Lecture 4
Testing in the Classical Linear Model

1

Hypothesis Testing: Brief Review

- In general, there are two kinds of hypotheses:
  (1) About the form of the probability distribution
  Example: Is the random variable normally distributed?

  (2) About the parameters of a distribution function
  Example: Is the mean of a distribution equal to 0?

- The second class is the traditional material of econometrics. We may test whether the effect of income on consumption is greater than one, or whether there is a size effect on the CAPM –i.e., the size coefficient on a CAPM regression is equal to zero.

2

Hypothesis Testing: Brief Review

• Some history:

- The modern theory of testing hypotheses begins with the Student's t-test in 1908.

- Fisher (1925) expands the applicability of the t-test (to the two-sample problem and the testing of regression coefficients). He generalizes it to an ANOVA setting. He pushes the 5% as the standard significance level.

- Neyman and Pearson (1928, 1933) consider the question: why these tests and not others? Or, alternatively, what is an optimal test? N&P's propose a testing procedure as an answer: the "best test" is the one that minimizes the probability of false acceptance (Type II Error) subject to a bound on the probability of false rejection (Type I Error).

- Fisher's and N&P's testing approaches can produce different results.

3

Hypothesis Testing: Brief Review

• We compare two competing hypothesis:

1) The null hypothesis, H_0, is the maintained hypothesis.

2) The alternative hypothesis, H_1, which we consider if H_0 is rejected.

• There are two types of hypothesis regarding parameters:

(1) A simple hypothesis. Under this scenario, we test the value of a parameter against a single alternative.

Example: H_0:$\theta=\theta_0$ against H_1:$\theta=\theta_1$.

(2) A composite hypothesis. Under this scenario, we test whether the effect of income on consumption is greater than one. Implicit in this test is several alternative values.

Example: H_0:$\theta>\theta_0$ against H_1:$\theta<\theta_1$.

4

Hypothesis Testing: Brief Review

• We compare two competing hypothesis: H_0 vs. H_1.

• Suppose the two hypothesis partition the universe: H_1 = Not H_0.

• Then, we collect a sample of data $X = \{X_1, \dots, X_n\}$ and device a decision rule, based on a statistic $T(X)$:

$T(X) \notin R$ ⇒ Reject H_0 (& we learn H_0! is not true).

$T(X) \in R$ ⇒ Fail to reject H_0. (No learning.)

The set R is called the *region of rejection* or the *critical region* of the test. We only, we only learn when $T(X)$ falls in this region –i.e., rejecting H_0:

"There are two possible outcomes: if the result confirms the hypothesis, then you've made a *measurement*. If the result is contrary to the hypothesis, then you've made a *discovery*." Enrico Fermi (Italy)

5

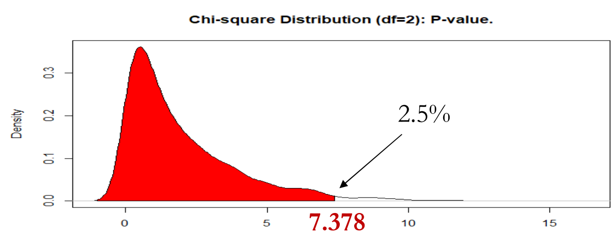
Hypothesis Testing: Brief Review - Fisher

• In this context, Fisher popularized a testing procedure known as *significance testing*. It relies on the *p-value*:

p-value is the probability of observing a result at least as extreme as the test statistic, under H_0.

Example: Suppose $T(X) \sim \chi_2^2$. We compute $\widehat{T(X)} = \mathbf{7.378}$. Then,

p-value($\widehat{T(X)}$ = **7.378**) = 1 – Prob[$T(X)$ < **7.378**] = 0.025



6

Hypothesis Testing: Brief Review - Fisher

• Fisher's Idea.

Steps for testing

1. Form H_0 and set a *significance level,* α.

2.a Collect a sample of data $X = \{X_1, \ldots, X_n\}$.

2.b Compute the test-statistics $T(X)$ used to test H_0.

3. Report the *p-value* -i.e., the probability, of observing a result at least as extreme as the test statistic, under H_0.

4. Decision rule:

If the *p-value* < α ⇒ result is *significant* & H_0 is rejected.

If the *p-value* > α ⇒ result is "not significant." No conclusions are reached. Gather more data/modify model.

7

Note: By setting α, we determine R.

Hypothesis Testing

Example: We want to test if the average net listings per year on the Mexican Bolsa de Valores, μ_{NL}, is 15.

1. H_0: $\mu_{NL} = 15$ & set $\alpha = .05$.

2a. Get a sample: $\{X_{1962}, X_{1963}, \ldots, X_{N=2022}\}$, with N=61.

2b. We use $T(X) = \bar{X}$, which is unbiased, consistent, and, assuming X is normally distributed, we know its distribution, $\bar{X} \sim N(\mu, \sigma^2/N)$.

3. Compute $\bar{X} = 36$ & *p-value*($\bar{X}$=36) = .005.

4. Decision Rule: *p-value* < α ⇒ result is *significant* & H_0 is rejected.

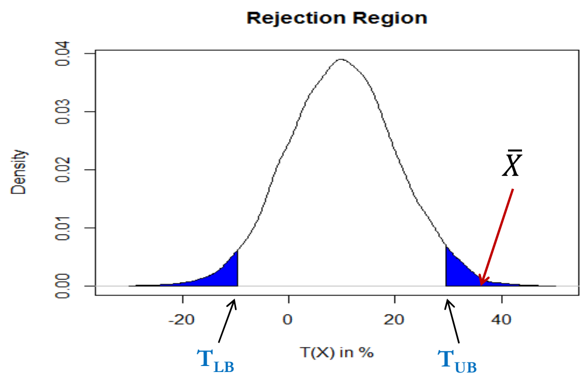
Instead of using a *p-value,* it is common to use a rejection region, R:

$$T(X) = \bar{X} \notin [T_{LB}, T_{UB}] \qquad \Rightarrow \text{Reject } H_0\text{: } \mu_{NL} = 15.$$

Hypothesis Testing

Example (continuation): That is,

$$R = [\bar{X} < T_{LB} \, , T_{UB} > \bar{X}]$$



• The blue area is the significance level, α.

9

Hypothesis Testing: Brief Review – N&P

• Under Fisher's testing procedure, declaring a result significant is subjective. Fisher pushed for a 5% (exogenous) significance level; but practical experience may play a role.

• Neyman and Pearson devised a different procedure, *hypothesis testing*, as a more objective alternative to Fisher's p-value.

Neyman's and Pearson's idea:

Consider two simple hypotheses (both with distributions). Calculate two probabilities and select the hypothesis associated with the higher probability (the hypothesis more likely to have generated the sample).

• Based on cost-benefit considerations, hypothesis testing determines the (fixed) rejection regions.

10

Hypothesis Testing: Brief Review – Summary

• The N&P's method always selects a hypothesis.

• There was a big debate between Fisher and N&P. In particular, Fisher believed that rigid rejection areas were not practical in science.

• Philosophical issues, like the difference between "inductive inference" (Fisher) and "inductive behavior" (N&P), clouded the debate.

• The dispute is unresolved. In practice, a hybrid of significance testing and hypothesis testing is used. Statisticians like the abstraction and elegance of the N&P's approach.

• Bayesian statistics using a different approach also assign probabilities to the various hypotheses considered.

11

Type I and Type II Errors

Definition: Type I and Type II errors

A *Type I error* is the error of rejecting H_0 when it is true. A *Type II error* is the error of "accepting" H_0 when it is false (that is when H_1 is true).

• <u>Notation</u>: Probability of Type I error: $\alpha = P[X \in R|H_0]$

Probability of Type II error: $\beta = P[X \in R^C|H_1]$

Definition: Power of the test

The probability of rejecting H_0 based on a test procedure is called the *power of the test.* It is a function of the value of the parameters tested, θ:

$$\pi = \pi(\theta) = P[X \in R].$$

<u>Note</u>: when $\theta \in H_1$ $\Rightarrow \pi(\theta) = 1 - \beta(\theta)$ -the usual application.

12

Type I and Type II Errors

• We want $\pi(\theta)$ to be near 0 for $\theta \in H_0$, and $\pi(\theta)$ to be near 1 for $\theta \in H_1$.

Definition: Level of significance

When $\theta \in H_0$, $\pi(\theta)$ gives you the probability of Type I error. This probability depends on θ. The maximum value of this when $\theta \in H_0$ is called *level of significance* of a test, denoted by α. Thus,

$$\alpha = \sup_{\theta \in H0} P[X \in R | H_0] = \sup_{\theta \in H0} \pi(\theta)$$

Define a *level α test* to be a test with $\sup_{\theta \in H0} \pi(\theta) \leq \alpha$.

