Intro to Econometrics: Recitation 8

Gustavo Pereira

November 11, 2019

Roadmap for today

- Chapter 6 quick review:
  - Asymptotic order notation
  - The basic toolkit: CMT, delta method and limit theorems
- BH Chapter 7:
  - Asymptotic properties in the LP/LR model

Asymptotic order notation

Deterministic sequences

- Consider the claim:

  "The sequence $a_n = n^2$ diverges to infinity an order of magnitude faster than $b_n = n$"

  - How do we make this rigorous?

Asymptotic order notation

Deterministic sequences

- Consider the claim:

  "The sequence $a_n = n^2$ diverges to infinity an order of magnitude faster than $b_n = n$"

  - How do we make this rigorous?

- "$|a_n - b_n|$ gets large" is not a good idea; e.g. $a_n = 5n$, $b_n = n$

Asymptotic order notation

Deterministic sequences

- Consider the claim:

  "The sequence $a_n = n^2$ diverges to infinity an order of magnitude faster than $b_n = n$"

  - How do we make this rigorous?
- "$|a_n - b_n|$ gets large" is not a good idea; e.g. $a_n = 5n$, $b_n = n$
- Let $r_n = \frac{a_n}{b_n}$:

Asymptotic order notation

Deterministic sequences

- Consider the claim:

  "The sequence $a_n = n^2$ diverges to infinity an order of magnitude faster than $b_n = n$"

  - How do we make this rigorous?
- "$|a_n - b_n|$ gets large" is not a good idea; e.g. $a_n = 5n$, $b_n = n$
- Let $r_n = \frac{a_n}{b_n}$:
  - $a_n \in O(b_n) \iff |r_n|$ is bounded

Asymptotic order notation

Deterministic sequences

- Consider the claim:

  "The sequence $a_n = n^2$ diverges to infinity an order of magnitude faster than $b_n = n$"

  - How do we make this rigorous?

- "$|a_n - b_n|$ gets large" is not a good idea; e.g. $a_n = 5n$, $b_n = n$
- Let $r_n = \frac{a_n}{b_n}$:
  - $a_n \in O(b_n) \iff |r_n|$ is bounded
  - $a_n \in o(b_n) \iff |r_n| \to 0$

Asymptotic order notation

Deterministic sequences

- Consider the claim:

  "The sequence $a_n = n^2$ diverges to infinity an order of magnitude faster than $b_n = n$"

  - How do we make this rigorous?

- "$|a_n - b_n|$ gets large" is not a good idea; e.g. $a_n = 5n$, $b_n = n$
- Let $r_n = \frac{a_n}{b_n}$:
  - $a_n \in O(b_n) \iff |r_n|$ is bounded
  - $a_n \in o(b_n) \iff |r_n| \to 0$
- Note: we're assuming for simplicity that $b_n > 0$

Asymptotic order notation

Deterministic sequences: examples

- Convince yourself (or even better, give a proof) that:
  1. $n \in o(n^2)$
  2. $log(n) \in O(n)$
  3. $sin(n) \in O(1)$

Asymptotic order notation

Deterministic sequences: examples

- Convince yourself (or even better, give a proof) that:
  1. $n \in o(n^2)$
  2. $log(n) \in O(n)$
  3. $sin(n) \in O(1)$
- Also, for arbitrary (positive) x_n and y_n,

Asymptotic order notation

Deterministic sequences: examples

- Convince yourself (or even better, give a proof) that:
  1. $n \in o(n^2)$
  2. $log(n) \in O(n)$
  3. $sin(n) \in O(1)$
- Also, for arbitrary (positive) x_n and y_n,
  1. $x_n \in O(x_n)$ but the same doesn't hold for $o(x_n)$

Asymptotic order notation

Deterministic sequences: examples

- Convince yourself (or even better, give a proof) that:
  1. $n \in o(n^2)$
  2. $log(n) \in O(n)$
  3. $sin(n) \in O(1)$
- Also, for arbitrary (positive) x_n and y_n,
  1. $x_n \in O(x_n)$ but the same doesn't hold for $o(x_n)$
  2. If $x_n \in O(An^2 + Bn + C)$, then $x_n \in O(n^2)$; moreover, $x_n \in o(n^3)$
     - Similar fact for higher degree polynomials goes through

