AFRE 835: Introductory Econometrics

Chapter 4: Multiple Regression Analysis: Inference

Spring 2017

Introduction

- In chapter 3, we covered the basic properties of the OLS estimator, focusing on its mean and variance;
- However, inference regarding the statistical significance of individual or combinations of parameter requires more.
- Specifically, we need to know something about the distribution of the OLS estimator.
- In this chapter, we rely on additional assumptions to pin down the distribution of $\hat{\boldsymbol{\beta}}$.
- Chapter 5 discusses a set of weaker assumptions to achieve similar results, but only asymptotically (i.e., in large samples).

Outline

Normality

- To pin down the sampling distribution for the OLS estimator, we add the assumption that the errors u are normally distributed in the population; i.e.,

  Assumption MLR.6 (Normality): The popular error u is *independent* of the explanatory variables $\boldsymbol{x} = (x_1, \ldots, x_k)'$ with zero mean and a variance of σ^2; i.e., $u|\boldsymbol{x} \sim \mathcal{N}(0, \sigma^2)$.
- Assumptions MLR.1 through MLR.6 are jointly referred to as the *Classical Linear Model (CLM)* assumptions.
- They imply that

$$y|\boldsymbol{x} \sim \mathcal{N}(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k, \sigma^2). \tag{1}$$

Implications of the CLM Assumptions

- An important implication of the CLM assumptions is that OLS is no longer *just* BLUE, but it is also the *minimum variance* unbiased estimator (i.e., it has the lowest variance among *all* unbiased estimators).
- It is also straightforward to show that the OLS estimator itself is normal.
- Specifically, we know that

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{y} \\ &= (\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'(\boldsymbol{x}\boldsymbol{\beta} + \boldsymbol{u}) \\ &= \boldsymbol{\beta} + (\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{u} \qquad (2)\end{aligned}$$

- Conditional on the independent variables ($\boldsymbol{x}$), the OLS estimator is just a linear combination of normal random variables and, hence, is itself normally distributed.

Implications of the CLM Assumptions (cont'd)

- Furthermore, since

$$E(\hat{\boldsymbol{\beta}}|\boldsymbol{x}) = \boldsymbol{\beta} + (\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'E(\boldsymbol{u}|\boldsymbol{x}) = \boldsymbol{\beta} \qquad (3)$$

and

$$\begin{aligned}Var(\hat{\boldsymbol{\beta}}|\boldsymbol{x}) &= E[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'|\boldsymbol{x}] \\ &= E[(\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{u}\boldsymbol{u}'\boldsymbol{x}(\boldsymbol{x}'\boldsymbol{x})^{-1}|\boldsymbol{x}] \\ &= (\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'E[\boldsymbol{u}\boldsymbol{u}'|\boldsymbol{x}]\boldsymbol{x}(\boldsymbol{x}'\boldsymbol{x})^{-1} \\ &= \sigma^2(\boldsymbol{x}'\boldsymbol{x})^{-1} \qquad (4)\end{aligned}$$

so that $\hat{\boldsymbol{\beta}}|\boldsymbol{x} \sim \mathcal{N}[\boldsymbol{\beta}, \sigma^2(\boldsymbol{x}'\boldsymbol{x})^{-1}]$

- Note: This implies that $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ are *jointly* normal, so that
  - Any subset of the $\hat{\beta}$'s are jointly normal;
  - Any linear combination of the $\hat{\beta}$ is normal.

Theorem 4.1

- Written in the more familiar form, we have

  Theorem 4.1 (Normal Sampling Distributions): Under the CLM assumptions, MLR.1 through MLR.6, conditional on the sample values of the independent variables:

$$\hat{\beta}_j \sim \mathcal{N}[\beta_j, Var(\hat{\beta}_j)] \quad (5)$$

  and

$$\frac{\hat{\beta}_j - \beta_j}{sd(\hat{\beta}_j)} \sim \mathcal{N}(0,1) \quad (6)$$

- Knowing the sampling distribution of the $\hat{\beta}$'s, we can now conduct hypothesis tests.
- Chapter 5 demonstrates that the normality of the OLS estimator holds approximately in large samples, even without MLR.6.

Testing Individual Parameters

- One of the most common class of hypotheses tested in econometrics are those focused on a single parameter.
- While Theorem 4.1 is helpful in this regard, the sampling distributions presented assume we know σ^2, which we usually don't.
- An additional result is helpful here:

  Theorem 4.2 (t-distribution of the standardized estimator): Under the CLM assumptions,

$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t_{n-k-1} = t_{df} \quad (7)$$

  where k is the number of slope parameters in the model and $n-k-1$ is the degrees of freedom (df).

Testing $H_0 : \beta_j = 0$

- Suppose that we want to know whether or not an independent variable belongs in the population regression function.
- This corresponds to hypothesizing

$$H_0 : \beta_j = 0. \tag{8}$$

- Theorem 4.2 implies that *under this null hypothesis* the t-statistic

$$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \sim t_{n-k-1} \tag{9}$$

- We are more likely to reject the null hypothesis if
  - $\hat{\beta}_j$ differs substantially from zero.
  - $se(\hat{\beta}_j)$ is small.

One-Side versus Two-Sided Tests

- The form of the hypothesis test depends upon what the relevant alternative hypothesis is.
- In many settings, we are only interested in whether a variable belongs in the population regression function, regardless of the sign; e.g.,
  - In a model of recreational travel, does the age or gender impact the number of trips taken in a season.
  - In a model of housing demand, do housing prices vary with proximity to the city center?

    In these cases, a two-sided alternative is appropriate; i.e., $H_A : \beta_j \neq 0$.
- However, we we often are interested only in departures from zero in one direction; e.g.,
  - In a model of loan approvals, is there racial discrimination.
  - In a model of health outcomes, does pollution increase mortality or morbidity rates?

    In these cases, a one-sided alternative is appropriate; i.e., $H_A : \beta_j < 0$ or $H_A : \beta_j > 0$.
- The alternative hypothesis should be set *prior* to looking at the data.