Sometimes, $\alpha = P[X \in R | H_0]$ is called the *size* of a test.

Practical Note: Usually, the distribution of *T(X)* is known only approximately. In this case, we need to distinguish between the *nominal* α and the actual *rejection probability (empirical size)*. They may differ greatly

13

Type I and Type II Errors

	State of World	
Decision	H_0 true	H_1 true (H_0 false)
Cannot reject H_0	*Correct decision*	Type II error
Reject H_0	Type I error	*Correct decision*

← *Learning*

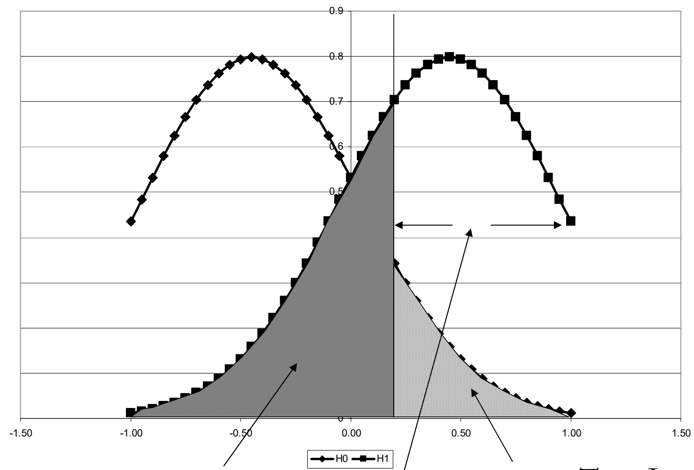
Need to control both types of error:

$\alpha = P(\text{rejecting } H_0 | H_0)$ <= Reject H_0 by "accident" or luck (a *false positive*).

$\beta = P(\text{not rejecting } H_0 | H_1)$ <= $1 - \beta$ = Power of test (under H_1).

14

Type I and Type II Errors



β = Type II error

α = Type I error

$\pi = 1 - \beta$ = Power of test (under H_1)

Note: Trade-off α & β.

15

Type I and Type II Errors - Example

- We conduct a 1,000 studies of some hypothesis (say, H_0: μ=0)
  - Use standard 5% significance level (**45** rejections under H_0).
  - Assume the proportion of false H_0 is 10% (**100** false cases).
  - Power 50% (50% correct rejections)

	State of World	
Decision	H_o true	H_1 true (H_o false)
Cannot reject H_o	855	50 *(Type II error)*
Reject H_o	**45** *(Type I error)*	50
	900	**100**

Note: Of the 95 studies which result in a "*statistically significant*" (i.e., $p<0.05$) result, **45** (**47.4%**) are true H_0 and so are "false positives."

16

Type I and Type II Errors: Example

- Now, with same proportion of false H_0 (**10%**) and same α = **5%**, assume the power is **80%** (80% correct rejections of H_0).

	State of World	
Decision	H_o true	H_1 true (H_o false)
Cannot reject H_o	855	20 *(Type II error)*
Reject H_o	**45** *(Type I error)*	**80**
	900	**100**

Now, of the 125 studies which result in a "*statistically significant*" (i.e., $p<0.05$) result, **45** (**36%**) are true H_0 and so are "false positives."

17

Type I and Type II Errors - Example

- Now, assume the power is 80% (80% correct rejections) and same α = **5%,** but the proportions of false H_0 is **50%** (**500** false cases).

	State of World	
Decision	H_o true	H_1 true (H_o false)
Cannot reject H_o	475	100 *(Type II error)*
Reject H_o	**25** *(Type I error)*	**400**
	500	**500**

Now, of the 425 studies which result in a "*statistically significant*" (i.e., $p<0.05$) result, **25** (**5.88%**) are true H_0 and so are "false positives."

Conclusion: The proportion of false positives depends on percentage of false H_0 and the power of test. Higher power, lower proportion.

18

Type I and Type II Errors - Example

• For a given α (P), higher power, lower % of false-positives –i.e., more true learning.

Proportion of ideas that are correct (null hypothesis false)	Power of study	Percentage of "significant" results that are false-positives		
		P=0.05	P=0.01	P=0.001
80%	20%	5.9	1.2	0.1
	50%	2.4	0.5	0.0
	80%	1.5	0.3	0.0
50%	20%	20.0	4.8	0.5
	50%	9.1	2.0	0.2
	80%	5.9	1.2	0.1
10%	20%	69.2	31.0	4.3
	50%	47.4	15.3	1.8
	80%	36.0	10.1	1.1
1%	20%	96.1	83.2	33.1
	50%	90.8	66.4	16.5
	80%	86.1	55.3	11.0

19

More Powerful Test

Definition: More Powerful Test

Let (α_1, β_1) and (α_2, β_2) be the characteristics of two tests. The first test is *more powerful* (better) than the second test if $\alpha_1 \leq \alpha_2$, and $\beta_1 \leq \beta_2$ with a strict inequality holding for at least one point.

Note: If we cannot determine that one test is better by the definition, we could consider the relative cost of each type of error. Classical statisticians typically do not consider the relative cost of the two errors because of the subjective nature of this comparison.

Bayesian statisticians compare the relative cost of the two errors using a loss function.

20

Most Powerful Test

Definition: Most powerful test of size α

R is the *most powerful test of size* α if $\alpha(R)=\alpha$ and for any test R_1 of size α, $\beta(R) \leq \beta(R_1)$.

Definition: Most powerful test of level α

R is the *most powerful test of level* α (that is, such that $\alpha(R) \leq \alpha$) and for any test R_1 of level α (that is, $\alpha(R_1) \leq \alpha$), if $\beta(R) \leq \beta(R_1)$.

21

UMP Test

Definition: Uniformly most powerful (UMP) test

R is the *uniformly most powerful test of level* α (that is, such that $\alpha(R) \leq \alpha$) and for *every test* R_1 of level α (that is, $\alpha(R_1) \leq \alpha$), if $\pi(R) \geq \pi(R_1)$.

For every test: for alternative values of θ_1 in H_1:$\theta=\theta_1$.

- Choosing between admissible test statistics in the (α, β) plane is similar to the choice of a consumer choosing a consumption point in utility theory. Similarly, the tradeoff problem between α and β can be characterized as a ratio.

- This idea is the basis of the *Neyman-Pearson Lemma* to construct a test of a hypothesis about θ: H_0:$\theta=\theta_0$ against H_1:$\theta=\theta_1$.

22

Neyman-Pearson Lemma

• Neyman-Pearson Lemma provides a procedure for selecting the best test of a simple hypothesis about θ: H_0:θ=θ_0 against H_1:θ=θ_1.

• Let $L(x|\theta)$ be the joint density function of X. We determine R based on the ratio $L(x|\theta_1)/L(x|\theta_0)$. (This ratio is called the *likelihood ratio*.) The bigger this ratio, the more likely the rejection of H_0.

• That is, the Neyman-Pearson lemma of hypothesis testing provides a good criterion for the selection of hypotheses: The ratio of their probabilities.

23

Neyman-Pearson Lemma

• Consider testing a simple hypothesis H_0: $\theta = \theta_0$ vs. H_1: $\theta = \theta_1$, where the pdf corresponding to θ_i is $L(\boldsymbol{x}|\theta_i)$, i=0,1, using a test with rejection region R that satisfies

(1) $\boldsymbol{x} \in R$ if $L(\boldsymbol{x}|\theta_1) > k\, L(\boldsymbol{x}|\theta_0)$

$\boldsymbol{x} \in R^c$ if $L(\boldsymbol{x}|\theta_1) < k\, L(\boldsymbol{x}|\theta_0)$,

for some $k \geq 0$, and

(2) $\alpha = P[X \in R|H_0]$

Then,

(a) Any test that satisfies (1) and (2) is a UMP level α test.

(b) If there exists a test satisfying (1) and (2) with $k > 0$, then every UMP level α test satisfies (2) and every UMP level α test satisfies (1) except perhaps on a set A satisfying $P[X \in A|H_0] = P[X \in A|H_1] = 0$.

24

Monotone Likelihood Ratio

• In general, we have no basis to pick θ_1. We need a procedure to test composite hypothesis, preferably with a UMP.

Definition: Monotone Likelihood Ratio

The model $f(X,\theta)$ has the *monotone likelihood ratio property in* $u(X)$ if there exists a real valued function $u(X)$ such that the likelihood ratio $\lambda = L(x|\theta_1)/L(x|\theta_0)$ is a non-decreasing function of $u(X)$ for each choice of θ_1 and θ_0, with $\theta_1 > \theta_0$.

If $L(x|\theta_1)$ satisfies the MLRP with respect to $L(x|\theta_0)$ the higher the observed value $u(X)$, the more likely it was drawn from distribution $L(x|\theta_1)$ rather than $L(x|\theta_0)$.