Asymptotic order notation

Deterministic sequences: examples

- Convince yourself (or even better, give a proof) that:
  1. $n \in o(n^2)$
  2. $log(n) \in O(n)$
  3. $sin(n) \in O(1)$
- Also, for arbitrary (positive) x_n and y_n,
  1. $x_n \in O(x_n)$ but the same doesn't hold for $o(x_n)$
  2. If $x_n \in O(An^2 + Bn + C)$, then $x_n \in O(n^2)$; moreover, $x_n \in o(n^3)$
     - ★ Similar fact for higher degree polynomials goes through
  3. $x_n \in O(y_n)$ iff $x_n/y_n \in o(1)$; same for O

Asymptotic order notation

Deterministic sequences: examples

- Convince yourself (or even better, give a proof) that:
  1. $n \in o(n^2)$
  2. $log(n) \in O(n)$
  3. $sin(n) \in O(1)$
- Also, for arbitrary (positive) x_n and y_n,
  1. $x_n \in O(x_n)$ but the same doesn't hold for $o(x_n)$
  2. If $x_n \in O(An^2 + Bn + C)$, then $x_n \in O(n^2)$; moreover, $x_n \in o(n^3)$
     - ★ Similar fact for higher degree polynomials goes through
  3. $x_n \in O(y_n)$ iff $x_n/y_n \in o(1)$; same for O
  4. If $x_n \to x$, then $x_n \in O(1)$

Asymptotic order notation

Deterministic sequences: results

Lemma

For any positive sequences x_n, y_n, z_n and w_n,

1. If $x_n \in o(z_n)$ and $y_n \in o(w_n)$, then

$$x_n y_n \in o(z_n w_n)$$

$$x_n + y_n \in o(z_n + w_n)$$

   and the same holds for O

2. If $x_n \in o(z_n)$ and $y_n \in O(w_n)$, then

$$x_n y_n \in o(z_n w_n)$$

$$x_n + y_n \in O(z_n + w_n)$$

Asymptotic order notation

Deterministic sequences: results

- Proof:
  - First part of (1) follows from limit arithmetic
  - First part of (2) is a consequence of the fact that $a_n \to 0$ and b_n bounded implies $a_n b_n \to 0$
  - Second part of (1) and (2): use

$$\frac{x_n + y_n}{z_n + w_n} = \frac{x_n}{z_n} \frac{z_n}{z_n + w_n} + \frac{y_n}{w_n} \frac{w_n}{z_n + w_n}$$

    the ratio $z_n/(z_n + w_n) \in [0, 1]$ is bounded

Asymptotic order notation

Sequences of random variables

- Analogs: let X_n be a sequence of random variables and a_n a positive deterministic sequence
  - $X_n \in O_p(a_n)$ iff $\dfrac{X_n}{a_n}$ is *stochastically* bounded
  - $X_n \in o_p(a_n)$ iff $\dfrac{X_n}{a_n} \xrightarrow{p} 0$

Asymptotic order notation

Sequences of random variables

- Analogs: let X_n be a sequence of random variables and a_n a positive deterministic sequence
  - $X_n \in O_p(a_n)$ iff $\dfrac{X_n}{a_n}$ is *stochastically* bounded
  - $X_n \in o_p(a_n)$ iff $\dfrac{X_n}{a_n} \xrightarrow{p} 0$
- Definition of stochastically bounded.

  For every $\delta > 0$ there exists M_δ such that

$$\Pr\{|X_n| \leq M_\delta\} > 1 - \delta$$

Asymptotic order notation

Sequences of random variables

- Analogs: let X_n be a sequence of random variables and a_n a positive deterministic sequence
  - $X_n \in O_p(a_n)$ iff $\dfrac{X_n}{a_n}$ is *stochastically* bounded
  - $X_n \in o_p(a_n)$ iff $\dfrac{X_n}{a_n} \xrightarrow{p} 0$
- Definition of stochastically bounded.

  For every $\delta > 0$ there exists M_δ such that

$$\Pr\{|X_n| \leq M_\delta\} > 1 - \delta$$

- Result: lemma goes through with o and O replaced with o_p and O_p

Asymptotic order notation

Sequences of random variables: an useful result

A sequence (X_n) that converges in distribution is necessarily $O_p(1)$

Asymptotic order notation

Sequences of random variables: an useful result

A sequence (X_n) that converges in distribution is necessarily $O_p(1)$

- Proof. Let Y be st $X_n \xrightarrow{d} Y$. Denote F_Y and F_n the CDF of Y and X_n, respectively. Fix $\delta > 0$.
  - Let M_δ be such that

$$F_Y(M_\delta) - F_Y(-M_\delta) = \Pr\{Y \in (-M_\delta, M_\delta]\} > 1 - \frac{\delta}{2}$$

  - In addition, we can choose M_δ st F_Y continuous at M_δ and $-M_\delta$ (why?)