Testing Against One-Sided Alternatives

- Suppose we are interested in the null hypothesis $H_0 : \beta_j \leq 0$ versus the alternative $H_A : \beta_j > 0$.
  ... This would be the case, for example, if we were interested in the impact of pollution on mortality or morbidity rates.
- We would want to reject the null hypothesis only if there is strong enough evidence against H_0.
  - It would not be enough to look at whether $\hat{\beta}_j$ itself is large, because it might be large by chance.
  - We would want it large relative to the precision with which it was estimated;
  - This is precisely what the t-statistic $t_{\hat{\beta}_j}$ measures.
  - Moreover, we know that distribution of the the t-stat, allowing us to calculate the probability of making a mistake.

Choosing the Critical Level

- Suppose we decide to reject the null hypothesis $H_0 : \beta_j \leq 0$ in favor of the alternative $H_A : \beta_j > 0$ if $t_{\hat{\beta}_j} > c$, where c is our *critical level*.
- We can use the information about the distribution of $t_{\hat{\beta}_j}$ to compute the probability of making a mistake by rejecting H_0 when in fact it is true (i.e., the probability of a Type I error).
- Specifically:

$$\begin{aligned} Pr(t_{\hat{\beta}_j} > c | H_0 \text{ is true}) &= Pr(t_{\hat{\beta}_j} > c) \\ &= Pr(t_{n-k-1} > c) \\ &= \alpha. \end{aligned} \tag{10}$$

  where α denotes the *significance level* of our test.
- If we want a 5% chance of a Type I error, we would choose a value of c such that $\alpha = 0.05$ (or 5%).
- Table G.2 provides critical values for one-tailed tests.
- For example, with $\alpha = 0.05$ and $n - k - 1 = 28$, $c = 1.701$.

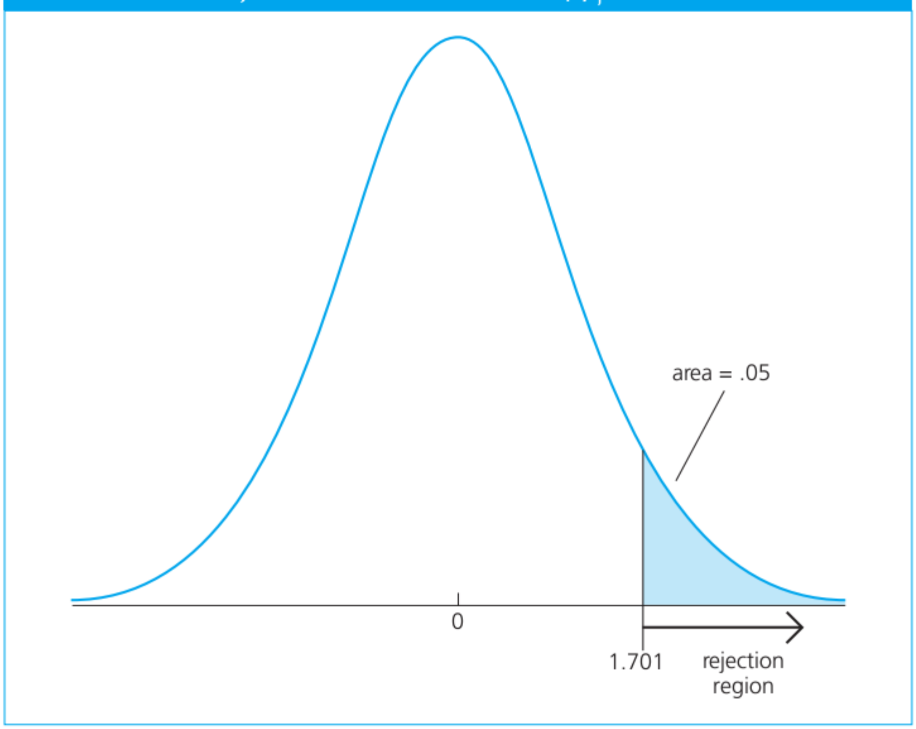
TABLE G.2 Critical Values of the *t* Distribution

		Significance Level			
1-Tailed:	**.10**	**.05**	**.025**	**.01**	**.005**
2-Tailed:	**.20**	**.10**	**.05**	**.02**	**.01**
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
90	1.291	1.662	1.987	2.368	2.632
120	1.289	1.658	1.980	2.358	2.617
∞	1.282	1.645	1.960	2.326	2.576

Degrees of Freedom

One-Sided Hypothesis $H_0 : \beta_j \leq 0$ vs. $H_A : \beta_j > 0$

FIGURE 4.2 5% rejection rule for the alternative H_1: $\beta_j > 0$ with 28 *df*.



One-Sided Hypothesis $H_0 : \beta_j \geq 0$ vs. $H_A : \beta_j < 0$

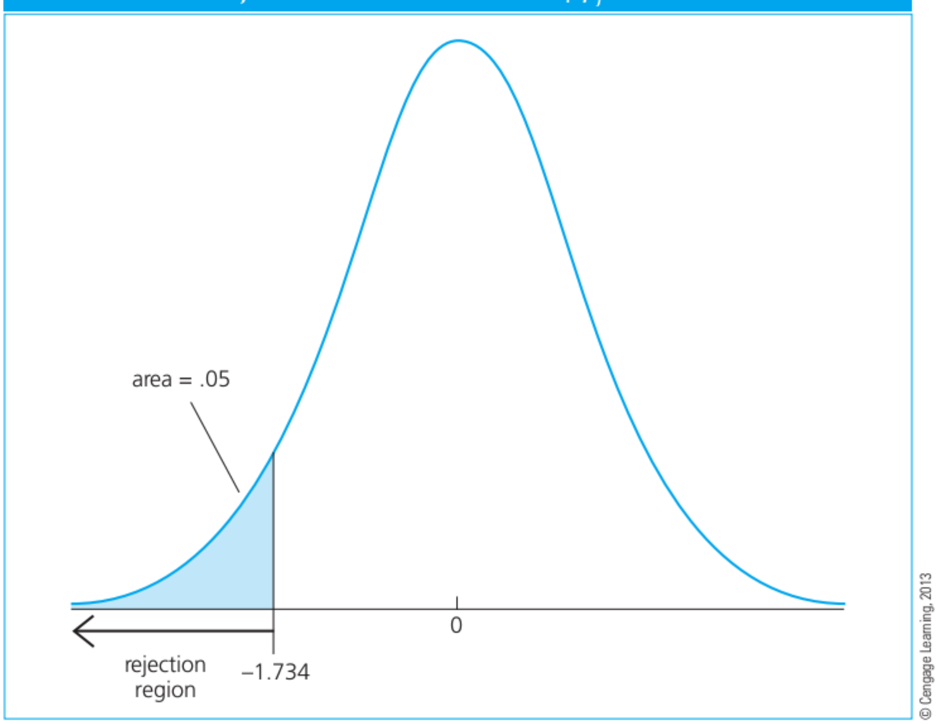
- There will, of course, also be situations in which the alternative hypothesis goes in the other direction.
- For example, this would be the case if we were studying the impact of race or gender on load approval rates.
- Now we want to reject the null hypothesis in favor of the alternative if $\hat{\beta}_j$ is sufficiently negative.
- Our rejection rule becomes:

$$t_{\hat{\beta}_j} < -c \tag{11}$$

where c is read off of Table G.2 using the one-tailed significance levels.