Note: In general, we think of $u(X)$ as a statistic.

25

Monotone Likelihood Ratio

• Under the MLRP there is a relationship between the magnitude of some observed variable, say $u(X)$, and the distribution it draws from it.

• Consider the exponential family:

$$L(X;\theta) = \exp\{\Sigma_i U(X_i) - A(\theta)\ \Sigma_i T(X_i) + n\ B(\theta)\}.$$

Then, $\ln \lambda = \Sigma_i T(X_i)\ [A(\theta_1) - A(\theta_0)] + nB(\theta_1) - nB(\theta_0).$

Let $u(X)=\Sigma_i T(X_i)$.

$$\Rightarrow \quad \delta \ln \lambda / \delta u = [A(\theta_1) - A(\theta_0)] > 0, \text{ if A(.) is monotonic in } \theta.$$

In addition, $u(X)$ is a sufficient statistic..

• Some distributions with MLRP in $T(X) = \Sigma_i x_i$: normal (with σ known), exponential, binomial, Poisson.

26

Karlin-Rubin Theorem

Theorem: Karlin-Rubin (KR) Theorem

Suppose we are testing $H_0: \theta \leq \theta_0$ vs. $H_1: \theta > \theta_0$.

Let T(*X*) be a sufficient statistic, and the family of distributions *g*(.) has the MLRP in T(*X*).

Then, for any t_0 the test with rejection region $T > t_0$ is UMP level α, where $\alpha = \Pr(T > t_0 | \theta_0)$.

27

KR Theorem: Practical Use

Goal: Find the UMP level α test of $H_0: \theta \leq \theta_0$ vs. $H_1: \theta > \theta_0$ (similar for $H_0: \theta \geq \theta_0$ vs. $H_1: \theta < \theta_0$)

1. If possible, find a univariate sufficient statistic T(*X*). Verify its density has an MLR (might be non-decreasing or non-increasing, just show it is monotonic).
2. KR states the UMP level α test is either 1) reject if $T > t_0$ or 2) reject if $T < t_0$. Which way depends on the direction of the MLR and the direction of H_1.
3. Derive E[T] as a function of θ. Choose the direction to reject ($T > t_0$ or $T < t_0$) based on whether E[T] is higher or lower for θ in H_1. If E[T] is higher for values in H_1, reject when $T > t_0$, otherwise reject for $T < t_0$.

28

KR Theorem: Practical Use

4. t_0 is the appropriate percentile of the distribution of T when $\theta=\theta_0$. This percentile is either the α percentile (if you reject for $T<t_0$) or the 1 – α percentile (if you reject for $T>t_0$).

29

Nonexistence of UMP tests

• For most two-sided hypotheses –i.e., H_0: $\theta = \theta_0$ vs. H_1: $\theta \neq \theta_0$–, no UMP level test exists.

Simple intuition: The test which is UMP for $\theta < \theta_0$ is not the same as the test which is UMP for $\theta > \theta_0$. A UMP test must be most powerful across *every* value in H_1.

Definition: Unbiased Test

A test is said to be *unbiased* when

$\pi(\theta) \geq \alpha$ for all $\theta \in H_1$

and P[Type I error]: $P[X \in R | H_0] = \pi(\theta) \leq \alpha$ for all $\theta \in H_0$.

Unbiased test $\Rightarrow \pi(\theta_0) < \pi(\theta_1)$ for all θ_0 in H_0 and θ_1 in H_1.

Most two-sided tests we use are UMP level α *unbiased* (UMPU) tests.

30

Some problems left for students

• So far, we have produced UMP level α tests for simple versus simple hypotheses (H_0:$\theta = \theta_0$ vs. H_1:$\theta = \theta_1$) and one sided tests with MLRP (H_0:$\theta \leq \theta_0$ vs. H_1:$\theta > \theta_0$).

• There are a lot of unsolved problems. In particular,

(1) We did not cover unbiased tests in detail, but they are often simply combinations of the UMP tests in each directions

(2) Karlin-Rubin discussed univariate sufficient statistics, which leaves out every problem with more than one parameter (for example testing the equality of means from two populations).

(3) Every problem without an MLRP is left out.

31

No UMP test

• Power function (again)

We define the power function as $\pi(\theta) = P[X \in R]$. Ideally, we want $\pi(\theta)$ to be near 0 for $\theta \in H_0$, and $\pi(\theta)$ to be near 1 for $\theta \in H_1$.

The classical (frequentist) approach is to look in the class of all level α tests (all tests with $\sup_{\theta \in H0} \pi(\theta) \leq \alpha$) and find the MP one available.

• In some cases there is a UMP level α test, as given by the Neyman Pearson Lemma (simple hypotheses) and the Karlin Rubin Theorem (one sided alternatives with univariate sufficient statistics with MLRP). But, in many cases, there is no UMP test.

• When no UMP test exists, we turn to general methods that produce good tests.

32

No UMP test

• Power is a function of three factors ($\theta - \theta_0$, n, & α):

– **Effect size**: True value (θ) – Hypothesized value. (Say, $\theta - \theta_0$). Bigger deviations from H_0 are easier to detect.

– **Sample size**: n. Higher n, smaller sampling error. Sampling distributions are more concentrated!

– **Statistical significance** –i.e., the α.

Example: We randomly collect 20 stock returns ($n = 20$), which are assumed $N(\theta, 0.2^2)$ (known σ^2 for simplicity). Set $\alpha = .05$. We want to test H_0: $\theta = \theta_0 = 0.1$ against H_1: $\theta > 0.1$.

Q: What is the power of the test if the true $\theta = 0.2$ (H_1: $\theta = 0.2$ is true)?

Test-statistitc: $z = (\bar{x} - \theta_0)/[\sigma/\text{sqrt}(n)]$.

Rejection rule: $z \geq z_{\alpha=.05} = 1.645$.

33

No UMP test

Example (continuation):

Test-statistic: z-*statistic* $= (\bar{x} - \theta_0)/[\sigma/\text{sqrt}(n)] = (\bar{x} - 0.1)/(.2/\text{sqrt}(20))$.

Rejection rule: $z \geq z_{\alpha=.05} = 1.645$, or, equivalently, when the observed $\bar{x} \geq .1736$ $[= z_{\alpha/2} * \sigma/\text{sqrt}(n) + \theta_0 = 1.645 * .2/\text{sqrt}(20) + .1]$

$$\Rightarrow \text{Power} = P[X \in R | H_1] = P[\bar{x} \geq .1736 \mid \theta = 0.2]$$
$$= P[z \geq (.1736 - 0.2)/(.2/\text{sqrt}(20))]$$
$$= P[z \geq -.591]$$
$$= 1 - P[z < -.591] = 0.722760$$

• Changing $\theta - \theta_0$

If (H_1: $\theta = 0.3$ **is true**)?, then the power of the test (under H_1):

$$\Rightarrow \text{Power} = P[X \in R | H_1] = P[z \geq (.1736 - 0.3)/(.2/\text{sqrt}(20))]$$
$$= P[z \geq -2.82713] = 0.997652$$

34

No UMP test

Example (continuation):

• Changing α ($\theta_1 = 0.2$; $n = 20$)

If $\alpha = .01$, then rejection rule: $z \geq z_{\alpha/2=.005} = 2.33$.

Or equivalently: $\bar{x} \geq 0.2042$ [$= 2.33$ *.2/sqrt(20) + 0.1]

$\Rightarrow$ Power = $P[X \in R | H_1]$ = $P[\bar{x} \geq (0.2042 - 0.2)/(.2/\text{sqrt}(20))]$

= $P[z \geq 0.093915] = .46259$

• Changing n ($\theta_1 = 0.2$; $\alpha = .05$)

If $n = 200$, then rejection rule: $\bar{x} \geq .12332$ [$= 1.645$ *.2/sqrt(200) + 0.1]

$\Rightarrow$ Power = $P[X \in R | H_1]$ = $P[\bar{x} \geq (.12323 - 0.2)/(.2/\text{sqrt}(200))]$

= $P[z \geq -5.4261] = .9999999$

Note: We can select n to achieve a given power (for given θ_1 & α). Say, set $n = 34$ to set $P[X \in R | H_1] = .90$.

35

General Methods

• Likelihood Ratio (LR) Tests

• Bayesian Tests - can be examined for their frequentist properties even if you are not a Bayesian.

• Pivot Tests - Tests based on a function of the parameter and data whose distribution does not depend on unknown parameters. Wald and Score tests are examples:

- Wald Tests - Based on the asymptotic normality of the MLE.
- Score Tests - Based on the asymptotic normality of the log-likelihood.

