbol{Z'X} = \boldsymbol{Q_{ZX}} = (\text{plim}(1/n)\boldsymbol{X'Z})' = \boldsymbol{Q'_{XZ}}$ non-zero and $\text{plim}(1/n)\boldsymbol{Z'Z} = \boldsymbol{Q_Z}$, we see consistency emerge

$$\text{plim}\,\hat{\boldsymbol{\beta}}^{IV} = \boldsymbol{\beta} + \text{plim}\left[\left(\frac{\boldsymbol{X'Z}}{n}\left(\frac{\boldsymbol{Z'Z}}{n}\right)^{-1}\frac{\boldsymbol{Z'X}}{n}\right)^{-1}\frac{\boldsymbol{X'Z}}{n}\left(\frac{\boldsymbol{Z'Z}}{n}\right)^{-1}\frac{\sum_{i=1}^{n}\boldsymbol{z_i}\varepsilon_i}{n}\right] \tag{140}$$

$$\begin{aligned}
&= \boldsymbol{\beta} + (\boldsymbol{Q_{XZ}Q_Z^{-1}Q_{ZX}})^{-1}\boldsymbol{Q_{XZ}Q_Z^{-1}0} && (141)\\
&= \boldsymbol{\beta}. && (142)
\end{aligned}$$

As for the limiting distribution, due to exogeneity of $\boldsymbol{Z}$, by CLT $\frac{\sum_{i=1}^{n}\boldsymbol{z_i}\varepsilon_i}{n} \xrightarrow{d} \mathcal{N}\left(\boldsymbol{0}, \frac{\text{Var}[\boldsymbol{z}\varepsilon]}{n}\right) = \mathcal{N}\left(\boldsymbol{0}, \frac{\sigma^2}{n}\boldsymbol{Q_Z}\right)$, and plim and convergence in distribution properties

$$\begin{aligned}
\hat{\boldsymbol{\beta}}^{IV} &\xrightarrow{d} \boldsymbol{\beta} + (\boldsymbol{Q_{XZ}Q_Z^{-1}Q_{ZX}})\boldsymbol{Q_{XZ}Q_Z^{-1}}\mathcal{N}\left(\boldsymbol{0}, \frac{\sigma^2}{n}\boldsymbol{Q_Z}\right) && (143)\\
&= \mathcal{N}\left(\boldsymbol{\beta}, (\boldsymbol{Q_{XZ}Q_Z^{-1}Q_{ZX}})^{-1}\boldsymbol{Q_{XZ}Q_Z^{-1}}\frac{\sigma^2}{n}\boldsymbol{Q_Z}\left((\boldsymbol{Q_{XZ}Q_Z^{-1}Q_{ZX}})^{-1}\boldsymbol{Q_{XZ}Q_Z^{-1}}\right)'\right) && (144)\\
&= \mathcal{N}\left(\boldsymbol{\beta}, \frac{\sigma^2}{n}\left(\boldsymbol{Q_{XZ}Q_Z^{-1}Q_{ZX}}\right)^{-1}\right) && (145)\\
&= \mathcal{N}(\boldsymbol{\beta}, \boldsymbol{0}) \qquad \text{as} \to \infty. && (146)
\end{aligned}$$

Asymptotic distribution, suitable for inference:

$$\sqrt{n}(\hat{\boldsymbol{\beta}}^{IV} - \boldsymbol{\beta}) \overset{A}{\sim} \mathcal{N}(\boldsymbol{0}, \sigma^2\left(\boldsymbol{Q_{XZ}Q_Z^{-1}Q_{ZX}}\right)^{-1}). \tag{147}$$

So for $K^* = K$, $\left(\boldsymbol{Q_{XZ}}\boldsymbol{Q_Z}^{-1}\boldsymbol{Q_{ZX}}\right)^{-1} = \boldsymbol{Q}_{\boldsymbol{ZX}}^{-1}\boldsymbol{Q_Z}\boldsymbol{Q}_{\boldsymbol{XZ}}^{-1}$.