Intro to Econometrics: Recitation 6

Hansen chapters 1-5

Gustavo Pereira

October 23, 2019

Roadmap

Hansen chapters 2-5 overview

- Chapter 2
  - Projection
  - Conditional expectation
  - Best linear predictor and linear regressions

- Take Y scalar rv and $\mathbf{X} = (X_1, \ldots, X_n)$. Consider following spaces:

$$\mathcal{L}(\mathbf{X}) = \{\hat{Y} : \hat{Y} = \sum_{i=1}^{n} \beta_i X_i\}$$

$$\mathcal{E}(\mathbf{X}) = \{\hat{Y} : \hat{Y} = f(\mathbf{X})\}$$

- Take Y scalar rv and $\mathbf{X} = (X_1, \ldots, X_n)$. Consider following spaces:

$$\mathcal{L}(\mathbf{X}) = \{\hat{Y} : \hat{Y} = \sum_{i=1}^{n} \beta_i X_i\}$$

$$\mathcal{E}(\mathbf{X}) = \{\hat{Y} : \hat{Y} = f(\mathbf{X})\}$$

- Let's restrict our analysis to variables Y, X such that $E[|Y|^2] < \infty$ and $E[\|\mathbf{X}\|^2] < \infty$. Moreover, assume $\mathbf{E}[\mathbf{X}\mathbf{X}'] > 0$

- Take Y scalar rv and $\mathbf{X} = (X_1, \ldots, X_n)$. Consider following spaces:

$$\mathcal{L}(\mathbf{X}) = \{\hat{Y} : \hat{Y} = \sum_{i=1}^{n} \beta_i X_i\}$$

$$\mathcal{E}(\mathbf{X}) = \{\hat{Y} : \hat{Y} = f(\mathbf{X})\}$$

- Let's restrict our analysis to variables Y, X such that $E[|Y|^2] < \infty$ and $E[\|\mathbf{X}\|^2] < \infty$. Moreover, assume $\mathbf{E}[\mathbf{X}\mathbf{X}'] > 0$
- That way the inner product $\langle X, Y \rangle := \mathbf{E}(YX)$ is well defined

- With that inner product: what's projection of Y (scalar valued) on $\mathcal{L}(\mathbf{X})$?

- With that inner product: what's projection of Y (scalar valued) on $\mathcal{L}(\mathbf{X})$?
- Orthogonality condition: $\langle Y - \mathbf{X}'\beta^*, \mathbf{X}'\beta \rangle = 0$ for all β

- With that inner product: what's projection of Y (scalar valued) on $\mathcal{L}(\mathbf{X})$?
- Orthogonality condition: $\langle Y - \mathbf{X}'\beta^*, \mathbf{X}'\beta \rangle = 0$ for all β
- Implies:
  1. $\beta^* = \mathbf{E}[\mathbf{X}\mathbf{X}']^{-1}\mathbf{E}[\mathbf{X}'Y]$
  2. Error term associated with projection is uncorrelated with $\mathbf{X}$. Let $u^* = Y - \mathbf{X}'\beta^*$

$$\langle u^*, \mathbf{X} \rangle = \mathbf{E}[u^*\mathbf{X}] = 0$$

- How about the projection of Y on $\mathcal{E}(\mathbf{X}) = \{\hat{Y} : \hat{Y} = f(\mathbf{X})\}$

- How about the projection of Y on $\mathcal{E}(\mathbf{X}) = \{\hat{Y} : \hat{Y} = f(\mathbf{X})\}$
- It's a *function* $f^*(\mathbf{X})$ such that for any function f,

$$\langle Y - f^*(\mathbf{X}), f(\mathbf{X}) \rangle = 0$$

- How about the projection of Y on $\mathcal{E}(\mathbf{X}) = \{\hat{Y} : \hat{Y} = f(\mathbf{X})\}$
- It's a *function* $f^*(\mathbf{X})$ such that for any function f,

$$\langle Y - f^*(\mathbf{X}), f(\mathbf{X})\rangle = 0$$

- The function

$$f^*(\mathbf{X}) = \mathbf{E}[Y|\mathbf{X}]$$

  satisfies the orthogonality conditions. (Check!)

