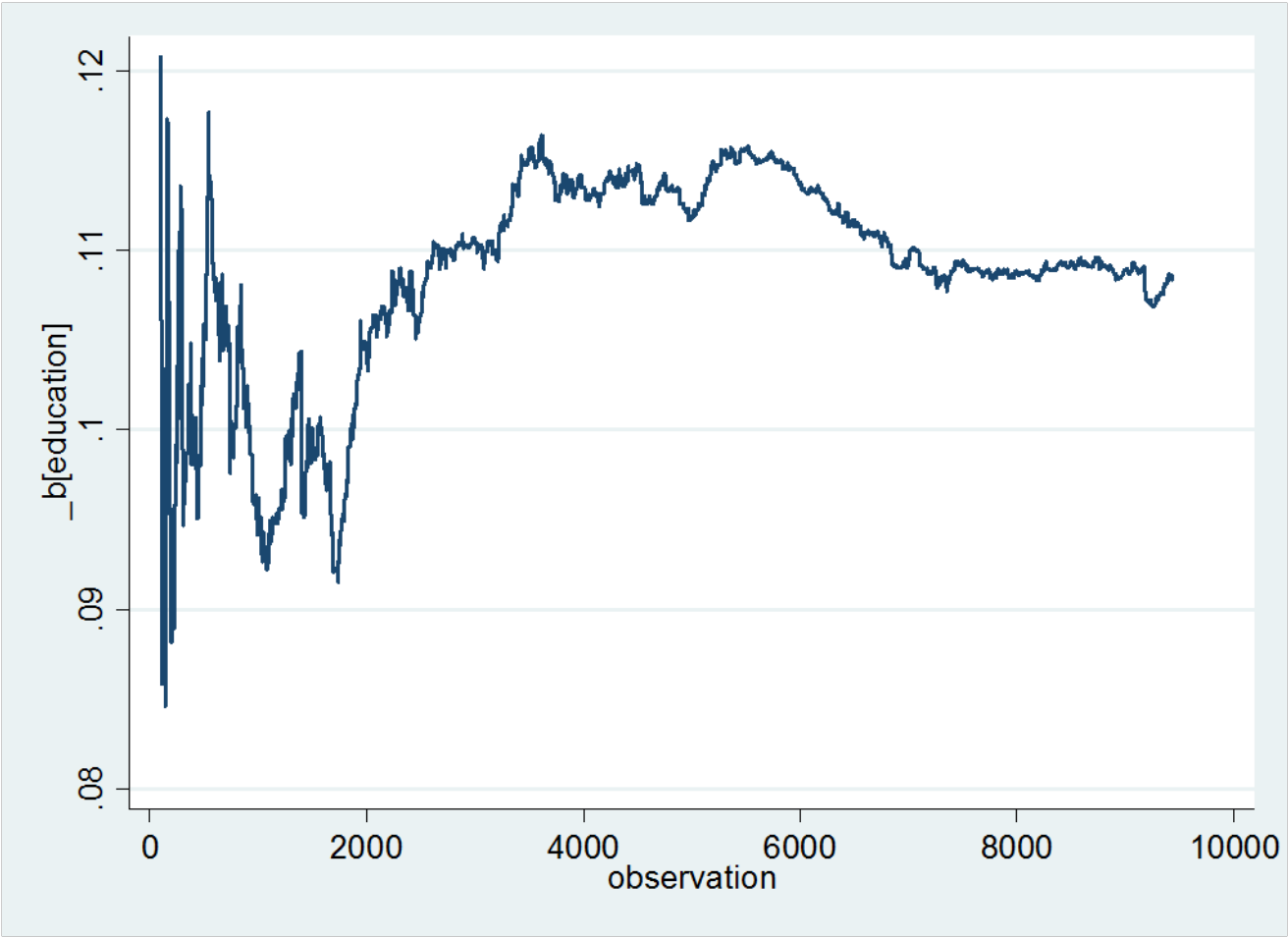
Distribution of Estimates

- 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



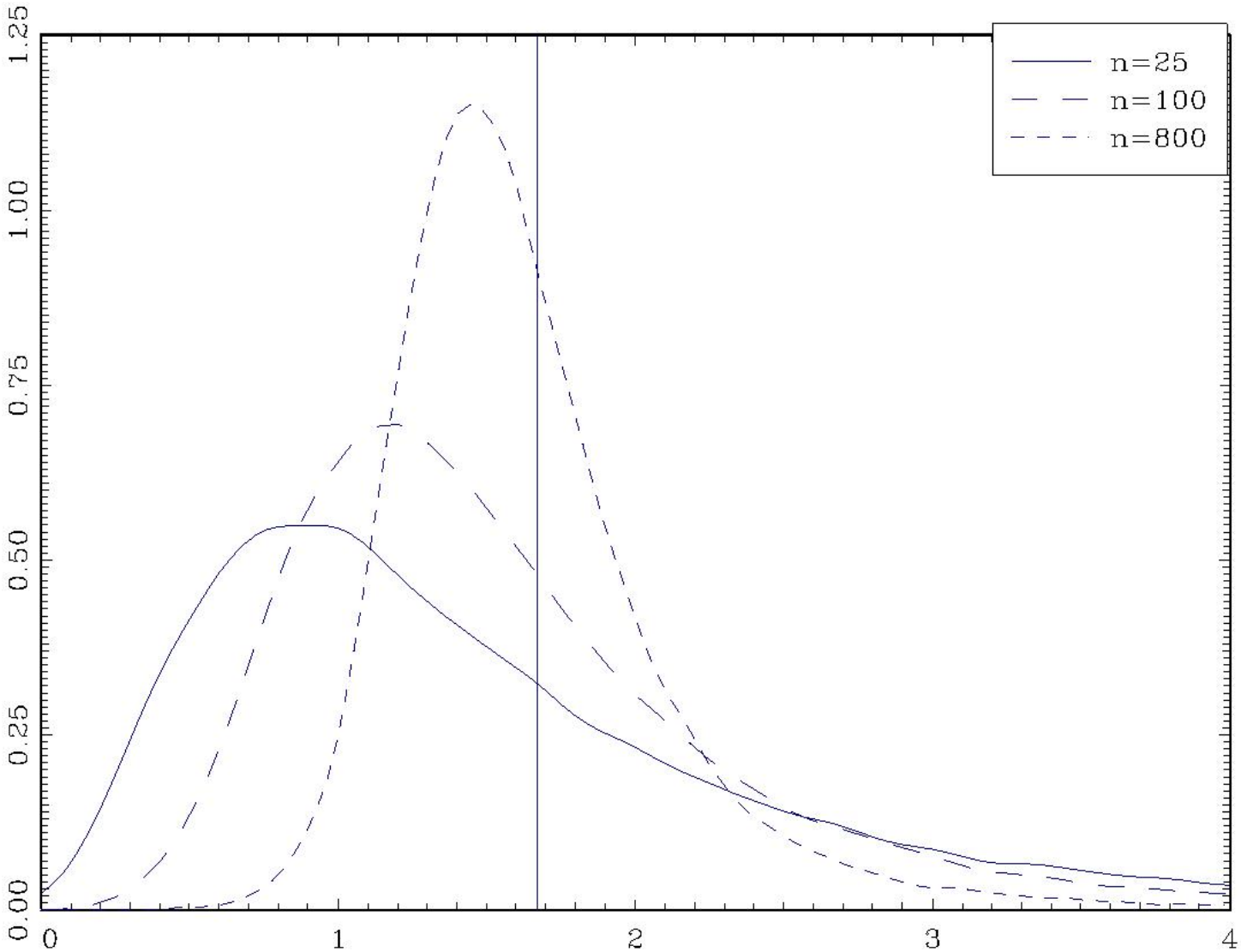
Asymptotic Normality

$$y_t = \alpha + \beta x_t + e_t$$

$$\hat{\beta} \overset{a}{\sim} N\left(\beta, \sigma^2_{\hat{\beta}}\right)$$

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

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

$$\sum_{t=1}^{T}\left(y_t - \beta x_t\right)^2 = \sum_{t=1}^{T} y_t^2 - 2\beta\sum_{t=1}^{T} x_t y_t + \beta^2\sum_{t=1}^{T} x_t^2$$

