Distribution of Estimates

• From Econometrics (410)
• Linear Regression Model \( y_t = \alpha + \beta x_t + e_t \)
  – Assume \((y_t, x_t)\) is iid and \(E(x_t e_t) = 0\)
• Estimation Consistency
  – The estimates approach the true values as the sample size increases
  – Estimation variance decreases as the sample size increases
Illustration of Consistency

• Take random sample of U.S. white men
• Estimate linear regression of log(wages) on education
• Total sample = 2089
• Start with 100 observations, sequentially increase to 2089
Sequence of Slope Coefficients

![Graph showing the sequence of slope coefficients with observations ranging from 0 to 10,000.]
Asymptotic Normality

\[ y_t = \alpha + \beta x_t + e_t \]

\[ \hat{\beta} \overset{a}{\sim} N(\beta, \sigma^2_{\hat{\beta}}) \]

\[ \sigma^2_{\hat{\beta}} = \frac{1}{T} \frac{\text{var}(x_t e_t)}{\text{var}(x_t)^2} \]
Illustration of Asymptotic Normality
Time Series

• Do these results apply to time-series data?
  – Consistency
  – Asymptotic normality
  – Variance formula

• Time-series models
  – AR models, i.e., \( x_t = y_{t-1} \)
  – Trend and seasonal models
  – One-step and multi-step forecasting
Derivation of Variance Formula

• For simplicity
  – Assume the variables have zero mean
  – The regression has no intercept

• Model with no intercept:

\[ y_t = \beta x_t + e_t \]
• Model with no intercept

\[ y_t = \beta x_t + e_t \]

• OLS minimizes the sum of squares

\[
\sum_{t=1}^{T} (y_t - \beta x_t)^2 = \sum_{t=1}^{T} y_t^2 - 2\beta \sum_{t=1}^{T} x_t y_t + \beta^2 \sum_{t=1}^{T} x_t^2
\]

• The first-order condition is

\[
0 = -2 \sum_{t=1}^{T} x_t y_t + 2 \hat{\beta} \sum_{t=1}^{T} x_t^2
\]
• Solution
\[ \hat{\beta} = \frac{\sum_{t=1}^{T} x_t y_t}{\sum_{t=1}^{T} x_t^2} = \frac{1}{T} \sum_{t=1}^{T} x_t y_t \]
\[ = \frac{1}{T} \sum_{t=1}^{T} x_t^2 \]

• Now substitute \( y_t = \beta x_t + e_t \)

\[ \hat{\beta} = \frac{1}{T} \sum_{t=1}^{T} x_t (x_t \beta + e_t) \]
\[ = \beta + \frac{1}{T} \sum_{t=1}^{T} x_t e_t \]
• We have

\[ \hat{\beta} = \beta + \frac{1}{T} \sum_{t=1}^{T} x_t e_t \]

\[ = \beta + \frac{1}{T} \sum_{t=1}^{T} \frac{1}{T} \sum_{t=1}^{T} x_t^2 \]

• The denominator is the sample variance (when \( x \) has mean zero), so

\[ \frac{1}{T} \sum_{t=1}^{T} x_t^2 \sim var(x_t) \]
Then
\[ \hat{\beta} \sim \beta + \frac{\sum_{t=1}^{T} v_t}{T \text{ var}(x_t)} \]

where \( v_t = x_t e_t \)

Since \( E(v_t) = E(x_t e_t) = 0 \)

then
\[ \text{var}(\hat{\beta}) \sim \frac{\text{var}\left( \sum_{t=1}^{T} v_t \right)}{[T \text{ var}(x_t)]^2} \]
• From the covariance formula

\[
\text{var}\left( \sum_{t=1}^{T} v_t \right) = \sum_{t=1}^{T} \text{var}(v_t) + \sum_{j \neq t} \text{cov}(v_t, v_j)
\]

• When the observations are independent, the covariances are zero.

• And since \( \text{var}(v_t) = \text{var}(x_t e_t) \)

we obtain

\[
\text{var}\left( \sum_{t=1}^{T} v_t \right) = T \text{var}(x_t e_t)
\]
• We have found

$$\text{var}(\hat{\beta}) \sim \frac{T \text{var}(x_t e_t)}{[T \text{var}(x_t)]^2} = \frac{\text{var}(x_t e_t)}{T[\text{var}(x_t)]^2}$$

as stated at the beginning
Extension to Time-Series

• The only place in this argument where we used the assumption of the *independence* of observations was to show that \( v_t = x_t e_t \) has zero covariance with \( v_j = x_j e_j \)

• This is saying that \( v_t \) is not autocorrelated.