One-Sided Hypothesis $H_0 : \beta_j \geq 0$ vs. $H_A : \beta_j < 0$

FIGURE 4.3 5% rejection rule for the alternative H_1: $\beta_j < 0$ with 18 *df*.



Two-Sided Hypothesis Test $H_0 : \beta_j = 0$ vs. $H_A : \beta_j \neq 0$

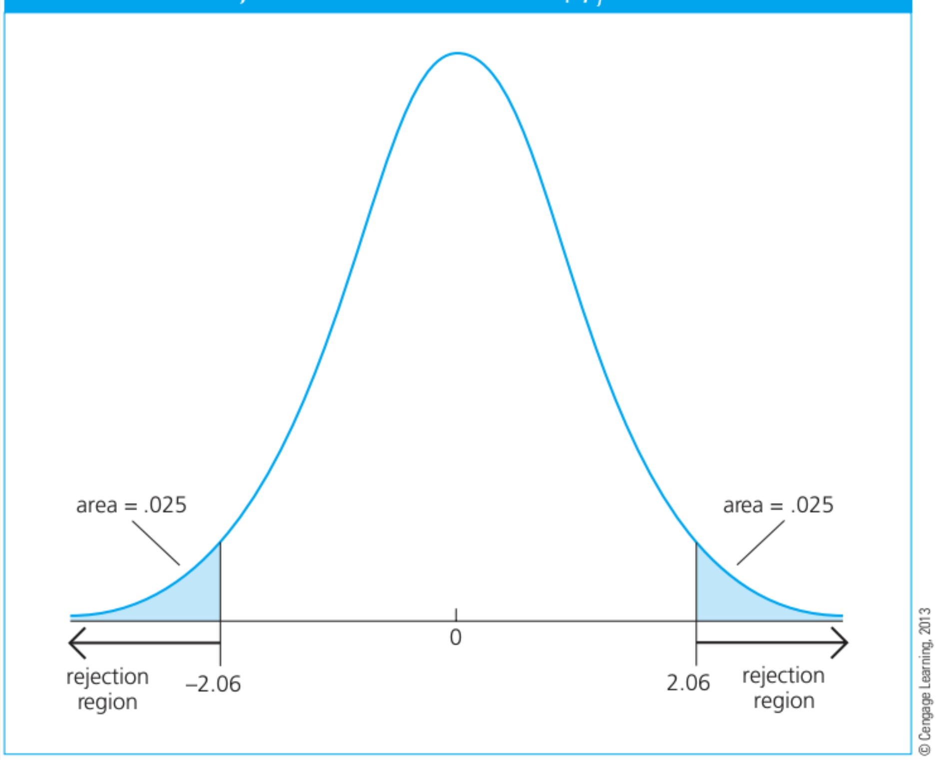
- If we only want to test whether a coefficient belongs in the population regression function, and not a particular direction of its effect, then a two-tailed test is appropriate.
- This would be the case, for example, if we wanted to know whether the cross-price elasticity of demand between two commodities in double-log model differed from zero.
- Our rejection rule becomes:

$$|t_{\hat{\beta}_j}| > c \tag{12}$$

  where c is read off of Table G.2 using the two-tailed significance levels.
- This allows us to reject the null hypothesis if either the t-stat is too big or the t-stat is too small, splitting these possibilities evenly.
- The next figure illustrates this for $n - k - 1$=25 and $\alpha = 0.05$.

Two-Sided Hypothesis $H_0 : \beta_j = 0$ vs. $H_A : \beta_j \neq 0$

FIGURE 4.4 5% rejection rule for the alternative $H_1: \beta_j \neq 0$ with 25 *df*.



Testing Other Possible Values for β_j

- Extending these testing procedures to allow for a non-zero hypothesized value for β_j is straightforward.
- For example, suppose we are in the two-sided setting, with $H_0 : \beta_j = a_j$ vs. $H_A : \beta_j \neq a_j$.
- This would be the case, for example, if we were testing whether or not the demand for a commodity had unitary income elasticity in a double-log model of demand, with β_j denoting the income elasticity and $a_j = 1$.
- The appropriate t-statistic in this case becomes

$$t = \frac{\hat{\beta}_j - a_j}{se(\hat{\beta}_j)} \sim t_{n-k-1} \tag{13}$$

Testing Other Possible Values for β_j (cont'd)

- This result should make intuitive sense if we consider defining $\theta_j = \beta_j - a_j$.
- Our hypotheses become: $H_0 : \theta_j = 0$ vs. $H_A : \theta_j \neq 0$.
- Moreover, $\hat{\theta}_j \equiv \hat{\beta}_j - a_j$, like $\hat{\beta}_j$, is normally distributed, just with a mean reduced by a constant a_j.
- Also, $\hat{\theta}_j$ and $\hat{\beta}_j$ will have the same variances, so that $se(\hat{\beta}_j)$ provides a consistent estimator for the standard deviation of $\hat{\theta}_j$.
- This implies that

$$t = \frac{\hat{\theta}_j}{se(\hat{\beta}_j)} = \frac{\hat{\beta}_j - a_j}{se(\hat{\beta}_j)} \sim t_{n-k-1} \tag{14}$$