- 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

$$\hat{\beta} = \frac{\sum_{t=1}^{T} x_t y_t}{\sum_{t=1}^{T} x_t^2} = \frac{\frac{1}{T}\sum_{t=1}^{T} x_t y_t}{\frac{1}{T}\sum_{t=1}^{T} x_t^2}$$

- Now substitute $y_t = \beta x_t + e_t$

$$\hat{\beta} = \frac{\frac{1}{T}\sum_{t=1}^{T} x_t \left(x_t \beta + e_t \right)}{\frac{1}{T}\sum_{t=1}^{T} x_t^2} = \beta + \frac{\frac{1}{T}\sum_{t=1}^{T} x_t e_t}{\frac{1}{T}\sum_{t=1}^{T} x_t^2}$$

- We have

$$\hat{\beta} = \beta + \frac{\frac{1}{T}\sum_{t=1}^{T} x_t e_t}{\frac{1}{T}\sum_{t=1}^{T} x_t^2}$$

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

$$\frac{1}{T}\sum_{t=1}^{T} x_t^2 \overset{a}{\sim} \mathrm{var}(x_t)$$

- Then $$\hat{\beta} \overset{a}{\sim} \beta + \frac{\sum_{t=1}^{T} v_t}{T \operatorname{var}(x_t)}$$

  where $v_t = x_t e_t$

- Since $E(v_t) = E(x_t e_t) = 0$

  then

$$\operatorname{var}(\hat{\beta}) \overset{a}{\sim} \frac{\operatorname{var}\left(\sum_{t=1}^{T} v_t\right)}{\left[T \operatorname{var}(x_t)\right]^2}$$

- From the covariance formula

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

- When the observations are independent, the covariances are zero.
- And since $\operatorname{var}(v_t) = \operatorname{var}(x_t e_t)$

  we obtain $\operatorname{var}\left(\sum_{t=1}^{T} v_t\right) = T \operatorname{var}(x_t e_t)$

- We have found

$$\operatorname{var}\left(\hat{\beta}\right) \overset{a}{\sim} \frac{T \operatorname{var}\left(x_t e_t\right)}{\left[T \operatorname{var}\left(x_t\right)\right]^2} = \frac{\operatorname{var}\left(x_t e_t\right)}{T\left[\operatorname{var}\left(x_t\right)\right]^2}$$

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 $v_t=x_te_t$ has zero covariance with $v_j=x_je_j$
- 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

$$\operatorname{var}\left(\hat{\beta}\right) \overset{a}{\sim} \frac{\operatorname{var}\left(x_t e_t\right)}{T\left[\operatorname{var}\left(x_t\right)\right]^2}$$

- 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 $t \neq j$

$$\operatorname{cov}(v_t, v_j) = E\left(e_t e_j\right) = 0$$

Summary

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

$$\hat{\beta} \overset{a}{\sim} N\left(\beta, \sigma_{\hat{\beta}}^2\right)$$

$$\sigma_{\hat{\beta}}^2 = \frac{1}{T} \frac{\mathrm{var}\left(x_t e_t\right)}{\left[\mathrm{var}\left(x_t\right)\right]^2}$$

- 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

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

- This formula is valid in time-series regression when the error is unforecastable

Classical Variance Formula

If we make the simplifying assumption

$$\mathrm{var}(x_t e_t) = \mathrm{var}(x_t)\mathrm{var}(e_t)$$

then

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

This can be a useful simplification

Homoskedasticity

- The variance simplification is valid under "conditional homoskedasticity"

$$E\left(e_t \mid \Omega_{t-1}\right) = 0$$
$$E\left(e_t^2 \mid \Omega_{t-1}\right) = \sigma^2$$

- 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

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

- Then the variance of the OLS estimate is

$$\sigma_{\hat{\beta}}^2 = \frac{1}{T}\frac{\mathrm{var}\left(e_t\right)}{\mathrm{var}\left(x_t\right)} = \frac{1}{T}\frac{\mathrm{var}\left(e_t\right)}{\mathrm{var}\left(y_{t-1}\right)}$$

since $x_t = y_{t-1}$ in this model

AR(1) Asymptotic Variance

- We know that

$$\operatorname{var}(y_{t-1}) = \frac{\operatorname{var}(e_t)}{1-\beta^2}$$

- So

$$\sigma^2_{\hat{\beta}} = \frac{1}{T}\frac{\operatorname{var}(e_t)}{\operatorname{var}(y_{t-1})} = \frac{1-\beta^2}{T}$$