36

Likelihood Ratio Tests

• Define the likelihood ratio (LR) statistic

$$\lambda(X) = \sup_{\theta \in H0} L(X|\theta) / \sup_{\theta} L(X|\theta)$$

Note:
Numerator: maximum of the LF within H_0
Denominator: maximum of the LF within the entire parameter space, which occurs at the MLE.

• Reject H_0 if $\lambda(X) < k$, where k is determined by

$$\text{Prob}[0 < \lambda(X) < k \mid \theta \in H_0] = \alpha.$$

37

Properties of the LR statistic $\lambda(X)$

• Properties of $\lambda(X) = \sup_{\theta \in H0} L(X|\theta) / \sup_{\theta} L(X|\theta)$
(1) $0 \leq \lambda(X) \leq 1$, with $\lambda(X) = 1$ if the supremum of the likelihood occurs within H_0.

Intuition of test: If the likelihood is much larger outside H_0 –i.e., in the unrestricted space–, then $\lambda(X)$ will be small and H_0 should be rejected.

(2) Under general assumptions, $-2 \ln \lambda(X) \sim \chi_p^2$, where p is the difference in *df* between the H_0 and the general parameter space.

(3) For simple hypotheses, the numerator and denominator of the LR test are simply the likelihoods under H_0 and H_1. The LR test reduces to a test specified by the NP Lemma.

38

Likelihood Ratio Tests: Example I

Example: λ(*X*) for a $\boldsymbol{X} \sim N(\theta,\sigma^2)$ for H_0: $\theta = \theta_0$ vs. H_1: $\theta \neq \theta_0$. Assume σ^2 is known.

$$\lambda(\mathrm{x}) = \frac{\mathrm{L}(\hat{\theta}_0 \,|\, \mathrm{x})}{\mathrm{L}(\bar{x} \,|\, \mathrm{x})} = \frac{(2\pi)^{-n/2} e^{-\sum_{i=1}^{n}(x_i-\theta_0)^2/2\sigma^2}}{(2\pi)^{-n/2} e^{-\sum_{i=1}^{n}(x_i-\bar{x})^2/2\sigma^2}} = \mathrm{e}^{\frac{-\sum_{i=1}^{n}(x_i-\theta_0)^2+\sum_{i=1}^{n}(x_i-\bar{x})^2}{2\sigma^2}} = \mathrm{e}^{\frac{-n(\bar{x}-\theta_0)^2}{2\sigma^2}}$$

$$\text{Reject } H_0 \text{ if } \quad \lambda(x) < k \quad \Rightarrow \ln\lambda(\mathrm{x}) = \frac{-n(\bar{x}-\theta_0)^2}{2\sigma^2} < \ln k \quad \Rightarrow \frac{(\bar{x}-\theta_0)^2}{\sigma^2/n} > -2\ln k$$

<u>Note</u>: Finding *k* is not needed.
Why? We know the left hand side is distributed as a χ_p^2, thus (-2 ln *k)* needs to be the 1 – α percentile of a χ_p^2. We need not solve explicitly for *k*, we just need the rejection rule.

Likelihood Ratio Tests: Example II

Example: λ(*X*) for a *X* ~ *exponential* (λ) for H_0: $\lambda = \lambda_0$ vs. H_1: $\lambda \neq \lambda_0$.

$L(X|\theta) = \lambda^n \exp(-\lambda\, \Sigma_i\, x_i) = \lambda^n \exp(-\lambda n \bar{x}) \qquad \Rightarrow \lambda_{MLE} = 1/\bar{x}$

$$\lambda(\mathrm{x}) = \frac{\lambda_0{}^n e^{-\lambda_0 n \bar{x}}}{(1/\bar{x})^n e^{-n}} = (\bar{x}\lambda_0)^n e^{\{n(1-\lambda_0\bar{x})\}}$$

$$\text{Reject } H_0 \text{ if } \quad \lambda(x) < k \quad \Rightarrow \ln\lambda(\mathrm{x}) = n\ln(\bar{x}\lambda_0) + n(1-\lambda_0\bar{x}) < \ln k$$

We need to find *k* such that P[λ(*X*)<*k*] = α. Unfortunately, this is not analytically feasible. We know the distribution of $\bar{x}$ is Gamma(*n*; λ/*n*), but we cannot get further.

It is, however, possible to determine the cutoff point, *k*, by simulation (set *n*, λ_0).

Testing in Economics

"The three golden rules of econometrics are test, test and test." David Hendry (1944, England)

"The only relevant test of the validity of a hypothesis is comparison of prediction with experience." Milton Friedman (1912-2006, USA)

41

Hypothesis Testing: Summary

• Hypothesis testing:

(1) We need a model. For example, $\mathbf{y} = f(\mathbf{X}, \theta) + \boldsymbol{\varepsilon}$

(2) We gather data ($\mathbf{y}$, $\mathbf{X}$) and estimate the model $\Rightarrow$ we get $\hat{\theta}$

(3) We formulate a hypotheses. For example, H_0: $\theta=\theta_0$ vs. H_1:$\theta\neq\theta_0$

(4) Find an appropriate test and know its distribution under H_0

(5) Decision Rule (Test H_0). Reject H_0: if θ_0 is too far from $\hat{\theta}$ ("the hypothesis is *inconsistent* with the sample evidence.")

The decision rule will be based on a statistic, $T(X)$. If the statistic is large, then, we reject H_0.

• To determine if the statistic is "large," we need a *null distribution*.

• Ideally, we use a test that is most powerful to test H_0.

42

Hypothesis Testing: Issues

• Logic of the Neyman-Pearson methodology:

If H_0 is true, then $T(X)$ will have a certain distribution (under H_0). We call this distribution *null distribution* or *distribution under the null.*

• It tells us how likely certain values are, if H_0 is true. Thus, we expect 'large values' for θ_0 to be unlikely.

• Decision rule.

If the observed value for $T(X)$ falls in rejection region R

$\Rightarrow$ Assumed distribution must be incorrect: H_0 should be rejected.

43

Hypothesis Testing: Issues

• Issues:

– What happens if the model is wrong?

– What is a testable hypothesis?

– Nested vs. Non-nested models

– Methodological issues

– Classical (frequentist approach): Are the data consistent with H_0?

– Bayesian approach: How do the data affect our prior odds? Use the posterior odds ratio.

44

Review – Testing in the CLM: Single Parameter

• We test a hypotheses about a single parameter, say β_k, of the DGP.

Example: The linear model (DGP): $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

1. Formulate H_0: $\mathbf{X}_k$ should not be in the DGP $\Rightarrow$ $H_0: \beta_k = \beta_k^0$
$H_1: \beta_k \neq \beta_k^0$.

2. Construct $T(X)$ test H_0: $t_k = (b_k - \beta_k^0)/\text{sqrt}\{s^2(\mathbf{X}'\mathbf{X})_{kk}^{-1}\}$
Distribution of $T(X)$ under H_0, with s^2 estimating σ^2 (unknown):
If (**A5**) $\boldsymbol{\varepsilon}|\mathbf{X} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_T)$, $\Rightarrow t_k \sim t_{T-k}$.
If (**A5**) not true, asymptotic results: $\Rightarrow t_k \xrightarrow{d} N(0, 1)$.

3. Using OLS, we estimate $b_1, b_2, \ldots, b_k, \ldots$,

4. <u>Decision Rule</u>: Set α level. We reject H_0 if p-value(t_k) < α.
Or, reject H_0, if $|t_k| > t_{T-k,\,\alpha/2}$.

45

Review – Testing in the CLM: *t-value*

• Special case: $H_0: \beta_k = 0$
$H_1: \beta_k \neq 0$.

Then,

$$t_k = b_k/\text{sqrt}\{s^2(\mathbf{X}'\mathbf{X})_{kk}^{-1}\} = b_k/\text{SE}[b_k] \Rightarrow t_k \sim t_{T-k}.$$

This special case of t_k is called the *t-value*. That is, the t-value is the ratio of the estimated coefficient and its SE.

• The t-value is routinely reported in all regression packages. In the lm() function, it is reported in the third row of numbers.

• Usually, α = 5%, then if $|t_k| > 1.96 \approx 2$, we say the coefficient b_k is "*significant*."

46

Hypothesis Testing: Confidence Intervals

• The OLS estimate **b** is a point estimate for **β**, meaning that **b** is a single value in R^k.

• Broader concept: Estimate a set C_n, a collection of values in R^k.

• When the parameter is real-valued, it is common to focus on intervals $C_n = [L_n; U_n]$, called an *interval estimate* for θ. The goal of C_n is to contain the true value, e.g. $\theta \in C_n$, with high probability.

• C_n is a function of the data. Therefore, it is a RV.

• The coverage probability of the interval $C_n = [L_n; U_n]$ is Prob$[\theta \in C_n]$.

47

Hypothesis Testing: Confidence Intervals

• The randomness comes from C_n, since θ is treated as fixed.