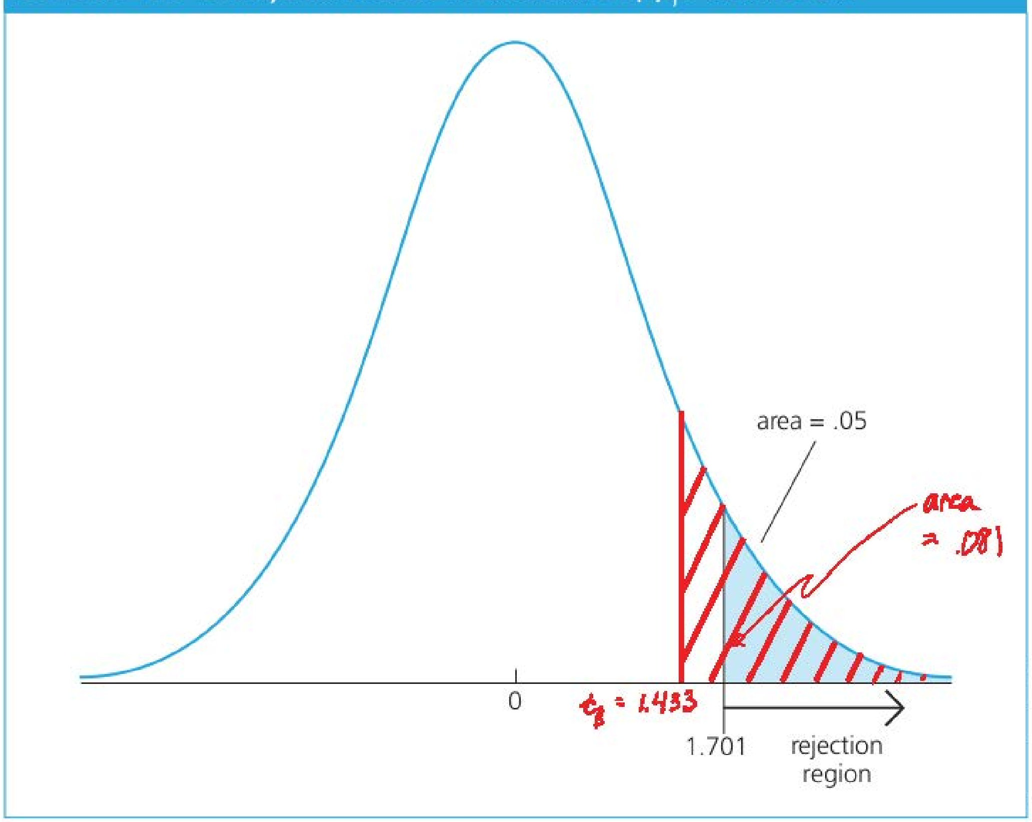
P-values

- The choice of the significance level is essentially arbitrary, with tradition more than anything else setting $\alpha = 0.10$, 0.05 or 0.01.
- An alternative approach is to report the so-called **p-value**: The lowest significance level at which the null hypothesis would be rejected.
- The p-value provides the probability of the Type I error if we reject the null hypothesis, so a smaller the p-value the more willing we should be to reject H_0.
- In the case of a one-tailed test, with $H_0 : \beta_j \leq 0$ vs. $H_A : \beta_j > 0$:

$$p - value = Pr(T_{n-k-1} > t_{\hat{\beta}_j}) = Pr(T_{n-k-1} \geq t_{\hat{\beta}_j}) \qquad (15)$$

  where T_{n-k-1} is a t-distributed random variable with $df = n - k - 1$.
- Essentially, we are finding how much of a t-distribution lies to the right of our observed t-statistic $t_{\hat{\beta}_j}$.
- Consider again a case with $df = 28$ and suppose that $t_{\hat{\beta}_j} = 1.433$.

FIGURE 4.2 5% rejection rule for the alternative H_1: $\beta_j > 0$ with 28 df.



Two-Sided P-value

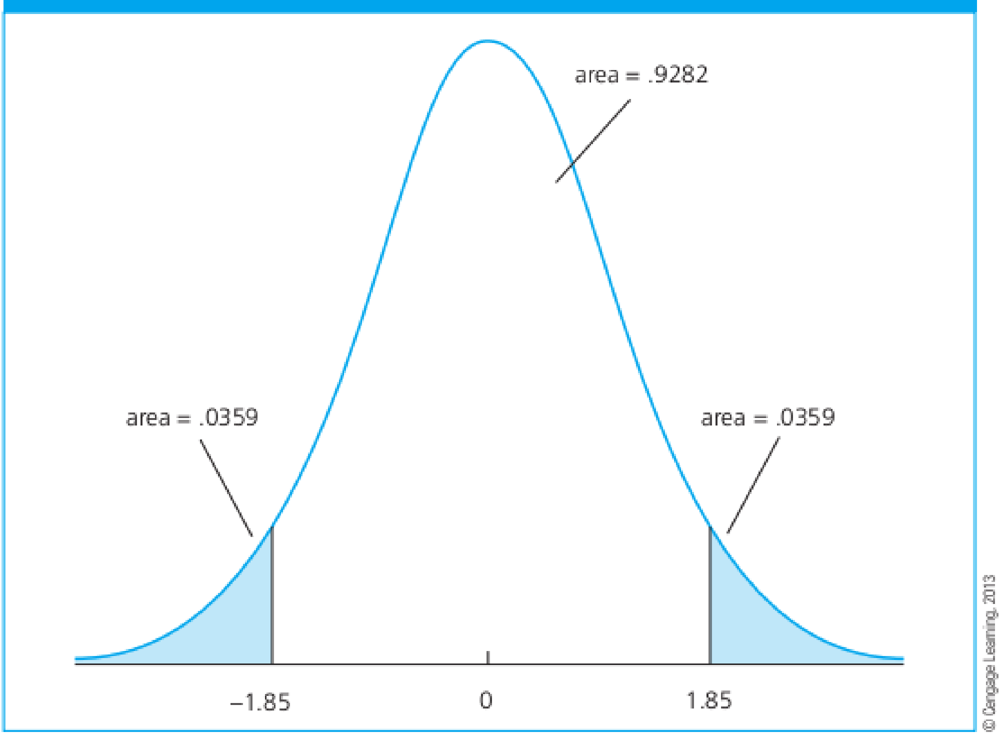
- The idea is the same in the case of a two-tailed hypothesis test, with $H_0 : \beta_j = 0$ vs. $H_A : \beta_j \neq 0$.
- We want to find how much of the t-distribution lies just to the right of $|t_{\hat{\beta}_j}|$ and to the left of $-|t_{\hat{\beta}_j}|$
- Formally, we want to find

$$\begin{aligned} p - value &= P\left(|T_{df}| > |t_{\hat{\beta}_j}|\right) \\ &= 2P\left(T_{df} > |t_{\hat{\beta}_j}|\right) \end{aligned} \tag{16}$$

- In the example in Figure 4.6, $t_{\hat{\beta}_j} = 1.85$ and $df = 40$, so that

$$\begin{aligned} p - value &= P\left(|T_{40}| > 1.85\right) \\ &= 2P\left(T_{40} > 1.85\right) \\ &= 2(0.0359) \\ &= 0.0718. \end{aligned} \tag{17}$$

FIGURE 4.6 Obtaining the *p*-value against a two-sided alternative, when $t = 1.85$ and $df = 40$.