- The asymptotic distribution is very simple

$$\hat{\beta} \overset{a}{\sim} N\left(\beta, \frac{1-\beta^2}{T}\right)$$

$$\hat{\beta} \overset{a}{\sim} N\left(\beta, \frac{1-\beta^2}{T}\right)$$

- The variance is a function of the unknown true value of β
- 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

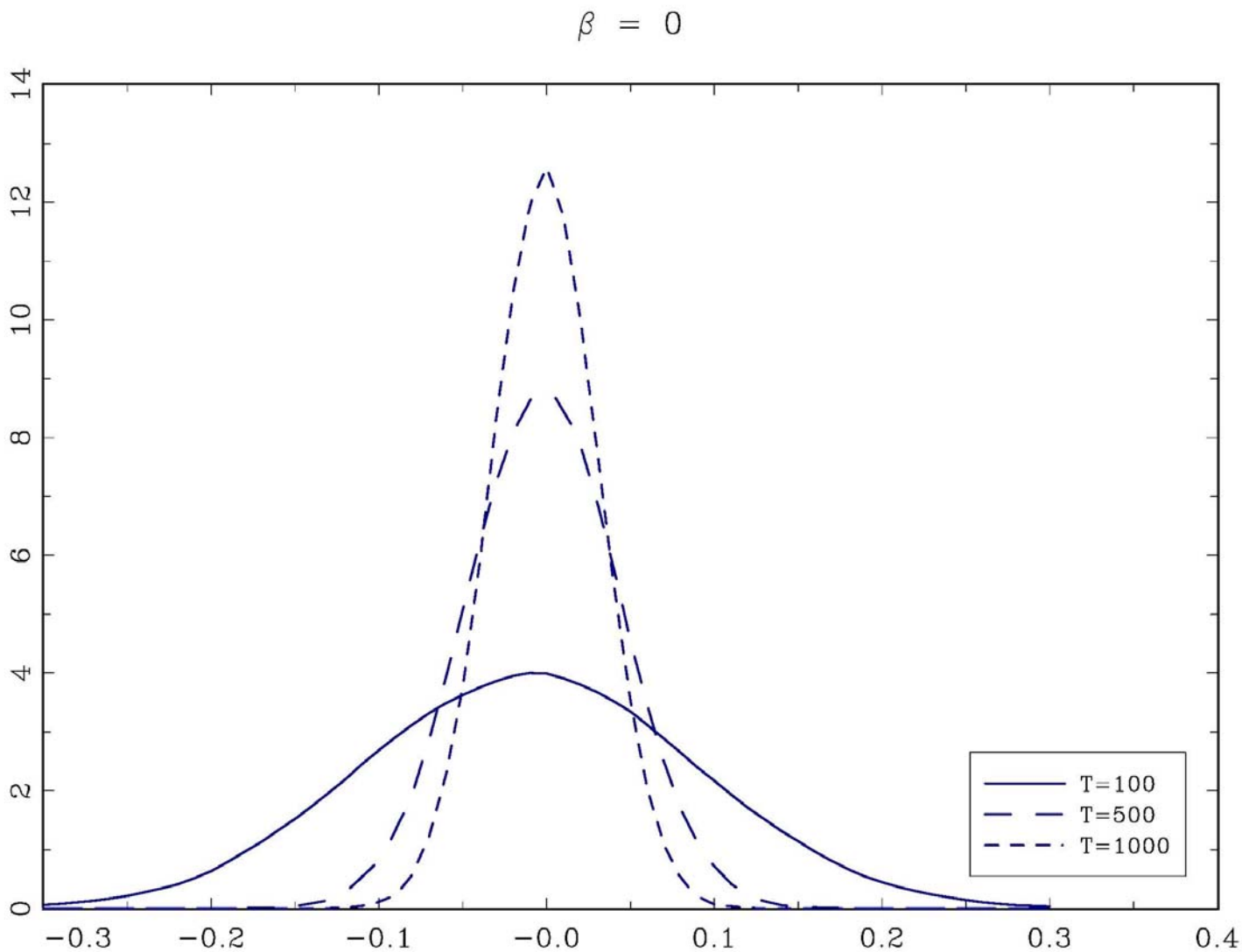
Example

- Take the AR(1) model with intercept

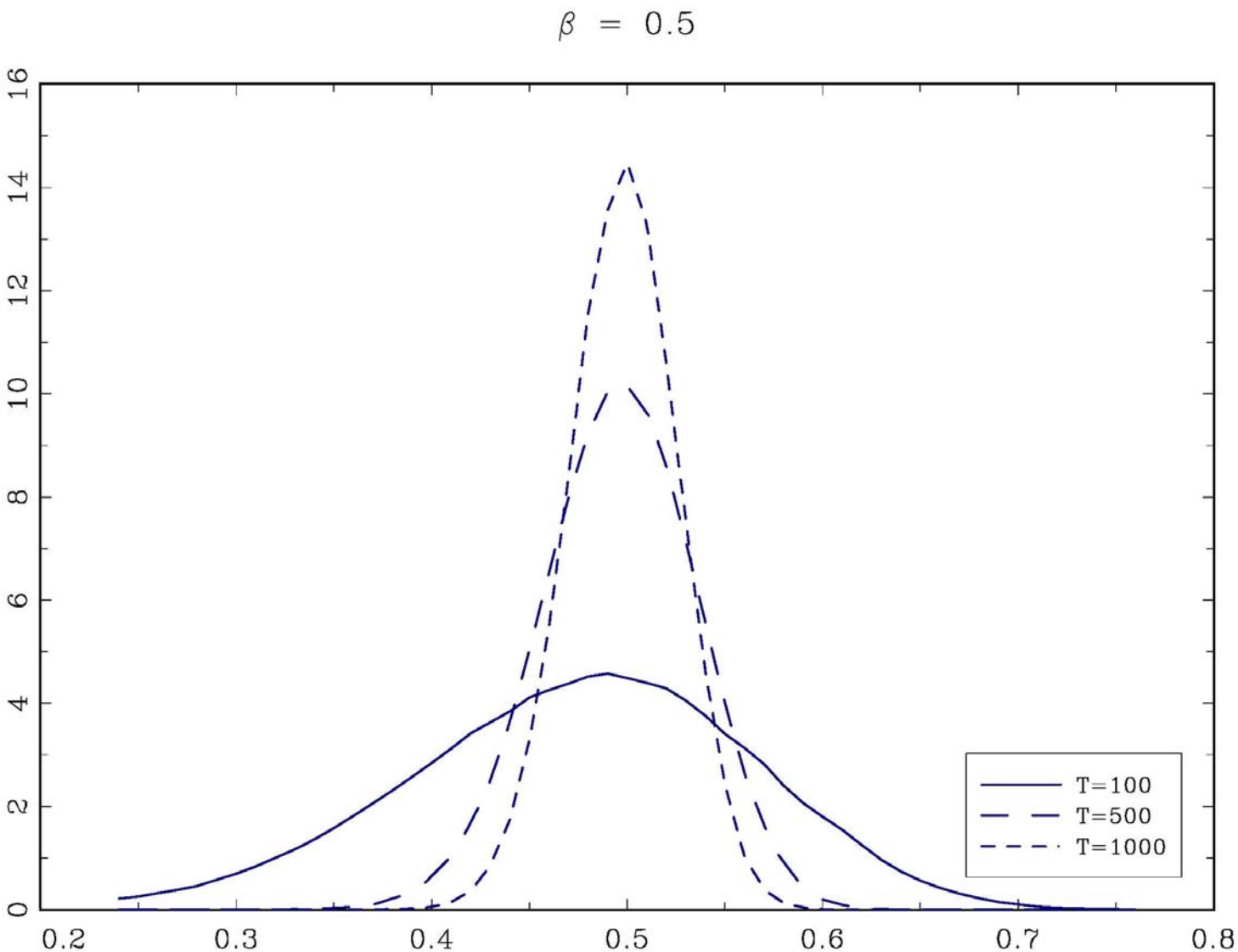
$$y_t = \alpha + \beta y_{t-1} + e_t$$

- e_t N(0,1)
- T=100, 500, 1000
- β=0.0, β=0.5, β=0.9,
- Numerically calculate distribution of least-squares estimate of β