• Interval estimates C_n are called *confidence intervals* (C.I.) as the goal is to set the coverage probability to equal a pre-specified target, usually 90% or 95%. C_n is called a $(1 - \alpha)$% C.I.

• When we know the distribution for the point estimate, it is easy to construct a C.I. For example, under (**A5**), the distribution of **b** is normal, then a 95% C.I. is given by:

$$C_n = [\, b_k - z_{.025} \times \text{Estimated SE}(b_k),\ b_k + z_{.025} \times \text{Estimated SE}(b_k)]$$

• This C.I. is symmetric around b_k. Its length is proportional to the SE(b_k).

48

Hypothesis Testing: Confidence Intervals

• Equivalently, C_n is the set of parameter values for b_k such that the z-statistic $z_n(b_k)$ is smaller (in absolute value) than $z_{\alpha/2}$. That is,

$C_n = \{b_k: | z_n(b_k) | \leq z_{\alpha/2}\}$ with coverage probability (1 - α)%.

• In general, the coverage probability of C.I.'s is unknown, since we do not know the distribution of the point estimates.

• In Lecture 8, we will use asymptotic distributions to approximate the unknown distributions. We will use these asymptotic distributions to get asymptotic coverage probabilities.

• Summary: C.I.'s are a simple but effective tool to assess estimation uncertainty.

49

Recall: A *t*-distributed variable

• Recall a t_ν-distributed variable is a ratio of two independent RV: a N(0,1) RV and the square root of a χ_ν^2 RV divided by ν.

Let $$z = \frac{(\bar{x} - \mu)}{\sigma / \sqrt{n}} = \sqrt{n}\,\frac{(\bar{x} - \mu)}{\sigma} \sim N(0,1)$$

Let $$U = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

Assume that Z and U are independent (check the middle matrices in the quadratic forms!). Then,

$$t = \frac{\sqrt{n}\,\frac{(\bar{x} - \mu)}{\sigma}}{\sqrt{\frac{(n-1)s^2}{\sigma^2}/(n-1)}} = \frac{\sqrt{n}\,(\bar{x} - \mu)}{s} = \frac{(\bar{x} - \mu)}{s / \sqrt{n}} \sim t_{n-1}$$

50

Hypothesis Testing: Testing Example in R

Example: 3 Factor Fama-French Model (continuation) for IBM:

```
Returns <- read.csv("http://www.bauer.uh.edu/rsusmel/phd/K-DIS-IBM.csv", head=TRUE, sep=",")

b <- solve(t(x)%*% x)%*% t(x)%*%y                  # b = (X'X)^-1 X' y  (OLS regression)
e <- y - x%*%b                                     # regression residuals, e
RSS <- as.numeric(t(e)%*%e)                        # RSS
R2 <- 1 - as.numeric(RSS)/as.numeric(t(y)%*%y)     # R-squared
Sigma2 <- as.numeric(RSS/(T-k))                    # Estimated σ^2 = s^2
SE_reg <- sqrt(Sigma2)                             # Estimated σ – Regression stand error
Var_b <- Sigma2*solve(t(x)%*% x)                   # Estimated Var[b|X] = s^2 (X'X)^-1
SE_b <- sqrt(diag(Var_b))                          # SE[b|X]
t_b <- b/SE_b                                      # t-stats (See Chapter 4)
```


OLS Estimation – Is IBM's Beta equal to 1?

```
> t(b)
                 Mkt_RF        SMB        HML
[1,] -0.005088944 0.9082989 -0.2124596 -0.1715002
> t(SE_b)
                 Mkt_RF        SMB        HML
[1,] 0.002487509 0.05672206 0.08411188 0.08468165
> t(t_b)
              Mkt_RF      SMB       HML
[1,] -2.045799 16.01315 -2.525917 -2.025235
```

$\Rightarrow$ all coefficients are *significant* ($|t|>2$).

- Q: Is the market beta (β_1) equal to 1? That is,

H_0: $\beta_1 = 1$ vs. H_1: $\beta_1 \neq 1$

$\Rightarrow t_k = (b_k - \beta_k^{0})/\text{Est. SE}(b_k)$

$t_1 = (\mathbf{0.9082989} - 1)/\ \mathbf{0.05672206} = -1.616674$

$\Rightarrow |t_1| < 1.96$ $\Rightarrow$ Cannot reject H_0 at 5% level

Testing: The Expectation Hypothesis (EH)

Example: EH states that forward/futures prices are good predictors of future spot rates: $E_t[S_{t+T}] = F_{t,T}$.

Implication of EH: $S_{t+T} - F_{t,T}$ = unpredictable.

That is, $E_t[S_{t+T} - F_{t,T}] = E_t[\varepsilon_t] = 0$!

Empirical tests of the EH are based on a regression:

$$(S_{t+T} - F_{t,T})/S_t = \alpha + \beta\, Z_t + \varepsilon_t, \qquad \text{(where } E[\varepsilon_t]=0\text{)}$$

where Z_t represents any economic variable that might have power to explain S_t, for example, $(i_d - i_f)$.

Then, under EH, H_0: $\alpha = 0$ and $\beta = 0$.
vs H_1: $\alpha \neq 0$ and/or $\beta \neq 0$.

53

Testing: The Expectation Hypothesis (EH)

Example (continuation): We will informally test EH using exchange rates (USD/GBP), 3-mo forward rates and 3-mo interest rates.

```
SF_da <- read.csv("http://www.bauer.uh.edu/rsusmel/4397/SpFor_prices.csv",
head=TRUE, sep=",")
summary(SF_da)
x_date <- SF_da$Date
x_S <- SF_da$GBPSP
x_F3m <- SF_da$GBP3M
i_us3 <- SF_da$Dep_USD3M
i_uk3 <- SF_da$Dep_UKP3M
T <- length(x_S)
prem <- (x_S[-1] - x_F3m[-T])/x_S[-1]
int_dif <- (i_us3 - i_uk3)/100
y <- prem
x <- int_dif[-T]
fit <- lm( y ~ x)
```

54

Testing: The Expectation Hypothesis (EH)

Example (continuation): We do two individual t-tests on α & β.

```
> summary(fit)
Call:
lm(formula = y ~ x)

Residuals:
 Min        1Q    Median      3Q       Max
-0.125672 -0.014576 -0.000439  0.017356  0.094283

Coefficients:
               Estimate   Std. Error  t value Pr(>|t|)
(Intercept)  -0.0001854  0.0016219  -0.114  0.90906      ⇒ constant not significant (|t|<2)
x            -0.2157540  0.0731553  -2.949  0.00339  ** ⇒ slope is significant (|t|>2). ⇒ Reject H0
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02661 on 361 degrees of freedom
Multiple R-squared:  0.02353,   Adjusted R-squared:  0.02082
F-statistic: 8.698 on 1 and 361 DF,  p-value: 0.003393
```

55

Testing: The Expectation Hypothesis (EH)

- 95% C.I. for b:

$C_n = [\, b_k - t_{k,\alpha/2} * \text{Estimated SE}(b_k),\ b_k + t_{k,\alpha/2} * \text{Estimated SE}(b_k)]$

Then,

$C_n = [\mathbf{-0.215754} - 1.96 * \mathbf{0.0731553}, \mathbf{-0.215754} + 1.96 * \mathbf{0.0731553}]$

$= [\mathbf{-0.3591384}, -0.07236961]$

Since $\beta = 0$ is not in C_n with 95% confidence ⇒ Reject H_0: $\beta_1 = 0$ at 5% level.

Note: The EH is a joint hypothesis, it should be tested with a joint test!

56

Testing a Hypothesis: Wald Statistic

• Most of our test statistics, including joint tests, are Wald statistics.

Wald = normalized distance measure:

One parameter: $t_b = (b_k - \beta^0_k)/s_{b,k}$ = distance/unit

More than one parameter.

Let $\mathbf{z}$ = (random vector – hypothesized value) be the distance

$W = \mathbf{z}' [\text{Var}(\mathbf{z})]^{-1} \mathbf{z}$ (a quadratic form)

• Distribution of W? We have a quadratic form.

– If $\mathbf{z}$ is normal and σ^2 known, $W \sim \chi^2_{\text{rank}(\text{Var}(\mathbf{z}))}$

– If $\mathbf{z}$ is normal and σ^2 unknown, $W \sim F$

– If $\mathbf{z}$ is not normal and σ^2 unknown, we rely on asymptotic theory, $W \xrightarrow{d} \chi^2_{\text{rank}(\text{Var}(\mathbf{z}))}$

Abraham Wald (1902–1950, Hungary)

57

Testing a Hypothesis: Wald Statistic

• Distribution of W? We have a quadratic form.