Confidence Intervals

- Confidence intervals (CI's) provide an alternative way of representing the uncertainty associated with estimators.
- Formally, the 95% confidence interval for $\hat{\beta}_j$ when we have $df = n - k - 1$ is given by $[\underline{\beta}_j, \bar{\beta}_j]$, where

$$\underline{\beta}_j = \hat{\beta}_j - c \cdot se(\hat{\beta}_j) \tag{18}$$

  and

$$\bar{\beta}_j = \hat{\beta}_j + c \cdot se(\hat{\beta}_j) \tag{19}$$

  with c is the two-tailed critical value for the 5%-significance level.
- It is importance to keep in mind that what is uncertain here is not the true value β_j, but our estimator of it.
- In repeated sampling, $[\underline{\beta}_j, \bar{\beta}_j]$ will contain the true value roughly 95% of the time.

Testing a Single Linear Restriction

- We are often interested in restriction on linear combinations of parameters.
- By a linear restriction we mean:

$$a_0\beta_0 + a_1\beta_1 + \cdots + a_k\beta_k = b \tag{20}$$

  where the a_j's and b are known constants.
- Examples include:
  - $\beta_1 = \beta_2$, which can be written using $a_1 = 1$, $a_2 = -1$, $a_j = 0 \ \forall j \notin \{1, 2\}$, and $b = 0$, or simply $1 \cdot \beta_1 + (-1)\beta_2 = 0$.
  - $\sum_{j=1}^{n} \beta_j = 1$, using $a_0 = 0$, $a_j = 1 \forall j \neq 0$ and $b = 1$.

The t-stat Approach

- One way to proceed is to use a t-statistic.
- For simplicity, we will restrict our attention to the simpler case in which $k = 2$ and $a_0 = 0$, but the approach generalizes.
- In this case, our linear restriction becomes: $a_1\beta_1 + a_2\beta_2 = b$
- Define: $\theta = a_1\beta_1 + a_2\beta_2 - b$.
- Suppose further that our hypothesis of interest is two-sided, with $H_0 : \theta = 0$ and $H_A : \theta \neq 0$.
- Our (unbiased) OLS estimator of θ would be

$$\hat{\theta} = a_1\hat{\beta}_1 + a_2\hat{\beta}_2 - b \tag{21}$$

- The relevant t-statistic in this case would be

$$t_{\hat{\theta}} = \frac{\hat{\theta}}{se(\hat{\theta})} \sim t_{n-k-1} \tag{22}$$

The t-stat Approach(cont'd)

- We could then proceed in the usual way *if* we had for $se(\hat{\theta})$, an estimator for the standard deviation of $\hat{\theta}$.
- Using the formula for the variance of linear combinations of random variables, we know that:

$$Var(\hat{\theta}) = a_1^2 Var(\hat{\beta}_1) + a_2^2 Var(\hat{\beta}_2) + 2a_1a_2Cov(\hat{\beta}_1, \hat{\beta}_2) \tag{23}$$

- From this, we can construct a consistent estimator for $sd(\hat{\theta})$ using:

$$se(\hat{\theta}) = \left\{a_1^2[se(\hat{\beta}_1)]^2 + a_2^2[se(\hat{\beta}_2]^2 + 2a_1a_2s_{12}\right\}^{\frac{1}{2}} \tag{24}$$

where s_{12} is an estimate of $Cov(\hat{\beta}_1, \hat{\beta}_2)$.
- We can then proceed in the usual way to test our hypothesis.

Example: Constant Returns to Scale Test

- Consider estimating a model:

$$ln(GDP_t) = \beta_0 + \beta_1 ln(Labor_t) + \beta_2 ln(Capital_t) + u_t \qquad (25)$$

with our hypotheses given by $H_0 : \theta = 0$ and $H_A : \theta \neq 0$ where

$$\theta = a_1\beta_1 + a_2\beta_2 - b = \beta_1 + \beta_2 - 1 \qquad (26)$$

i.e., $a_1 = a_2 = 1$ and $b = 1$.

- The following table provide the OLS estimates of our model


```
. reg lngdp lnemp lncap

      Source |       SS       df       MS              Number of obs =      20
-------------+------------------------------           F(  2,    17) = 1719.20
       Model |  2.75165006     2  1.37582503           Prob > F      =  0.0000
    Residual |  .01360456     17  .000800268           R-squared     =  0.9951
-------------+------------------------------           Adj R-squared =  0.9945
       Total |  2.76525462    19  .145539717           Root MSE      =  .02829

------------------------------------------------------------------------------
       lngdp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lnemployment |   .3397362   .1856928     1.83   0.085    -.0520414    .7315138
   lncapital |   .8459951    .093352     9.06   0.000     .6490397    1.042951
       _cons |  -1.652429   .6062017    -2.73   0.014    -2.931402   -.3734547
------------------------------------------------------------------------------

. matrix cov=e(V)

. matrix list cov

symmetric cov[3,3]
              lnemployment     lncapital         _cons
lnemployment     .03448182
   lncapital    -.01703459     .00871459
       _cons    -.10494718      .0480547     .36748054
```

Example: Constant Returns to Scale Test

- Using our results, we then have

$$\hat{\theta} = \beta_1 + \beta_2 - 1 = 0.3397 + 0.8460 - 1 = 0.1857 \quad (27)$$

and

$$\begin{aligned} Var(\hat{\theta}) &= Var(\hat{\beta}_1) + Var(\hat{\beta}_2) + 2Cov(\hat{\beta}_1, \hat{\beta}_2) \\ &= 0.345 + 0.0087 + 2 * (-0.017) = 0.0091 \end{aligned} \quad (28)$$

so that $se(\hat{\theta}) = \sqrt{0.0091} = 0.0955$

- This gives us a t-stat of $t_{\hat{\theta}} = 0.1857/0.0955 = 1.944$.
- The critical level using a significance level of 5% and $df = 17$ is $c = 2.110$, so we would not reject the null hypothesis in this case.