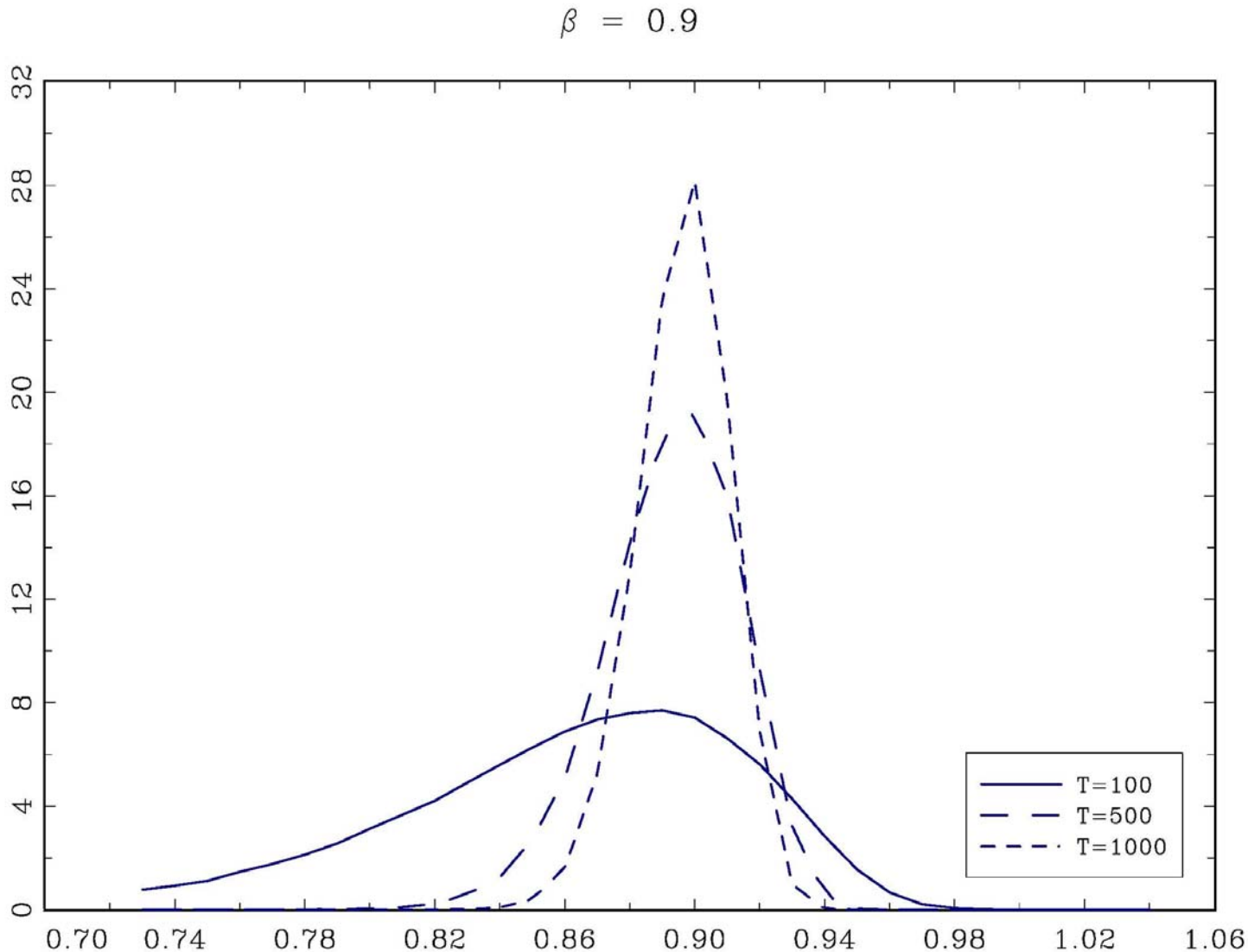
Distribution, β=0.0



Distribution, β=0.5



Distribution, β=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
$$\hat{\beta} \overset{a}{\sim} N\left(\beta, \sigma_{\hat{\beta}}^2\right)$$
$$\sigma_{\hat{\beta}}^2 = \frac{1}{T} \frac{\operatorname{var}\left(x_t e_t\right)}{\left[\operatorname{var}\left(x_t\right)\right]^2}$$