• When does this happen in time-series?
Unforecastable one-step errors

- **Claim**: In one-step-ahead forecasting, if the regression error is **unforecastable** then $v_t$ is not autocorrelated.
- In this case, the variance formula for the least-squares estimate is the same as in regression:

\[
\text{var}(\hat{\beta}) \sim \frac{\text{var}(x_t e_t)}{T[\text{var}(x_t)]^2}
\]
• Why is the claim true?
• The error is unforecastable if $E(e_t \mid \Omega_{t-1})=0$
• For simplicity suppose $x_t=1$
• Then for $t \neq j$

$$\text{cov}(v_t, v_j) = E(e_t e_j) = 0$$
Summary

• In one-step-ahead time-series models, if the error is unforecastable, then least-squares estimates satisfy the asymptotic (approximate) distribution

\[ \hat{\beta} \sim N(\beta, \sigma_\beta^2) \]

\[ \sigma_\beta^2 = \frac{1}{T} \frac{\text{var}(x_t e_t)}{[\text{var}(x_t)]^2} \]

• As the sample size $T$ is in the denominator, the variance decreases as the sample size increases.

• This means that least-squares is consistent
Variance Formula

- The variance formula for the least-squares estimate takes the form

\[ \sigma^2_{\hat{\beta}} = \frac{1}{T} \frac{\text{var}(x_t e_t)}{[\text{var}(x_t)]^2} \]

- This formula is valid in time-series regression when the error is unforecastable.
Classical Variance Formula

If we make the simplifying assumption

\[ \text{var}(x_t e_t) = \text{var}(x_t) \text{var}(e_t) \]

then

\[ \sigma^2_{\hat{\beta}} = \frac{1}{T} \frac{\text{var}(x_t e_t)}{[\text{var}(x_t)]^2} = \frac{1}{T} \frac{\text{var}(e_t)}{\text{var}(x_t)} \]

This can be a useful simplification
HOMOSKEDASTICITY

• The variance simplification is valid under “conditional homoskedasticity”

\[
E(e_t \mid \Omega_{t-1}) = 0 \\
E(e_t^2 \mid \Omega_{t-1}) = \sigma^2
\]

• This is a simplifying assumption made to make calculations easier, and is a conventional assumption in introductory econometrics courses

• It is not used in serious econometrics
Variance Formula : AR(1) Model

• Take the AR(1) model with unforecastable homoskedastic errors

\[ y_t = \alpha + \beta y_{t-1} + e_t \]
\[ E(e_t \mid \Omega_{t-1}) = 0 \]
\[ E(e_t^2 \mid \Omega_{t-1}) = \sigma^2 \]

• Then the variance of the OLS estimate is

\[ \sigma_{\hat{\beta}}^2 = \frac{1}{T} \frac{\text{var}(e_t)}{\text{var}(x_t)} = \frac{1}{T} \frac{\text{var}(e_t)}{\text{var}(y_{t-1})} \]

since \( x_t = y_{t-1} \) in this model
AR(1) Asymptotic Variance

• We know that

\[ \text{var}(y_{t-1}) = \frac{\text{var}(e_t)}{1 - \beta^2} \]

• So

\[ \sigma_{\beta}^2 = \frac{1}{T} \frac{\text{var}(e_t)}{\text{var}(y_{t-1})} = \frac{1 - \beta^2}{T} \]

• The asymptotic distribution is very simple

\[ \hat{\beta} \sim N\left( \beta, \frac{1 - \beta^2}{T} \right) \]
\[\hat{\beta} \sim N\left(\beta, \frac{1 - \beta^2}{T}\right)\]

- The variance is a function of the unknown true value of \(\beta\)
- As \(|\beta|\) increases, the variance decreases, so the OLS estimate is actually more precise
Distribution of Least-Squares

• In classic regression, if the errors are iid normal, and independent of the regressors, then the least-squares estimates have an exact normal distribution, not just asymptotic.