Asymptotic order notation

Sequences of random variables: an useful result

A sequence (X_n) that converges in distribution is necessarily $O_p(1)$

- Proof. Let Y be st $X_n \xrightarrow{d} Y$. Denote F_Y and F_n the CDF of Y and X_n, respectively. Fix $\delta > 0$.
  - Let M_δ be such that

$$F_Y(M_\delta) - F_Y(-M_\delta) = \Pr\{Y \in (-M_\delta, M_\delta]\} > 1 - \frac{\delta}{2}$$

  - In addition, we can choose M_δ st F_Y continuous at M_δ and $-M_\delta$ (why?)
- Convergence in distribution implies $F_n(M_\delta) \to F_Y(M_\delta)$ and $F_n(-M_\delta) \to F_Y(-M_\delta)$

Asymptotic order notation

Sequences of random variables: an useful result

A sequence (X_n) that converges in distribution is necessarily $O_p(1)$

- Proof. Let Y be st $X_n \xrightarrow{d} Y$. Denote F_Y and F_n the CDF of Y and X_n, respectively. Fix $\delta > 0$.
  - Let M_δ be such that

$$F_Y(M_\delta) - F_Y(-M_\delta) = \Pr\{Y \in (-M_\delta, M_\delta]\} > 1 - \frac{\delta}{2}$$

  - In addition, we can choose M_δ st F_Y continuous at M_δ and $-M_\delta$ (why?)
- Convergence in distribution implies $F_n(M_\delta) \to F_Y(M_\delta)$ and $F_n(-M_\delta) \to F_Y(-M_\delta)$
- Hence for n_0 sufficiently large, $F_n(M_\delta) > F_Y(M_\delta) - \delta/4$ and $F_n(-M_\delta) < F_Y(-M_\delta) + \delta/4$ for all $n \geq n_0$

Asymptotic order notation

Sequences of random variables: an useful result

A sequence (X_n) that converges in distribution is necessarily $O_p(1)$

- Proof. Let Y be st $X_n \xrightarrow{d} Y$. Denote F_Y and F_n the CDF of Y and X_n, respectively. Fix $\delta > 0$.
  - Let M_δ be such that

$$F_Y(M_\delta) - F_Y(-M_\delta) = \Pr\{Y \in (-M_\delta, M_\delta]\} > 1 - \frac{\delta}{2}$$

  - In addition, we can choose M_δ st F_Y continuous at M_δ and $-M_\delta$ (why?)
- Convergence in distribution implies $F_n(M_\delta) \to F_Y(M_\delta)$ and $F_n(-M_\delta) \to F_Y(-M_\delta)$
- Hence for n_0 sufficiently large, $F_n(M_\delta) > F_Y(M_\delta) - \delta/4$ and $F_n(-M_\delta) < F_Y(-M_\delta) + \delta/4$ for all $n \geq n_0$
- That implies

$$\Pr\{|X_n| \leq M_\delta\} \geq F_n(M_\delta) - F_n(-M_\delta) > 1 - \delta$$

The asymptotics toolkit

- You need to know the following by heart:
  1. General results
     - Continuous mapping theorem
     - Slutzky's theorem
     - Delta method
  2. Results about averages
     - Law of large numbers
     - Central Limit Theorem

Continuous mapping theorem

- Let $F : U \to V$ have a set of discontinuities U_0
- If $X_n \xrightarrow{d} X$ or $X_n \xrightarrow{p} X$, then $F(X_n) \to F(X)$ as long as

$$\Pr\{X \in U_0\} = 0$$

Slutzky's theorem

- If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c$ where c is non-random, then
  1. $X_n + Y_n \xrightarrow{d} X + c$
  2. $X_n Y_n \xrightarrow{d} cX$
  3. $X_n / Y_n \xrightarrow{d} X/c$ as long as $c > 0$
- These results follow from CMT, plus the fact that

$$(X_n, Y_n) \xrightarrow{d} (X, c)$$

provided that c is non-random.