- How about the projection of Y on $\mathcal{E}(\mathbf{X}) = \{\hat{Y} : \hat{Y} = f(\mathbf{X})\}$
- It's a *function* $f^*(\mathbf{X})$ such that for any function f,

$$\langle Y - f^*(\mathbf{X}), f(\mathbf{X}) \rangle = 0$$

- The function

$$f^*(\mathbf{X}) = \mathbf{E}[Y|\mathbf{X}]$$

  satisfies the orthogonality conditions. (Check!)
- Residual u^* satisfies exogeneity $\mathbf{E}[u^*|X] = 0$

- Take away:
  1. For any variables $(Y, \mathbf{X})$, you can *always* find β^* such that
  $$Y = \mathbf{X}\beta^* + u^*$$
  and $E[u^*\mathbf{X}] = 0$

- Take away:
  1. For any variables $(Y, \mathbf{X})$, you can *always* find β^* such that

     $$Y = \mathbf{X}\beta^* + u^*$$

     and $E[u^*\mathbf{X}] = 0$
  2. You can always write

     $$Y = f^*(\mathbf{X}) + u^*$$

     where $\mathbf{E}[u^*|X] = 0$

- Wait a second: *how about all the resources people spend trying to argue for 'exogeneity'*?

- Wait a second: *how about all the resources people spend trying to argue for 'exogeneity'*?
  - The point is *exactly* that in empirical applications, estimating

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

    will give you β^* in the limit

- Wait a second: *how about all the resources people spend trying to argue for 'exogeneity'*?
  - The point is *exactly* that in empirical applications, estimating
  $$y_i = x_i'\beta + u_i$$
  will give you β^* in the limit
  - Sometimes β^* is not the object of interest, but
  $$y_i = x_i'\tilde{\beta} + v_i$$
  where $\mathbf{E}[x_i v_i] \neq 0$

- Data $(y_i, \mathbf{x}_i)$, $i = 1, \ldots, n$ identically distributed from some joint distribution F

- Data $(y_i, \mathbf{x}_i)$, $i = 1, \ldots, n$ identically distributed from some joint distribution F
- Least squares problem:

$$\min_{\beta} \sum_{i=1}^{n} (y_i - \mathbf{x}_i' \beta)^2$$

- Data $(y_i, \mathbf{x}_i)$, $i = 1, \ldots, n$ identically distributed from some joint distribution F
- Least squares problem:

$$\min_{\beta} \sum_{i=1}^{n} (y_i - \mathbf{x}_i'\beta)^2$$

- In matrix notation:

$$\min_{\beta} \|\mathbf{y} - \mathbf{X}\beta\|^2$$

- Data $(y_i, \mathbf{x}_i)$, $i = 1, \ldots, n$ identically distributed from some joint distribution F
- Least squares problem:

$$\min_{\beta} \sum_{i=1}^{n} (y_i - \mathbf{x}_i'\beta)^2$$

- In matrix notation:

$$\min_{\beta} \|\mathbf{y} - \mathbf{X}\beta\|^2$$

- Solution: $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$

- Data $(y_i, \mathbf{x}_i)$, $i = 1, \ldots, n$ identically distributed from some joint distribution F
- Least squares problem:

$$\min_{\beta} \sum_{i=1}^{n} (y_i - \mathbf{x}_i' \beta)^2$$

- In matrix notation:

$$\min_{\beta} \|\mathbf{y} - \mathbf{X}\beta\|^2$$

- Solution: $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$
- Orthogonality condition: $\mathbf{X}'[\mathbf{y} - \mathbf{X}\hat{\beta}] = 0$

- Notation:

$$\hat{\mathbf{Q}}_{\mathbf{xx}} = \frac{1}{n}\mathbf{X}'\mathbf{X} = \sum_{i=1}^{n} \mathbf{x}_i x_i'$$

$$\hat{\mathbf{Q}}_{\mathbf{xy}} = \frac{1}{n}\mathbf{X}'\mathbf{y} = \sum_{i=1}^{n} \mathbf{x}_i y_i$$

$$\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

$$\mathbf{M} = \mathbf{I}_n - \mathbf{P} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