Recall **Theorem 7.4.** Let the $N\times 1$ vector $y \sim N(\mu_y, \Sigma_y)$. Then,

$(y - \mu_y)' \Sigma_y^{-1} (y - \mu_y) \sim \chi^2_N$. –note: $N = \text{rank}(\Sigma_y)$.

$\Rightarrow$ If $\mathbf{z} \sim N(0,\text{Var}(\mathbf{z})) \Rightarrow W$ is distributed as $\chi^2_{\text{rank}(\text{Var}(\mathbf{z}))}$

In general, Var($\mathbf{z}$) is unknown, we need to use an estimator of Var($\mathbf{z}$). In our context, we need an estimator of σ^2. Suppose we use s^2. Then, we have the following result:

Let $\mathbf{z} \sim N(0,\text{Var}(\mathbf{z}))$. We use s^2 instead of σ^2 to estimate Var($\mathbf{z}$)

$\Rightarrow W \sim F$ distribution.

Recall the F distribution arises as the ratio of two χ^2 variables divided by their degrees of freedom.

58

Recall: An *F*-distributed variable

Let $F = \dfrac{\chi_J^2 / J}{\chi_T^2 / T} \sim F_{J,T}$

Let $z = \dfrac{(\bar{x} - \mu)}{\sigma / \sqrt{n}} = \sqrt{n}\,\dfrac{(\bar{x} - \mu)}{\sigma} \sim N(0,1)$

Let $U = \dfrac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$

If Z and U are independent, then

$$F = \frac{\left[\sqrt{n}\,\dfrac{(\bar{x} - \mu)}{\sigma}\right]^2 / 1}{\dfrac{(n-1)s^2}{\sigma^2} / (n-1)} = \frac{(\bar{x} - \mu)^2}{s^2 / n} \sim F_{1,n-1}$$

59

Recall: An *F*-distributed variable

• There is a relationship between *t* and *F* when testing one restriction.
- For a single restriction, *m* = r'b – *q*. The variance of *m* is: r Var[b] r.
- The distance measure is *t* = *m* / Est. SE(*m*) ~ t_{T-k}.
- This *t*-ratio is the sqrt{*F*-ratio}.

• *t*-ratios are used for individual restrictions, while *F*-ratios are used for joint tests of several restrictions.

60

The General Linear Hypothesis: H_0: $R\beta - q = 0$

• Suppose we are interested in testing J joint hypotheses.

Example: We want to test that in the 3 FF factor model that the SMB and HML factors have the same coefficients, $\beta_{SMB} = \beta_{HML} = \beta^0$.

We can write linear restrictions as H_0: $\mathbf{R}\boldsymbol{\beta} - \mathbf{q} = \mathbf{0}$, where $\mathbf{R}$ is a Jxk matrix and $\mathbf{q}$ a Jx1 vector.

In the above example (J=2), we write:

$$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} \beta_1 \\ \beta_{Mkt} \\ \beta_{SMB} \\ \beta_{HML} \end{bmatrix} = \begin{bmatrix} \beta^0 \\ \beta^0 \end{bmatrix}$$

61

The General Linear Hypothesis: H_0: $R\beta - q = 0$

• Q: Is $\mathbf{Rb} - \mathbf{q}$ close to $\mathbf{0}$? There are two different approaches to this questions. Both have in common the property of unbiasedness for $\mathbf{b}$.

(1) We base the answer on the discrepancy vector:

$$\mathbf{m} = \mathbf{Rb} - \mathbf{q}.$$

Then, we construct a Wald statistic:

$$W = \mathbf{m}' \,(\text{Var}[\mathbf{m}\,|\,\mathbf{X}])^{-1}\, \mathbf{m}$$

to test if $\mathbf{m}$ is different from 0.

(2) We base the answer on a model loss of fit when restrictions are imposed: RSS must increase and R^2 must go down. Then, we construct an F test to check if the unrestricted RSS (RSS_U) is different from the restricted RSS (RSS_R).

62

The General Linear Hypothesis: H_0: $\mathbf{R\beta} - \mathbf{q} = \mathbf{0}$

- Q: Is $\mathbf{Rb} - \mathbf{q}$ close to $\mathbf{0}$? There are two different approaches to this questions. Both have in common the property of unbiasedness for $\mathbf{b}$.

(1) We base the answer on the discrepancy vector:

$$\mathbf{m} = \mathbf{Rb} - \mathbf{q}.$$

Then, we construct a Wald statistic:

$$W = \mathbf{m}' \,(\text{Var}[\mathbf{m}\,|\,\mathbf{X}])^{-1}\, \mathbf{m}$$

to test if $\mathbf{m}$ is different from 0.

(2) We base the answer on a model loss of fit when restrictions are imposed: RSS must increase and R^2 must go down. Then, we construct an F test to check if the unrestricted RSS (RSS_U) is different from the restricted RSS (RSS_R).

63

Wald Test Statistic for H_0: $\mathbf{R\beta} - \mathbf{q} = \mathbf{0}$

- To test H_0, we calculate the discrepancy vector:

$$\mathbf{m} = \mathbf{Rb} - \mathbf{q}.$$

Then, we compute the Wald statistic:

$$W = \mathbf{m}' \,(\text{Var}[\mathbf{m}\,|\,\mathbf{X}])^{-1}\, \mathbf{m}$$

It can be shown that $\text{Var}[\mathbf{m}\,|\,\mathbf{X}] = \mathbf{R}[\sigma^2(\mathbf{X}'\mathbf{X})^{-1}]\mathbf{R}'$. Then,

$$W = (\mathbf{Rb} - \mathbf{q})' \,\{\mathbf{R}[\sigma^2(\mathbf{X}'\mathbf{X})^{-1}]\mathbf{R}'\}^{-1}\, (\mathbf{Rb} - \mathbf{q}) \sim \chi^2_J$$

Under H_0, assuming (**A5**) & estimating σ^2 with $s^2 = \mathbf{e}'\mathbf{e}/(T-k)$:

$$W^* = (\mathbf{Rb} - \mathbf{q})' \,\{\mathbf{R}[s^2(\mathbf{X}'\mathbf{X})^{-1}]\mathbf{R}\}^{-1}\, (\mathbf{Rb} - \mathbf{q})$$

Recall that $(T-k)\, s^2/\sigma^2 \sim \chi^2_{T-k}$. Then,

$$F = W^*/J \sim F_{J,T-k}$$

If (**A5**) is not assumed, the results are only asymptotic: $J^*F \xrightarrow{d} \chi^2_J$

64

Wald Test Statistic for H_0: $R\beta - q = 0$

Example: We want to test that in the 3 FF factor model (T=569)

1. H_0: $\beta_{SMB} = 0.2$ and $\beta_{HML} = 0.6$.
   H_1: $\beta_{SMB} \neq 0.2$ and/or $\beta_{HML} \neq 0.6$. $\Rightarrow J = 2$

We define **R** (2x4) below and write $\mathbf{m} = \mathbf{R}\beta - \mathbf{q} = \mathbf{0}$:

$$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} \beta_1 \\ \beta_{Mkt} \\ \beta_{SMB} \\ \beta_{HML} \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.6 \end{bmatrix}$$

2. Test-statistic: $F = W^*/J = (\mathbf{Rb} - \mathbf{q})' \{\mathbf{R}[s^2(\mathbf{X'X})^{-1}]\mathbf{R}'\}^{-1} (\mathbf{Rb} - \mathbf{q})$

   Distribution under H_0: $F = W^*/2 \sim F_{2,T-2}$ (asymptotic, $2^*F \xrightarrow{d} \chi^2_2$)

3. Get OLS results, compute F.

Wald Test Statistic for H_0: $R\beta - q = 0$

Example (continuation):

4. Decision Rule: $\alpha = 0.05$ level. We reject H_0 if p-value(F) < .05.
   Or, reject H_0, if $F > F_{J=2,T-2,.05}$.

```
J <- 2                                   # number of restriction
R <- matrix(c(0,0,0,0,1,0,0,1), nrow=2)  # matrix of restrictions
q <- c(.2,1)                             # hypothesized values
m <- R%*%b - q                           # m = Estimated R*Beta - q
Var_m <- R %*% Var_b %*% t(R)            # Variance of m
det(Var_m)                               # check for non-singularity
W <- t(m)%*%solve(Var_m)%*%m
F_t <- as.numeric(W/J)                   # F-test statistic

qf(.95, df1=J, df2=(T - k))              # exact distribution (F-dist) if errors normal
p_val <- 1 - pf(F_t, df1=J, df2=(T - k)) # p-value(F_t) under errors normal
p_val
```

Wald Test Statistic for H_0: $R\beta - q = 0$

Example (continuation):

```
> F_t
[1] 49.21676
>
> qf(.95, df1=J, df2=(T - k))                    # exact distribution (F-dist) if errors normal
[1] 3.011672                                     F_t > 3.011672 ⇒ reject H0 at 5% level
> p_val <- 1 - pf(F_t, df1=J, df2=(T - k))       # p-value(F_t) under errors normal
> p_val
[1] 0                                            very low chance H0 is true.
```

67

Wald Test Statistic for H_0: Does EH hold?