• This is not true in most time-series regressions.
Non-Classical Distributions

• Estimates in autoregressive models
  – Biased downwards
  – Skewed
  – Thick tails

• Especially
  – When autoregressive coefficients are large
  – Sample sizes are small

• These issues diminish in large samples
Example

• Take the AR(1) model with intercept

\[ y_t = \alpha + \beta y_{t-1} + e_t \]

• \( e_t \sim N(0,1) \)
• \( T=100, 500, 1000 \)
• \( \beta=0.0, \beta=0.5, \beta=0.9, \)
• Numerically calculate distribution of least-squares estimate of \( \beta \)
Distribution, $\beta=0.0$
Distribution, $\beta=0.5$
Distribution, $\beta=0.9$
Interpretation

- Estimates of autoregressive parameters are random
- Even if regression error is normal, the parameter estimates are not normally distributed
- Distributions are less normal when AR coefficient is large
- Distributions are more concentrated and normal when sample size is large
Asymptotic Standard Deviation

• The least-squares estimate is asymptotically (approximately) normally distributed
• In the simple model \( y_t = \beta x_t + e_t \)

then

\[ \hat{\beta} \sim ^a N(\beta, \sigma^2_{\hat{\beta}}) \]

\[ \sigma^2_{\hat{\beta}} = \frac{1}{T} \frac{\text{var}(x_t e_t)}{[\text{var}(x_t)]^2} \]

• The standard deviation measures the precision of the estimate, but it is unknown.
Standard Errors

• Estimates of the standard deviations are called **standard errors**, and are reported in regression output.
• They are used to measure estimation precision.
Classical standard errors

A classic standard error is an estimate of the standard deviation from the formula

$$\hat{\sigma}^2_{\beta} = \frac{1}{n} \frac{\text{var}(e_t)}{\text{var}(x_t)}$$

This formula is valid under conditional homoskedasticity

$$E(e_t | \Omega_{t-1}) = 0$$
$$E(e_t^2 | \Omega_{t-1}) = \sigma^2$$
Robust standard errors

• “Robust” standard errors are estimates of

\[ \sigma_\beta = \sqrt{\frac{1}{n} \frac{\text{var}(x_t e_t)}{[\text{var}(x_t)]^2}} \]

• These are the conventional standard errors for regression analysis

• Also known as “White” standard errors
Halbert White

• Professor Hal White, UCSD (1950-2012)
• Leading contributor to econometric methods, especially time series analysis
• Introduced robust standard errors into econometrics (1980)
  – Most referenced paper in economics
• Founded Bates-White consulting firm, a leader in economic policy analysis
Have you seen robust standard errors?

• If you took an econometrics course other than 410, you may not be familiar with robust standard errors

• If you are currently taking 410, you won’t cover robust standard errors until later in the course
  – Wooldridge uses the homoskedasticity assumption in the early part of his text

• Stock-Watson use robust standard errors throughout
Does the Choice Matter?

• Classic standard errors are for the assumption of conditional homoskedasticity

\[ E(e_t^2 \mid \Omega_{t-1}) = \sigma^2 \]

• This is **unforecastability in the variance**
  – This is not implied by conventional unforecastability
  – It may be a convenient approximation for macro data
  – It is a bad assumption (quite false) in financial data
The robust standard error on the AR(1) coefficient is almost twice as large as the conventional standard error.
Computation

• In STATA, the default is conventional standard errors.
• They are automatically reported with the `regress (reg)` command
• For robust standard errors, use the “r” option
  `.reg return L.return, r`
Example: Real GDP Growth

```
. reg gdp L(1/4).gdp
```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 267</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>653.311643</td>
<td>4</td>
<td>163.327911</td>
<td>F(4, 262) = 12.14</td>
</tr>
<tr>
<td>Residual</td>
<td>3525.83667</td>
<td>262</td>
<td>13.4573919</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>4179.14831</td>
<td>266</td>
<td>15.7110839</td>
<td>R-squared = 0.1563</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adj R-squared = 0.1434</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Root MSE = 3.6684</td>
</tr>
</tbody>
</table>

| gdp       | Coef.   | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|-----------|---------|-----------|-------|------|---------------------|
| gdp       | .3324889 | .0616142  | 5.40  | 0.000 | .2111668 .4538109   |
| L1.       | .1406371 | .0647062  | 2.17  | 0.031 | .0132267 .2680475   |
| L2.       | -.0682077 | .0646485 | -1.06 | 0.292 | -.1955046 .0590891 |
| L3.       | -.0744968 | .0616124 | -1.21 | 0.228 | -.1958153 .0468216 |
| _cons     | 2.18892  | .3609659 | 6.06  | 0.000 | 1.478157 2.899684  |
With Robust st. errors