- Note:

$$
\begin{aligned}
\mathbf{y} &= \mathbf{X}\hat{\beta} + \overbrace{(\mathbf{y} - \mathbf{X}\hat{\beta})}^{\text{LS residuals}} \\
&= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} + \left[\mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{y}\right] \\
&= \mathbf{P}\mathbf{y} + \mathbf{M}\mathbf{y}
\end{aligned}
$$

- Hence $\mathbf{Py}$ is the predicted part and $\mathbf{My}$ is the residual
- Matrices $\mathbf{P}$ and $\mathbf{M}$ are both *symmetric*, and satisfiy:

$$
\begin{aligned}
\mathbf{PP} &= \mathbf{P} \\
\mathbf{MM} &= \mathbf{M} \\
\mathbf{PM} &= \mathbf{MP} = \mathbf{0} \\
\mathbf{PX} &= \mathbf{X} \\
\mathbf{MX} &= \mathbf{0}
\end{aligned}
$$

- Let's apply this machinery. Two components:

$$y_i = \mathbf{x}_{1i}'\beta_1 + \mathbf{x}_{2i}'\beta_2 + u_i$$

- Let's apply this machinery. Two components:

$$y_i = \mathbf{x}_{1i}'\beta_1 + \mathbf{x}_{2i}'\beta_2 + u_i$$

- In matrix notation:

$$\mathbf{y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \mathbf{u}$$

- Let's apply this machinery. Two components:

$$y_i = \mathbf{x}_{1i}'\beta_1 + \mathbf{x}_{2i}'\beta_2 + u_i$$

- In matrix notation:

$$\mathbf{y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \mathbf{u}$$

- At the least squares solution,

$$\mathbf{y} = \mathbf{X}_1\hat{\beta}_1 + \mathbf{X}_2\hat{\beta}_2 + \mathbf{e}$$

where

$$\mathbf{0} = \mathbf{X}'\mathbf{e} = \begin{bmatrix} \mathbf{X}_1' \\ \mathbf{X}_2' \end{bmatrix} \mathbf{e}$$

- Define $\mathbf{P}_j, \mathbf{M}_j$ for $j = 1, 2$ accordingly
- Suppose we want to find expression for $\hat{\beta}_1$. Can get rid of $\mathbf{X}_2$ by multiplying $\mathbf{M}_2$!

$$\mathbf{M}_2\mathbf{y} = \mathbf{M}_2\mathbf{X}_1\hat{\beta}_1 + \mathbf{M}_2\mathbf{X}_2\hat{\beta}_2 + \mathbf{M}_2\mathbf{e}$$

- Note: $\mathbf{M}_2\mathbf{e} = \mathbf{e} - \mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{e} = \mathbf{e}$
- Hence

$$\mathbf{M}_2\mathbf{y} = \mathbf{M}_2\mathbf{X}_1\hat{\beta}_1 + \mathbf{e}$$

- Denote $\tilde{\mathbf{y}} = \mathbf{M}_2\mathbf{y}$ and $\tilde{\mathbf{X}}_1 = \mathbf{M}_2\mathbf{X}_1$

- Denote $\tilde{\mathbf{y}} = \mathbf{M}_2\mathbf{y}$ and $\tilde{\mathbf{X}}_1 = \mathbf{M}_2\mathbf{X}_1$
- We have

$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}_1\hat{\beta}_1 + \mathbf{e}$$

Moreover,

$$\tilde{\mathbf{X}}_1'\mathbf{e} = \mathbf{X}_1'\mathbf{M}_2\mathbf{e} = \mathbf{0}$$

Thus (as long as $\tilde{X}_2$ is full row rank):

$$\hat{\beta}_1 = (\tilde{\mathbf{X}}_1'\tilde{\mathbf{X}}_1)^{-1}\tilde{\mathbf{X}}_1'\mathbf{y} = (\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{M}_2\mathbf{y}$$