Delta method

- If μ is non-random,
$$\sqrt{n}(X_n - \mu) \xrightarrow{d} Z$$
and g is differentiable at μ, then
$$\sqrt{n}(g(X_n) - g(\mu)) \xrightarrow{d} g'(\mu)Z$$
- Remarks.
  1. No assumption about normality is needed for Z.
  2. The theorem goes through in the multivariate case
     - Example: $F(\gamma_1, \gamma_2) = \gamma_1\gamma_2$

Results about averages

- Let $\{X_i\}$ be sampled iid from some distribution. Denote

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Results about averages

- Let $\{X_i\}$ be sampled iid from some distribution. Denote

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- Law of large numbers. If $\{X_i\}$ are sampled iid from distribution with finite mean, then

$$\bar{X}_n \xrightarrow{p} \mathbf{E}X_1$$

Results about averages

- Let $\{X_i\}$ be sampled iid from some distribution. Denote

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

- Law of large numbers. If $\{X_i\}$ are sampled iid from distribution with finite mean, then

$$\bar{X}_n \xrightarrow{p} \mathbf{E} X_1$$

- Central limit theorem. If $\{X_i\}$ are sampled iid from distribution with finite mean μ and variance σ^2,

$$\sqrt{n} \left(\bar{X}_n - \mu \right) \xrightarrow{d} Z \sim N(0, \sigma^2)$$

Results about averages

- Remark. If X_i have finite mean and variance, the central limit theorem gives a bound on the rate of convergence of $\bar{X}_n$

Results about averages

- Remark. If X_i have finite mean and variance, the central limit theorem gives a bound on the rate of convergence of $\bar{X}_n$
- In fact:

$$\bar{X}_n - \mu = \frac{1}{\sqrt{n}}\left[\sqrt{n}(\bar{X}_n - \mu)\right] = O_p(n^{-1/2})O_p(1) = O_p(n^{-1/2})$$

Results about averages

- Remark. If X_i have finite mean and variance, the central limit theorem gives a bound on the rate of convergence of $\bar{X}_n$
- In fact:

$$\bar{X}_n - \mu = \frac{1}{\sqrt{n}}\left[\sqrt{n}(\bar{X}_n - \mu)\right] = O_p(n^{-1/2})O_p(1) = O_p(n^{-1/2})$$

- Hence (with finite second moments) the convergence of $\bar{X}_n$ to μ is no slower than $n^{-1/2}$ goes to 0.

OLS asymptotics: the linear projection model

- Consider the following assumptions:

  (1) (y_i, x_i) sampled iid

  (2) $\mathbf{E}(y_i^2) < \infty$

  (2') $\mathbf{E}(y_i^4) < \infty$

  (3) $\mathbf{E}(\|x_i\|^2) < \infty$

  (3') $\mathbf{E}(\|x_i\|^4) < \infty$

  (4) $\mathbf{E}[x_i x_i']$ positive definite

- Notation:

$$\mathbf{E}[x_i x_i'] =: Q_{xx}$$

$$\mathbf{E}[x_i y_i] =: Q_{xy}$$

OLS asymptotics: the linear projection model

- Results:
  1. Under (1), (2), (3) and (4), OLS is consistent for the population projection coefficient:
  $$\hat{\beta} = \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1} \sum_{i=1}^{n} x_i y_i \xrightarrow{p} \beta := Q_{xx}^{-1} Q_{xy}$$
  2. Under (1), (2'), (3') and (4), OLS is asymptotically normal:
  $$\sqrt{n}\left[\left(\sum_{i=1}^{n} x_i x_i'\right)^{-1} \sum_{i=1}^{n} x_i y_i - \beta\right] \xrightarrow{d} N\left(0, Q_{xx}^{-1} \Omega Q_{xx}^{-1}\right)$$
     - ⋆ β is again the population projection coefficient
     - ⋆ $\Omega = \mathbf{E}[u_i^2 x_i x_i']$
     - ⋆ $u_i := y_i - x_i'\beta$
- Notation:
$$V_\beta = Q_{xx}^{-1} \Omega Q_{xx}^{-1}$$

OLS asymptotics: the linear regression model

- Consider the following assumptions:

(1) (y_i, x_i) sampled iid

(2) $\mathbf{E}(y_i^2) < \infty$

(3) $\mathbf{E}(\|x_i\|^2) < \infty$

(4) $\mathbf{E}[x_i x_i']$ positive definite

(5) $\mathbf{E}(y_i|x_i) = x_i'\beta$

- Assumptions (1), (2), (3), (4) and (5) are called the linear regression model

OLS asymptotics: the linear regression model

- Consider the following assumptions:

(1) (y_i, x_i) sampled iid

(2') $\mathbf{E}(y_i^4) < \infty$

(3') $\mathbf{E}(\|x_i\|^4) < \infty$

(4) $\mathbf{E}[x_i x_i']$ positive definite

(5) $\mathbf{E}(y_i | x_i) = x_i' \beta$

- Assumptions (1), (2), (3), (4) and (5) are called the linear regression model
- Again (2') and (3') are required for asymptotic normality

OLS asymptotics: the linear regression model

- Consider the following assumptions:

(1) (y_i, x_i) sampled iid

(2) $\mathbf{E}(y_i^2) < \infty$

(2') $\mathbf{E}(y_i^4) < \infty$

(3) $\mathbf{E}(\|x_i\|^2) < \infty$

(3') $\mathbf{E}(\|x_i\|^4) < \infty$

(4) $\mathbf{E}[x_i x_i']$ positive definite

(5) $\mathbf{E}(y_i|x_i) = x_i'\beta$

(6) $\mathbf{V}[y_i|x_i] \equiv \sigma^2$

- Assumptions (1), (2), (3), (4) and (5) are called the linear regression model
- Again (2') and (3') are required for asymptotic normality
- If (6) is assumed, we have the homoskedastic linear regression model

OLS asymptotics: the linear regression model

- Finite sample variance of OLS:

$$\hat{\beta} = (X'X)^{-1}X'Y \implies \mathbf{V}(\hat{\beta}|X) = (X'X)^{-1}X'\mathbf{V}(Y|X)X(X'X)^{-1}$$

OLS asymptotics: the linear regression model

- Finite sample variance of OLS:

$$\hat{\beta} = (X'X)^{-1}X'Y \implies \mathbf{V}(\hat{\beta}|X) = (X'X)^{-1}X'\mathbf{V}(Y|X)X(X'X)^{-1}$$

- $\mathbf{V}(Y|X)$ is a diagonal matrix. Why?

OLS asymptotics: the linear regression model

- Finite sample variance of OLS:

$$\hat{\beta} = (X'X)^{-1}X'Y \implies \mathbf{V}(\hat{\beta}|X) = (X'X)^{-1}X'\mathbf{V}(Y|X)X(X'X)^{-1}$$

- $\mathbf{V}(Y|X)$ is a diagonal matrix. Why?
- Denote $D = \mathbf{V}(Y|X)$. Hansen defines:

$$V_{\hat{\beta}} = (X'X)^{-1}X'DX(X'X)^{-1}$$

OLS asymptotics: the linear regression model

- Finite sample variance of OLS:

$$\hat{\beta} = (X'X)^{-1}X'Y \implies \mathbf{V}(\hat{\beta}|X) = (X'X)^{-1}X'\mathbf{V}(Y|X)X(X'X)^{-1}$$

- $\mathbf{V}(Y|X)$ is a diagonal matrix. Why?
- Denote $D = \mathbf{V}(Y|X)$. Hansen defines:

$$V_{\hat{\beta}} = (X'X)^{-1}X'DX(X'X)^{-1}$$

- This is not the same thing as

$$V_{\beta} = Q_{xx}^{-1}\mathbf{E}[u_i^2 x_i x_i']Q_{xx}^{-1}$$

OLS asymptotics: the linear regression model

- Finite sample variance of OLS:

$$\hat{\beta} = (X'X)^{-1}X'Y \implies \mathbf{V}(\hat{\beta}|X) = (X'X)^{-1}X'\mathbf{V}(Y|X)X(X'X)^{-1}$$

- $\mathbf{V}(Y|X)$ is a diagonal matrix. Why?
- Denote $D = \mathbf{V}(Y|X)$. Hansen defines:

$$V_{\hat{\beta}} = (X'X)^{-1}X'DX(X'X)^{-1}$$

- This is not the same thing as

$$V_{\beta} = Q_{xx}^{-1}\mathbf{E}[u_i^2 x_i x_i']Q_{xx}^{-1}$$

- It is true however that

$$nV_{\hat{\beta}} \xrightarrow{p} V_{\beta}$$

(why?)