Rewriting the Model

- Another way to proceed is to incorporate θ into the model.
- In particular, note that $\theta = a_1\beta_1 + a_2\beta_2 - b$ can be rewritten as:

$$\beta_1 = \tilde{a}_\theta \theta + \tilde{a}_2\beta_2 + \tilde{b} \quad (29)$$

where $\tilde{a}_\theta = 1/a_1$, $\tilde{a}_2 = -a_2/a_1$, and $\tilde{b} = b/a_1$.

- Substituting this into our model, we get

$$\begin{aligned} y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \\ &= \beta_0 + (\tilde{a}_\theta\theta + \tilde{a}_2\beta_2 + \tilde{b})x_1 + \beta_2 x_2 + u \\ &= \beta_0 + \theta \cdot (\tilde{a}_\theta x_1) + \beta_2 \cdot (\tilde{a}_2 x_1 + x_2) + \tilde{b}x_1 + u \\ &\Rightarrow \tilde{y} = \beta_0 + \theta\tilde{x}_1 + \beta_2\tilde{x}_2 + u \end{aligned} \quad (30)$$

where $\tilde{y} = y - \tilde{b}x_1$, $\tilde{x}_1 = \tilde{a}_\theta x_1$, and $\tilde{x}_2 = \tilde{a}_2 x_1 + x_2$.

- We can now directly estimate θ using (30) and construct t-stats for it.

Testing Multiple Linear Restrictions: The F-test

- In many settings, we will want to simultaneous test whether or not a group of independent variables (say, socio-demographic characteristics) impact our dependent variable (recreation demand).
- In general, consider a model with k independent variables, so that

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u \tag{31}$$

- Without loss of generality, suppose we are interested in testing the hypothesis that the last q of these independent variables do not belong in the population regression function; i.e.,

$$H_0 : \beta_{k-q-1} = 0, \ldots, \beta_k = 0 \tag{32}$$

with

$$H_A : H_0 \text{ is not true} \tag{33}$$

- Equation (31) is referred to as the **unrestricted model**, while the **restricted model** would be:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_{k-q} x_{k-q} + u \tag{34}$$

The F-Statistic

- One test statistic that can be used in this context is the **F-statistic** given by:

$$F = \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)} \tag{35}$$

where SSR_r denotes the sum of squared residuals from the restricted model and SSR_{ur} denotes the sum of squared residuals from the unrestricted model.
- Note that since $SSR_r \geq SSR_{ur}$, we know that $F \geq 0$.
- $q = df_r - df_{ur}$ is referred to as the numerator degrees of freedom and $n - k - 1 = df_{ur}$ is referred to as the denominator degrees of freedom.
- Under the CLM assumptions, F is distributed as an F random variable with $(q, n-k-1)$ degrees of freedom.

The F-Statistic (cont'd)

- The null hypothesis in this case is rejected if F is large enough; i.e., if the SSR increases "too much" when we move to the restricted model.
- Much like in the case of the one-sided t-test, we reject the null if $F > c$, where c denotes the critical value for a given significance level and a given set of numerator and denominator degrees of freedom.
- Table G.3 in Wooldridge provide one table of critical values.
- The F-Statistic can be written as a function of R^2's, i.e.,

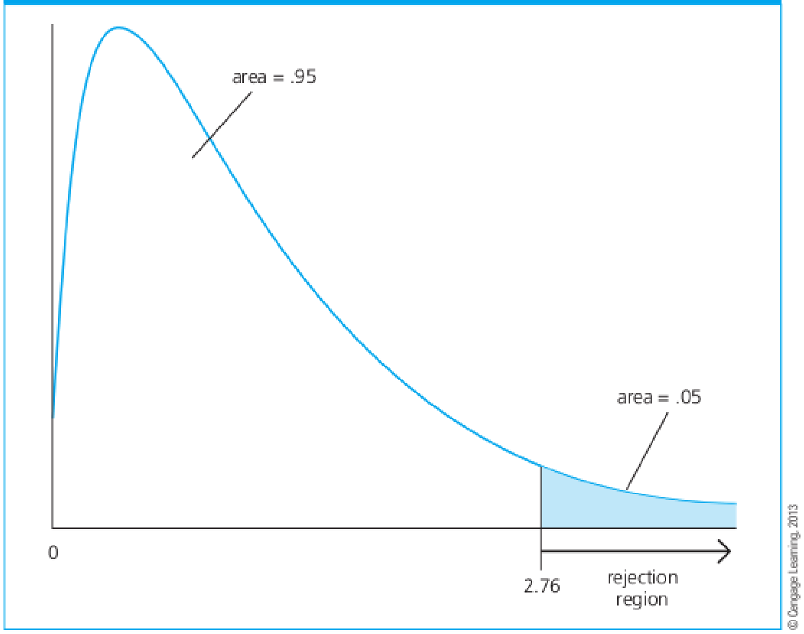
$$F = \frac{(R^2_{ur} - R^2_r)/q}{(1 - R^2_{ur})/(n-k-1)} \tag{36}$$

... Note: This formulation only holds if the dependent variable is the same when constructing both R^2's.

TABLE G.3b 5% Critical Values of the F Distribution

		Numerator Degrees of Freedom									
		1	2	3	4	5	6	7	8	9	10
Denominator Degrees of Freedom	10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
	11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
	12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
	13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
	14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
	15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
	16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
	17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
	18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
	19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
	20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
	21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
	22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
	23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
	24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
	25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
	26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
	27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
	28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
	29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
	30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
	40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
	60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
	90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94
	120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91

FIGURE 4.7 The 5% critical value and rejection region in an $F_{3,60}$ distribution.