. reg gdp L(1/4).gdp, r

Linear regression

Number of obs = 267
F(  4,  262) =  9.51
Prob > F    =  0.0000
R-squared   =  0.1563
Root MSE    =  3.6684

| Coef. | Std. Err. | t     | P>|t|  |  [95% Conf. Interval] |
|-------|-----------|-------|------|-----------------------|
| gdp   |           |       |      |                       |
| gdp   |  0.3324889| 0.074552| 4.46 | 0.0000                |
| L1.   | -0.0682077| 0.0721403| -0.95| 0.345                 |
| L2.   | -0.0744968| 0.0749086| -0.99| 0.321                 |
| L3.   | -0.0682077| 0.0721403| -0.95| 0.345                 |
| L4.   | -0.0744968| 0.0749086| -0.99| 0.321                 |
| _cons|  2.18892  | 0.4365194| 5.01 | 0.0000                |
Robust st. errors

• With the “r” option
  .reg y x, r

• You get robust
  – Standard errors
  – t statistics and p-values
  – test statistics
Annoyance

• In STATA, with the “r” option, STATA omits sum of squared error table
  – Yet this can be useful
• So both commands may be useful
  .reg y x
  .reg y x, r
Interpretation of standard errors

- The standard errors measure precision of the estimate
  - Forecasts use *estimated* coefficients.
- Small standard errors mean the estimate is precise
  - Good for forecasting
- Large standard errors mean the estimate is not precise
  - Bad for forecasting
  - Inaccurate estimates leads to inaccurate forecasts

|        | Coef. | Std. Err. | t    | P>|t|  | [95% Conf. Interval] |
|--------|-------|-----------|------|------|----------------------|
| gdp    | 0.3324889 | 0.074552  | 4.46 | 0.000 | [0.1856916, 0.4792862] |
| L1.    | 0.1406371 | 0.0833804 | 1.69 | 0.093 | [-0.0235438, 0.304818] |
| L2.    | -0.0682077 | 0.0721403 | -0.95 | 0.345 | [-0.2102563, 0.0738409] |
| L3.    | -0.0744968 | 0.0749086 | -0.99 | 0.321 | [-0.2219964, 0.0730027] |
| L4.    | 2.18892    | 0.4365194 | 5.01 | 0.000 | [1.329387, 3.048453]   |
Interpretation of t-statistics

• “t” is the coefficient estimate divided by the standard error.
• It is used to test if the coefficient is zero
  – “P>|t|” is the p-value of the t-statistic
  – If p<.05 you “reject” the hypothesis of a zero coefficient
• Hypothesis tests are useful for assessing economic theories
  – But are less useful for picking good forecasting models

|     | Coef.  | Std. Err. | t      | P>|t| | [95% Conf. Interval] |
|-----|--------|-----------|--------|------|---------------------|
| gdp | .3324889 | .074552   | 4.46   | 0.000 | .1856916 .4792862   |
| L1. | .1406371 | .0833804  | 1.69   | 0.093 | -.0235438 .304818   |
| L2. | -.0682077 | .0721403  | -.95   | 0.345 | -.2102563 .0738409  |
| L3. | -.0744968 | .0749086  | -.99   | 0.321 | -.2219964 .0730027  |
| L4. | 2.18892  | .4365194  | 5.01   | 0.000 | 1.329387 3.048453   |
| _cons |        |           |        |      |         |
Interpretation of Confidence Interval

- The 95% interval is the coefficient estimate plus and minus 1.96 times the standard error
- Helps gauge possible values for the true coefficient
- Useful tool

|       | Coef.  | Std. Err. |   t   | P>|t|  | [95% Conf. Interval] |
|-------|--------|-----------|-------|------|----------------------|
|   gdp | 0.3324889 | 0.074552 | 4.46  | 0.000 | 0.1856916 - 0.4792862 |
| L1.   | 0.1406371 | 0.0833804| 1.69  | 0.093 | -0.0235438 - 0.304818 |
| L2.   | -0.0682077| 0.0721403| -0.95 | 0.345 | -0.2102563 - 0.0738409 |
| L3.   | -0.0744968| 0.0749086| -0.99 | 0.321 | -0.2219964 - 0.0730027 |
| _cons | 2.18892  | 0.4365194| 5.01  | 0.000 | 1.329387 - 3.048453   |
Summary

• In one-step-ahead forecast regressions with unforecastable errors
  – Robust standard errors generally appropriate
  – Classical standard errors appropriate under conditional homoskedasticity