Example: Now, we do a joint test of the EH. H_0: $\alpha = 0$ and $\beta = 0$. Using the previous program but with:

```
J <- 2                                           # number of restriction
R <- matrix(c(1,0,0,1), nrow=2)                  # matrix of restrictions
q <- c(0,0)                                      # hypothesized values
> F_t
[1] 4.1024
>
> qf(.95, df1=J, df2=(T - k))                    # exact distribution (F-dist) if errors normal
[1] 3.020661                                     F_t > 3.020661 ⇒ reject H0 at 5% level
>
> p_val <- 1 - pf(F_t, df1=J, df2=(T - k))       # p-value(F_t) under errors normal
> p_val
[1] 0.01731                                      very low chance H0 is true.
```

68

The F Test: H_0: $\mathbf{R}\beta - \mathbf{q} = 0$

(2) We know that imposing the restrictions leads to a loss of fit. R^2 must go down. Does it go down a lot? –i.e., significantly?

Recall (i) $\mathbf{e}^* = \mathbf{y} - \mathbf{X}\mathbf{b}^* = \mathbf{e} - \mathbf{X}(\mathbf{b}^* - \mathbf{b})$

(ii) $\mathbf{b}^* = \mathbf{b} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b} - \mathbf{q})$

$$\Rightarrow \quad \mathbf{e}^{*\prime}\mathbf{e}^* = \mathbf{e}'\mathbf{e} + (\mathbf{b}^* - \mathbf{b})'\mathbf{X}'\mathbf{X}(\mathbf{b}^* - \mathbf{b})$$

$$\mathbf{e}^{*\prime}\mathbf{e}^* = \mathbf{e}'\mathbf{e} + (\mathbf{R}\mathbf{b}-\mathbf{q})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b}-\mathbf{q})$$

$$\mathbf{e}^{*\prime}\mathbf{e}^* - \mathbf{e}'\mathbf{e} = (\mathbf{R}\mathbf{b} - \mathbf{q})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b} - \mathbf{q})$$

Recall

– $W = (\mathbf{R}\mathbf{b} - \mathbf{q})'\{\mathbf{R}[\sigma^2(\mathbf{X}'\mathbf{X})^{-1}]\mathbf{R}'\}^{-1}(\mathbf{R}\mathbf{b} - \mathbf{q}) \sim \chi^2_J$ (if σ^2 is known)

– $\mathbf{e}'\mathbf{e}/\sigma^2 \sim \chi^2_{T-k}$.

Then,

$$F = (\mathbf{e}^{*\prime}\mathbf{e}^* - \mathbf{e}'\mathbf{e})/J \;/\; [\mathbf{e}'\mathbf{e}/(T-k)] \sim F_{J,T-k}.$$

69

The F Test: H_0: $\mathbf{R}\beta - \mathbf{q} = 0$

- $F = (\mathbf{e}^{*\prime}\mathbf{e}^* - \mathbf{e}'\mathbf{e})/J \;/\; [\mathbf{e}'\mathbf{e}/(T-k)] \sim F_{J,T-k}.$

Let R^2 = unrestricted model = 1 – RSS/TSS

R^{*2} = restricted model fit = 1 – RSS*/TSS

Then, dividing and multiplying F by TSS we get

$$F = ((1 - R^{*2}) - (1 - R^2))/J \;/\; [(1 - R^2)/(T-k)] \sim F_{J,T-k}$$

or

$$F = \{ (R^2 - R^{*2})/J \} \;/\; [(1 - R^2)/(T-k)] \sim F_{J,T-k}.$$

70

The F Test: F-test of goodness of fit

- In the linear model

$$\mathbf{y} = \mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{X}_1\,\boldsymbol{\beta}_1 + \mathbf{X}_2\,\boldsymbol{\beta}_2 + \ldots + \mathbf{X}_k\,\boldsymbol{\beta}_k + \boldsymbol{\varepsilon}$$

- We want to test if the slopes $\mathbf{X}_2, \ldots, \mathbf{X}_k$ are equal to zero. That is,

H_0: $\beta_2 = \ldots = \beta_k = 0$

H_1: at least one $\beta \neq 0$ $\Rightarrow J = k - 1$

- We can write H_0: $\mathbf{R}\boldsymbol{\beta} - \mathbf{q} = \mathbf{0}$ $\Rightarrow \begin{bmatrix} 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \ldots \\ \beta_k \end{bmatrix} = \begin{bmatrix} 0 \\ \ldots \\ 0 \end{bmatrix}$

- We have $J = k - 1$. Then,

$F = \{ (R^2 - R^{*2})/(k-1) \} \,/\, [(1 - R^2)/(T-k)] \sim F_{k-1,T-k}$.

- For the restricted model, $R^{*2} = 0$.

71

10

The F Test: F-test of goodness of fit

Then, $F = \{ R^2/(k-1) \}/[(1 - R^2)/(T-k)] \sim F_{k-1,T-K}$.

- Recall ESS/TSS is the definition of R^2. RSS/TSS is equal to $(1 - R^2)$.

$$F(k-1, T-k) = \frac{R^2/(k-1)}{(1-R^2)/(T-k)} = \frac{\frac{ESS}{TSS}/(k-1)}{\frac{RSS}{TSS}/(T-k)} = \frac{ESS/(k-1)}{RSS/(T-k)}$$

- This test statistic is called the *F-test of goodness of fit*.

72

10

The F Test: F-test of goodness of fit

Example: We want to test if all the FF factors (Market, SMB, HML) are significant, using monthly data 1973 – 2020 (T=569).

```
y <- ibm_x
T <- length(x)
x0 <- matrix(1,T,1)
x <- cbind(x0,Mkt_RF, SMB, HML)
k <- ncol(x)
b <- solve(t(x)%*% x)%*% t(x)%*%y #OLS regression
e <- y - x%*%b
RSS <- as.numeric(t(e)%*%e)
R2 <- 1 - as.numeric(RSS)/as.numeric(t(y)%*%y)    #R-squared
> R2
[1] 0.338985
F_goodfit <- (R2/(k-1))/((1-R2)/(T-k))              #F-test of goodness of fit.
> F_goodfit
[1] 96.58204
```

$\Rightarrow$ F_goodfit > $F_{2,565,.05}$ = 2.387708 $\Rightarrow$ Reject H_0.

73

10

The F Test: General Case - Example

- In the linear model

$$\mathbf{y} = \mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \beta_1 + \mathbf{X}_2\,\beta_2 + \mathbf{X}_3\,\beta_3 + \mathbf{X}_4\,\beta_4 + \boldsymbol{\varepsilon}$$

- We want to test if the slopes $\mathbf{X}_3$, $\mathbf{X}_4$ are equal to zero. That is,

$$H_0: \beta_3 = \beta_4 = 0$$
$$H_1: \beta_3 \neq 0 \text{ or } \beta_4 \neq 0 \quad \text{or both } \beta_3 \text{ and } \beta_4 \neq 0$$

- We can use, $F = (\mathbf{e}^{*\prime}\mathbf{e}^* - \mathbf{e}'\mathbf{e})/J \,/\, [\mathbf{e}'\mathbf{e}/(T-k)] \sim F_{J,T-K}$.

Define $Y = \beta_1 + \beta_2 X_2 + \varepsilon$ $\quad RSS_R$

$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$ $\quad RSS_U$

$$\boldsymbol{F}(\text{cost in } df, \text{unconstr } df) = \frac{RSS_R - RSS_U \,/\, k_U - k_R}{RSS_U \;/\; T - k_U}$$

74

32

The F Test: General Case - Example

Example: We want to test if the additional FF factors (SMB, HML) are significant, using monthly data 1973 – 2020 (T=569).