Exclusion Restrictions

- In testing a set of exclusion restrictions (e.g., $H_0 : \beta_{k-q-1} = 0, \ldots, \beta_k = 0$) it is important to keep in mind that the corresponding individual restrictions may not indicate how the joint test will turn out.
  - One can reject the joint hypothesis, even when all of the individual hypothesis tests turn out to be insignificant.
  - One can reject individual hypotheses and still not reject the joint test

  ... though typically rejecting several of the individual hypotheses will signal rejection of the joint test.
- One special version of the exclusion restrictions jointly tests whether *all* of the independent variables can be excluded from the population regression function; i.e., $H_0 : \beta_1 = 0, \ldots, \beta_k = 0$
- The F-statistic in this case reduces to:

$$F = \frac{(R^2_{ur} - R^2_r)/q}{(1 - R^2_{ur})/(n-k-1)} = \frac{R^2/k}{(1 - R^2)/(n-k-1)} \tag{37}$$

Testing General Linear Restrictions

- Often times we may be interested in testing a series of linear restrictions (including several exclusion restrictions).
- The general form of the F-statistic becomes:

$$F = \frac{(R^2_{ur} - R^2_r)/(df_{ur} - df_r}{(1 - R^2_{ur})/(df_{ur})} \tag{38}$$

where df_r and df_{ur} denotes the number of degrees of freedom in the restricted and unrestricted models, respectively, with $df_{ur} - df_r = k_{ur} - k_r$ denoting the *effective* number of restrictions being imposed.

Reporting Results

- Tabular presentation of the results is usually appropriate, including:
  - Coefficient estimates and corresponding standard errors, with enough digits to enable the calculation of approximate t-statistics.
  - Clearly labeled variables.
  - Sample sizes used.
  - R^2's.
- In presenting your results you want to make it easy as possible for the reader to *see* the results.
- Among other things, this means
  - Avoiding the use of too many digits.
  - Grouping common factors.
  - Making comparisons of interest appear vertically and adjacent, rather than horizontally.

Example #1

TABLE 4.1 Testing the Salary-Benefits Tradeoff

Independent Variables	Dependent Variable: log(*salary*) (1)	(2)	(3)
b/s	−.825 (.200)	−.605 (.165)	−.589 (.165)
log(*enroll*)	——	.0874 (.0073)	.0881 (.0073)
log(*staff*)	——	−.222 (.050)	−.218 (.050)
droprate	——	——	−.00028 (.00161)
gradrate	——	——	.00097 .00066)
intercept	10.523 (0.042)	10.884 (0.252)	10.738 (0.258)
Observations	408	408	408
R-squared	.040	.353	.361

Example #1

Table 4: IV Coefficients

Variables	Dependent variable [redacted], IV Coefficient (SE)		
	ALL	Females	Males
[redacted]	0.383(0.047)***	(omitted)	(omitted)
[redacted]	0.186(0.034)***	0.189(0.032)***	−1.222(7.153)
[redacted]	−0.002(0.0004)***	−0.002(0.0003)***	0.018(0.104)
[redacted]	−0.237(0.063)***	−0.191(0.045)***	1.688(9.007)
[redacted]	0.087(0.030)***	0.095(0.031)***	0.770(3.427)
[redacted]	0.037(0.017)**	0.034(0.015)**	−0.042(0.200)
[redacted]	0.012(0.005)**	0.012(0.005)**	0.033(0.215)
[redacted]	0.049(0.005)***	0.046(0.004)***	−0.173(1.033)
[redacted]	0.381(0.108)***	0.347(0.085)***	−2.282(11.940)
[redacted]	0.526(0.164)***	0.509(0.137)***	−2.181(10.858)
[redacted]	0.205(0.0978)**	0.161(0.122)	1.579(6.289)
[redacted]	0.046(0.087)	0.011(0.115)	0.400(1.732)
[redacted]	−0.127(0.057)**	−0.177(0.056)***	−1.718(10.211)
[redacted]	−0.113(0.067)*	−0.168(0.065)***	−0.395(3.244)
[redacted]	−0.292(0.09)***	−0.306(0.080)***	−0.935(5.078)
[redacted]	0.0946(0.117)	−0.002(0.099)	0.514(6.041)
[redacted]	8.276(1.054)***	7.164(0.701)***	−14.39(99.005)
[redacted]	19549	14362	5187

*significant at 10%; **significant at 5%; ***significant at 1%.

Example #1 (cont'd)

---- Coefficients ----

	(b) IV	(B) OLS	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
[redacted]	-.2367433	-.0060941	-.2306492	.0252444
[redacted]	.1857527	.0804354	.1053173	.0115269
[redacted]	-.0020682	-.0008727	-.0011956	.0001309
[redacted]	.3836584	.2450026	.1386559	.0151758
[redacted]	.0878427	.1234722	-.0356294	.0038996
[redacted]	.0371918	-.0092951	.0464869	.0050879
[redacted]	.0119557	-.0059211	.0178767	.0019566
[redacted]	.0488842	.0310265	.0178577	.0019545
[redacted]	.3812876	-.0018738	.3831614	.0419367
[redacted]	.5263013	-.063601	.5899023	.0645643
[redacted]	.2048787	.3821584	-.1772797	.0194031
[redacted]	.0455292	.1437374	-.0982082	.0107488
[redacted]	-.1268754	.0255169	-.1523923	.0166792
[redacted]	-.1134802	.0147098	-.1281899	.0140303
[redacted]	-.2924411	-.0076508	-.2847903	.03117
[redacted]	.094608	-.0373376	.1319457	.0144413
[redacted]	8.275678	4.687988	3.58769	.3926695

Example #2

[redacted] Mean Estimation Result across 100 Monte Carlo Iterations [redacted])