OLS asymptotics: estimating the limiting variance

- We know that the limiting variance of OLS in the LPM is

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

- A valid estimator for $\Omega = \mathbf{E}[u_i^2 x_i x_i']$ is

$$\hat{\Omega} = \frac{1}{n} \sum_{i=1}^{n} \hat{u}_i^2 x_i x_i'$$

- One estimator for V_β is

$$\hat{V}_\beta^{\text{HC0}} = \hat{Q}_{xx}^{-1} \hat{\Omega} \hat{Q}_{xx}^{-1}$$

Prediction: regression intervals

- One object of interest in the LRM is

$$m(x) = \mathbf{E}[y_i|x] = x'\beta$$

Prediction: regression intervals

- One object of interest in the LRM is

$$m(x) = \mathbf{E}[y_i|x] = x'\beta$$

- How does one estimate this function? Reasonable candidate is:

$$\hat{m}(x) = x'\hat{\beta}$$

Prediction: regression intervals

- One object of interest in the LRM is

$$m(x) = \mathbf{E}[y_i|x] = x'\beta$$

- How does one estimate this function? Reasonable candidate is:

$$\hat{m}(x) = x'\hat{\beta}$$

- Note that for fixed x,

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

so that

$$\left(x'\hat{V}_\beta x\right)^{-1/2}\sqrt{n}[\hat{m}(x) - m(x)] \xrightarrow{d} N(0,1)$$

Prediction: regression intervals

- One object of interest in the LRM is

$$m(x) = \mathbf{E}[y_i|x] = x'\beta$$

- How does one estimate this function? Reasonable candidate is:

$$\hat{m}(x) = x'\hat{\beta}$$

- Note that for fixed x,

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

so that

$$\left(x'\hat{V}_\beta x\right)^{-1/2}\sqrt{n}[\hat{m}(x) - m(x)] \xrightarrow{d} N(0,1)$$

- That motivates the approximation

$$\hat{m}(x) \approx N\left(m(x), \frac{1}{n}x'\hat{V}_\beta x\right)$$

Prediction: forecast errors

- Suppose you expect to obtain one additional data point: (y_{n+1}, x_{n+1})

Prediction: forecast errors

- Suppose you expect to obtain one additional data point: (y_{n+1}, x_{n+1})
- How would you predict y_{n+1}? Quadratic loss function implies best prediction is

$$y^*_{n+1} = \mathbf{E}[y_i | x_{n+1}] = m(x_{n+1}) = x'_{n+1}\beta$$

Prediction: forecast errors

- Suppose you expect to obtain one additional data point: (y_{n+1}, x_{n+1})
- How would you predict y_{n+1}? Quadratic loss function implies best prediction is

$$y^*_{n+1} = \mathbf{E}[y_i | x_{n+1}] = m(x_{n+1}) = x'_{n+1}\beta$$

- The above is infeasible; use plug-in estimate $\hat{m}(x)$

$$\tilde{y}_{n+1} = x'_{n+1}\hat{\beta}$$

Prediction: forecast errors

- Suppose you expect to obtain one additional data point: (y_{n+1}, x_{n+1})
- How would you predict y_{n+1}? Quadratic loss function implies best prediction is

$$y^*_{n+1} = \mathbf{E}[y_i|x_{n+1}] = m(x_{n+1}) = x'_{n+1}\beta$$

- The above is infeasible; use plug-in estimate $\hat{m}(x)$

$$\tilde{y}_{n+1} = x'_{n+1}\hat{\beta}$$

- Forecast error is

$$\tilde{e}_{n+1} = y_{n+1} - x'\hat{\beta} = x'\beta + e_i - x'\hat{\beta} = e_{n+1} + x'(\beta - \hat{\beta})$$

  - The true regression has an error e_{n+1}
  - In addition to that, there is an estimation noise coming from $\hat{\beta}$ being different than β

Prediction: forecast errors

- Finite sample variance of e_{n+1}:

$$\mathbf{E}[\hat{e}_{n+1}^2 | x_{n+1}] = \sigma^2(x_{n+1}) + x_{n+1}' V_{\hat{\beta}} x_{n+1}$$

- Can we build a confidence interval for the forecast?