- Denote $\tilde{\mathbf{y}} = \mathbf{M}_2\mathbf{y}$ and $\tilde{\mathbf{X}}_1 = \mathbf{M}_2\mathbf{X}_1$
- We have
$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}_1\hat{\beta}_1 + \mathbf{e}$$
Moreover,
$$\tilde{\mathbf{X}}_1'\mathbf{e} = \mathbf{X}_1'\mathbf{M}_2\mathbf{e} = \mathbf{0}$$
Thus (as long as $\tilde{X}_2$ is full row rank):
$$\hat{\beta}_1 = (\tilde{\mathbf{X}}_1'\tilde{\mathbf{X}}_1)^{-1}\tilde{\mathbf{X}}_1'\mathbf{y} = (\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{M}_2\mathbf{y}$$
- Interpretation? Frisch-Waugh-Lovell

- Data $(y_i, \mathbf{x}_i)$ independently drawn from $F(y, \mathbf{x})$

- Data $(y_i, \mathbf{x}_i)$ independently drawn from $F(y, \mathbf{x})$
- Statistical model will put further restrictions on F.

- Data $(y_i, \mathbf{x}_i)$ independently drawn from $F(y, \mathbf{x})$
- Statistical model will put further restrictions on F.
- Note: not assuming deterministic $\mathbf{x}_i$ anymore. Analysis will strongly rely on conditioning

- Depending on what properties you want to get for OLS, different assumptions are required

- Depending on what properties you want to get for OLS, different assumptions are required
  1. Linear regression model.
     - $\mathbf{E}[y_i|\mathbf{x}_i] = \mathbf{x}_i'\beta$
     - Finite second moments, and $\mathbf{E}[\mathbf{x}_i\mathbf{x}_i']$ invertible

- Depending on what properties you want to get for OLS, different assumptions are required
  1. Linear regression model.
     - $\mathbf{E}[y_i|\mathbf{x}_i] = \mathbf{x}_i'\beta$
     - Finite second moments, and $\mathbf{E}[\mathbf{x}_i\mathbf{x}_i']$ invertible
- With the above assumption, we get an unbiased OLS estimator.

- Depending on what properties you want to get for OLS, different assumptions are required
    1. Linear regression model.
        - $\mathbf{E}[y_i|\mathbf{x}_i] = \mathbf{x}_i'\beta$
        - Finite second moments, and $\mathbf{E}[\mathbf{x}_i\mathbf{x}_i']$ invertible
- With the above assumption, we get an unbiased OLS estimator.
- What about 'optimality' in any sense? Need restriction on second moments.

- Another assumption:

  2. Homoskedasticity. In addition to linear regression hypohtesis,

$$\mathbf{V}[y_i|\mathbf{x}_i] \equiv \sigma^2$$

- Then we get the Gauss-Markov result.

Theorem (Gauss-Markov)

In the homoskedastic linear regression model, $\hat{\beta}$ is the best linear unbiased estimator of β (with L^2 loss).
That means that any other unbiased estimator $\tilde{\beta} = \tilde{A}\mathbf{y}$ satisfies

$$\mathbf{V}[\tilde{\beta}|\mathbf{X}] \geq \mathbf{V}[\hat{\beta}|\mathbf{X}]$$

- When homoskedasticity is not assumed, we can sometimes do better than OLS
- For example, if we abandon the iid assumption, and instead only impose finite second moments and

$$\mathbf{E}[\mathbf{y}|\mathbf{X}] = \mathbf{X}\beta$$

$$\mathbf{V}[\mathbf{y}|\mathbf{X}] = \Omega$$

- If Ω is known, the *Generalized Least Squares estimator* is the way to go.

- Another important case that we frequently find in applied work:

$$\mathbf{V}[y_i|\mathbf{x}_i] = \varsigma(x_i)^2 = \sigma_i^2$$

- With the above form for residual variace, we have a *heteroskedastic linear regression model*

Chapter 4

Least squares: variance estimation

- TBI