	[redacted]		[redacted]			[redacted]			[redacted]			[redacted]		
		$\beta =$	-0.02	-0.05	-0.1	-0.02	-0.05	-0.1	-0.02	-0.05	-0.1	-0.02	-0.05	-0.1
		[redacted]	6.2	5.5	0.3	223.6	337.1	57.1	161.8	126.5	0.3	197.8	250.5	36.2
		[redacted]	0.1	0.1	0.1	1.6	4.2	9.3	1	1.7	1.8	0.7	0.4	0.5
	[redacted]	[redacted]	5.4	3.2	1	-	-	-	66	23.8	28.5	-	-	-
		[redacted]	441.4	157.3	122	449.1	167.3	133.3	444.9	156	121.9	442.8	154.6	121.3
		[redacted]	0.2	0.1	0.1	1.9	6.4	9.2	1	0.9	0.2	0.5	1.8	0.7
		[redacted]	0.1	1.7	0.1	65	1168.9	80.2	55.7	734.5	32.1	57.8	821.5	37
		[redacted]	0	0.1	0.1	1.9	5.4	10.7	0.1	2.6	4.7	0.6	2.1	4.6
[redacted]	[redacted]	[redacted]	0.2	0.5	0.1	-	-	-	90.4	81	82	-	-	-
		[redacted]	60.6	49.3	16.1	61.8	51.8	17.1	60.6	51	17.2	60.3	50.5	17.1
		[redacted]	0.2	0.2	0.1	2.1	4.9	6.5	0.1	3.3	7.1	0.5	2.4	6.5
		[redacted]	1.1	0.1	0.9	57.8	202.7	149.6	61	245.1	202.1	50.8	153	101
		[redacted]	0.1	0.1	0.1	2.5	6.8	15.1	1.3	2.4	3.9	0.9	1.8	3.2
	[redacted]	[redacted]	4	0.6	0.1	-	-	-	142	200.7	233.1	-	-	-
		[redacted]	35.9	15	6.5	36.8	16	7.7	36.4	15.3	6.8	36.2	15.3	6.7
		[redacted]	0.4	0.7	0.4	3	7.7	16.9	1.7	2.9	3.4	1.2	2.5	2.5
		[redacted]	0.3	0.3	0.1	166.2	262.5	43.6	139.9	157.9	16	153.2	224.4	37.3
		[redacted]	0.1	0.1	0.1	1.3	4.7	13.3	0.6	0.6	1.9	0.4	0.3	1.8
	[redacted]	[redacted]	0.3	0.1	0.6	-	-	-	84.4	52.8	24.2	-	-	-
		[redacted]	280.2	140.4	149.6	272.1	143.5	157.1	277.9	141.1	150.5	279.3	141.4	152.2
		[redacted]	0.2	0.2	0.1	2.8	2.1	5.1	0.7	0.4	0.7	0.2	0.6	1.8
		[redacted]	0.2	1.8	0.3	71.2	1014.9	45.3	63.6	778.8	31.2	64	815.6	36.2
		[redacted]	0.2	0.1	0.1	2.6	6.2	14.7	0.2	0.2	1.5	0.2	0.1	1.5
[redacted]	[redacted]	[redacted]	1.2	1.1	0.6	-	-	-	98.7	91.3	80.8	-	-	-
		[redacted]	85.6	53.3	23.4	87.8	56.7	28.5	85.3	53.3	23	85.3	53.3	22.9
		[redacted]	0.1	0.2	0.1	2.8	6.6	21.7	0.3	0.3	1.9	0.3	0.2	2.3
		[redacted]	1.9	0.4	0.1	70.5	216.8	69.9	73.9	248.5	116.3	67.3	211.7	89.8
		[redacted]	0.1	0.1	0.1	3.1	9	18.9	0.2	0.5	1.2	0.2	0.3	0.9
	[redacted]	[redacted]	6.2	0.2	0.8	-	-	-	123.2	136.7	157.3	-	-	-
		[redacted]	74.8	24.9	16	77	27.3	19.8	74.9	24.9	16.2	74.8	24.9	16
		[redacted]	0.1	0.3	0.3	3	9.5	24.3	0.2	0.2	1.5	0.1	0.5	0.7

Example #2 (cont'd)

Table [redacted] Mean Absolute Percentage Error in Estimated β [redacted]

[redacted]	[redacted]	[redacted]b.			[redacted]			S[redacted]		
[redacted]	[redacted]	$\beta =$ -0.01	-0.05	-0.10	-0.01	-0.05	-0.10	-0.01	-0.05	-0.10
[redacted]	[redacted]	0.1	0.1	0.1	1.2	5.9	8.5	0.3	0.1	0.0
		0.1	0.1	0.1	1.2	6.5	13.7	1.2	1.0	1.0
		0.1	0.1	0.0	1.4	6.8	15.8	0.2	1.5	1.6
[redacted]	[redacted]	0.2	0.1	0.1	1.8	5.4	13	0.3	0.3	0.3
		0.3	0.1	0.1	1.6	8.2	17.9	0.6	0.5	1.8
		0.1	0.1	0.2	1.3	8.2	17.3	0.5	0.7	0.6
[redacted]	[redacted]	0.1	0.1	0.1	1.4	7.0	13.6	0.7	4.0	5.1
		0.2	0.1	0.2	1.9	7.4	14.9	0.5	2.4	4.1
		0.1	0.1	0.1	1.2	9.0	19.6	1.5	3.0	3.5
[redacted]	[redacted]	0.1	0.1	0.1	0.4	6.1	8.6	0.6	2.4	2.7
		0.1	0.1	0.2	0.9	8.1	15.4	1.0	0.5	2.6
		0.1	0.1	0.1	1.5	9.4	18.1	0.6	1.0	1.2

[redacted] Mean Absolute Percentage Error in Estimated [redacted]

[redacted]	[redacted]	[redacted]b.			[redacted]			[redacted]		
[redacted]	[redacted]	$\beta =$ -0.01	-0.05	-0.10	-0.01	-0.05	-0.10	-0.01	-0.05	-0.10
[redacted]	[redacted]	0.1	0.1	0.1	1.3	6.5	10.9	0.3	0.3	1.0
		0.1	0.1	0.1	1.5	7.2	16.1	1.5	1.1	1.0
		0.2	0.1	0.1	1.1	7.9	20.8	0.1	1.7	2.5
[redacted]	[redacted]	0.2	0.0	0.1	1.8	7.5	12.4	0.3	0.8	0.5
		0.1	0.3	0.4	1.4	9.1	16.9	0.4	0.1	1.1
		0.1	0.1	0.2	1.3	8.2	17.3	0.5	0.7	0.6
[redacted]	[redacted]	0.1	0.2	0.1	1.6	8.3	15.5	0.9	4.7	5.4
		0.2	0.2	0.3	2.0	8.5	17.7	0.5	3.0	5.0
		0.4	0.1	0.2	0.9	10.5	25.4	1.2	3.3	4.4
[redacted]	[redacted]	0.5	0.2	0.1	0.5	4.7	7.0	1.0	2.7	2.7
		0.2	0.1	0.1	1.0	10.0	19.6	0.9	0.6	3.2
		0.3	0.1	0.3	1.3	10.8	23.9	0.4	0.7	1.4