- 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

$$\sigma_{\hat{\beta}}^2 = \frac{1}{n}\frac{\mathrm{var}(e_t)}{\mathrm{var}(x_t)}$$

This formula is valid under conditional homoskedasticity

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

Robust standard errors

- "Robust" standard errors are estimates of

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

- 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
- Introducted 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

$$E\left(e_t^2 \mid \Omega_{t-1}\right) = \sigma^2$$

- 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

Example: Stock Returns, AR(1)

r	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
r						
L1.	-.0163895	.0179105	-0.92	0.360	-.051507	.0187279
_cons	.0013626	.0003728	3.66	0.000	.0006317	.0020935

r	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
r						
L1.	-.0163895	.032377	-0.51	0.613	-.0798718	.0470927
_cons	.0013626	.0003878	3.51	0.000	.0006022	.002123

- 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 y x, r

Example: Real GDP Growth

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

      Source |       SS       df       MS              Number of obs =     247
-------------+------------------------------           F(  4,   242) =   11.39
       Model |  662.232234     4  165.558059           Prob > F      =  0.0000
    Residual |  3518.78213   242   14.540422           R-squared     =  0.1584
-------------+------------------------------           Adj R-squared =  0.1445
       Total |  4181.01437   246  16.9959934           Root MSE      =  3.8132

------------------------------------------------------------------------------
         gdp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gdp |
         L1. |    .327656   .0640344     5.12   0.000     .2015202    .4537919
         L2. |   .1466135   .0670302     2.19   0.030     .0145764    .2786506
         L3. |  -.0980287    .066934    -1.46   0.144    -.2298764    .0338189
         L4. |  -.0889209   .0644466    -1.38   0.169    -.2158689     .038027
             |
       _cons |   2.378427    .389677     6.10   0.000     1.610836    3.146019
------------------------------------------------------------------------------
```

With Robust st. errors

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

Linear regression                                      Number of obs =     247
                                                       F(  4,   242) =    8.85
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.1584
                                                       Root MSE      =  3.8132
```

gdp	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
gdp						
L1.	.327656	.076895	4.26	0.000	.1761871	.479125
L2.	.1466135	.0858808	1.71	0.089	-.0225558	.3157828
L3.	-.0980287	.0728951	-1.34	0.180	-.2416186	.0455611
L4.	-.0889209	.0790354	-1.13	0.262	-.244606	.0667641
_cons	2.378427	.4731312	5.03	0.000	1.446447	3.310408

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

gdp	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
gdp						
L1.	.327656	.076895	4.26	0.000	.1761871	.479125
L2.	.1466135	.0858808	1.71	0.089	-.0225558	.3157828
L3.	-.0980287	.0728951	-1.34	0.180	-.2416186	.0455611
L4.	-.0889209	.0790354	-1.13	0.262	-.244606	.0667641
_cons	2.378427	.4731312	5.03	0.000	1.446447	3.310408

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

gdp	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
gdp						
L1.	.327656	.076895	4.26	0.000	.1761871	.479125
L2.	.1466135	.0858808	1.71	0.089	-.0225558	.3157828
L3.	-.0980287	.0728951	-1.34	0.180	-.2416186	.0455611
L4.	-.0889209	.0790354	-1.13	0.262	-.244606	.0667641
_cons	2.378427	.4731312	5.03	0.000	1.446447	3.310408

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

gdp	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
gdp						
L1.	.327656	.076895	4.26	0.000	.1761871	.479125
L2.	.1466135	.0858808	1.71	0.089	-.0225558	.3157828
L3.	-.0980287	.0728951	-1.34	0.180	-.2416186	.0455611
L4.	-.0889209	.0790354	-1.13	0.262	-.244606	.0667641
_cons	2.378427	.4731312	5.03	0.000	1.446447	3.310408

Summary

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