Unrestricted Model:

(U) $IBM_{Ret} - r_f = \beta_0 + \beta_1 (Mkt_{Ret} - r_f) + \beta_2 SMB + \beta_3 HML + \varepsilon$

Hypothesis: H_0: $\beta_2 = \beta_3 = 0$

H_1: $\beta_2 \neq 0$ and/or $\beta_3 \neq 0$

Then, the Restricted Model:

(R) $IBM_{Ret} - r_f = \beta_0 + \beta_1 (Mkt_{Ret} - r_f) + \varepsilon$

Test: $F = \frac{(RSS_R - RSS_U)/J}{RSS_U/(T-k_u)} \sim F_{J,T-K}$. with $J = k_U - k_R = 4 - 2 = 2$

75

The F Test: General Case - Example

Example (continuation): The unrestricted model was already estimated in Lecture 3. For the restricted model:

```
y <- ibm_x
x0 <- matrix(1,T,1)
x_r <- cbind(x0,Mkt_RF)                          # Restricted X vector
T <- nrow(x)
k2 <- ncol(x)

b2 <- solve(t(x_r)%*% x_r)%*% t(x_r)%*%y         # Restricted OLS regression
e2 <- y – x_r%*%b2
RSS2 <- as.numeric(t(e2)%*%e2)
> RSS = 1.932442                                 # RSS_U
> RSS2 = 1.964844                                # RSS_R
J <- k - k2                                      # J = degrees of freedom of numerator
F_test <- ((RSS2 - RSS)/J)/(RSS/(T-k))
```

76

The F Test: General Case - Example

Example (continuation):

F_test <- ((RSS2 - RSS)/J)/(RSS/(T-k))
> F_test
[1] **4.736834**
> qf(.95, df1=J, df2=(T-k)) # $F_{2,565,.05}$ value (≈ 3)
[1] 3.011672 ⇒ Reject H_0.
p_val <- 1 - pf(F_test, df1=J, df2=(T-k)) # p-value of F_test
> p_val
[1] **0.009117494** ⇒ p-value is small ⇒ Reject H_0.

77

Lagrange Multiplier Statistics

• Specific to the classical model.
Recall the Lagrange multipliers:

$$\boldsymbol{\lambda} = [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\,\mathbf{m}$$

Suppose we just test H_0: $\boldsymbol{\lambda} = \mathbf{0}$, using the Wald criterion.

$$W = \boldsymbol{\lambda}'(\text{Var}[\boldsymbol{\lambda}\,|\,\mathbf{X}])^{-1}\boldsymbol{\lambda}$$

where

$$\text{Var}[\boldsymbol{\lambda}\,|\,\mathbf{X}] = [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\text{Var}[\mathbf{m}\,|\,\mathbf{X}]\,[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}$$

$$\text{Var}[\mathbf{m}\,|\,\mathbf{X}] = \mathbf{R}[\sigma^2(\mathbf{X}'\mathbf{X})^{-1}]\mathbf{R}'$$

$$\text{Var}[\boldsymbol{\lambda}\,|\,\mathbf{X}] = [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\,\mathbf{R}[\sigma^2(\mathbf{X}'\mathbf{X})^{-1}]\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}$$
$$= \sigma^2\,[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}$$

Then,

$$W = \mathbf{m}'\,[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\,\{\sigma^2\,[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\}^{-1}\,[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\,\mathbf{m}$$
$$= \mathbf{m}'\,[\sigma^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\}\;\mathbf{m}$$

78

Application (Greene): Gasoline Demand

• Time series regression,

$$\text{Log}\mathbf{G} = \beta_1 + \beta_2\log\mathbf{Y} + \beta_3\log\text{PG} + \beta_4\log\mathbf{PNC} + \beta_5\log\mathbf{PUC} + \beta_6\log\mathbf{PPT} + \beta_7\log\mathbf{PN} + \beta_8\log\mathbf{PD} + \beta_9\log\mathbf{PS} + \boldsymbol{\varepsilon}$$

Period = 1960 - 1995.

• A significant event occurs in October 1973: the first oil crash. In the next lecture, we will be interested to know if the model 1960 to 1973 is the same as from 1974 to 1995.

Note: All coefficients in the model are elasticities.

79

Application (Greene): Gasoline Demand

```
Ordinary     least squares regression ............
LHS=LG       Mean                 =        5.39299
             Standard deviation   =         .24878
             Number of observs.   =             36
Model size   Parameters           =              9
             Degrees of freedom   =             27
Residuals    Sum of squares       =         .00855  <*******
             Standard error of e  =         .01780  <*******
Fit          R-squared            =         .99605  <*******
             Adjusted R-squared   =         .99488  <*******
--------+--------------------------------------------------------------
Variable| Coefficient    Standard Error  t-ratio  P[|T|>t]   Mean of X
--------+--------------------------------------------------------------
Constant|   -6.95326***      1.29811       -5.356   .0000
      LY|    1.35721***       .14562        9.320   .0000       9.11093
     LPG|    -.50579***       .06200       -8.158   .0000        .67409
    LPNC|    -.01654          .19957        -.083   .9346        .44320
    LPUC|    -.12354*         .06568       -1.881   .0708        .66361
    LPPT|     .11571          .07859        1.472   .1525        .77208
     LPN|    1.10125***       .26840        4.103   .0003        .60539
     LPD|     .92018***       .27018        3.406   .0021        .43343
     LPS|   -1.09213***       .30812       -3.544   .0015        .68105
--------+--------------------------------------------------------------
```

80

Application (Greene): Gasoline Demand

• Q: Is the price of public transportation really relevant? $H_0 : \beta_6 = 0$.

(1) Distance measure: $t_6 = (b_6 - 0) \,/\, s_{b6} = (.11571 - 0) \,/\, .07859$

$= 1.472 < 2.052 \quad \Rightarrow$ cannot reject H_0.

(2) Confidence interval: $b_6 \pm t_{(.95,27)} \times$ Standard error

$= .11571 \pm 2.052 \times (.07859)$

$= .11571 \pm .16127 = (-.045557 ,.27698)$

$\Rightarrow$ C.I. contains 0 $\quad \Rightarrow$ cannot reject H_0.

(3) Regression fit if $\mathbf{X}_6$ drop? Original $R^2 = .99605$,

Without LPPT, $R^{*2} = .99573$

$F(1,27) = [(.99605 - .99573)/1]/[(1 - .99605)/(36 - 9)] = 2.187$

$= 1.472^2$ (with some rounding) $\quad \Rightarrow$ cannot reject[8] H_0.

Gasoline Demand (Greene) - Hypothesis Test: Sum of Coefficients

• Do the three aggregate price elasticities sum to zero?

$H_0 : \beta_7 + \beta_8 + \beta_9 = 0$

$\mathbf{R} = [0, 0, 0, 0, 0, 0, 1, 1, 1], \quad \mathbf{q} = 0$

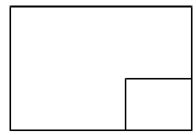
```
Variable| Coefficient    Standard Error  t-ratio  P[|T|>t]
------------+--------------------------------------------------
 LPN|    1.10125***        .26840         4.103   .0003        .60539
 LPD|      .92018***       .27018         3.406   .0021        .43343
 LPS|   -1.09213***        .30812        -3.544   .0015        .68105
```

	1	2	3	4	5	6	7	8	9
1	1.6851	-0.189024	-0.0256198	-0.218091	-0.0240267	-0.0295907	-0.0261772	0.197857	0.176068
2	-0.189024	0.0212045	0.00290895	0.0243971	0.00269963	0.0032894	0.00280174	-0.0222154	-0.0195876
3	-0.0256198	0.00290895	0.00384368	-0.000682307	-0.000413822	-0.00176052	-0.0114883	-0.0044953	0.0108144
4	-0.218091	0.0243971	-0.000682307	0.0398293	0.00350897	0.00824835	0.0236143	-0.0311143	-0.0453555
5	-0.0240267	0.00269963	-0.000413822	0.00350897	0.00431411	0.001419	0.00979376	-0.0118214	-0.00970482
6	-0.0295907	0.0032894	-0.00176052	0.00824835	0.001419	0.00617673	0.0134911	-0.00740557	-0.0198458
7	-0.0261772	0.00280174	-0.0114883	0.0236143	0.00979376	0.0134911	0.0720371	-0.0335608	-0.0705545
8	0.197857	-0.0222154	-0.0044953	-0.0311143	-0.0118214	-0.00740557	-0.0335608	0.0729982	0.0346625
9	0.176068	-0.0195876	0.0108144	-0.0453555	-0.00970482	-0.0198458	-0.0705545	0.0346625	0.0949391

82

Gasoline Demand (Greene) - Hypothesis Test: Sum of Coefficients – Wald Test

```
--> MATRIX ; list ;R = [0,0,0,0,0,0,1,1,1] ; q = [0]
    ; m = R*b - q
    ; Varm = R*Varb*R'
    ; Wald = m' <Varm> m $
```

Var[m] = R × Var[b] × R′ = [0 0 0 0 0 0 1 1 1]  $\begin{pmatrix}0\\0\\0\\0\\0\\0\\1\\1\\1\end{pmatrix}$

$\sum_{i=1}^{9}\sum_{j=1}^{9} R_i R_j Cov(b_i, b_j)$ = 0.10107

m′ [Var(m)]$^{-1}$ m = 8.5446

The critical chi squared with 1 degree of freedom is 3.84, so the hypothesis is rejected.

83

Gasoline Demand (Greene) - Imposing the Restriction

```
Linearly restricted regression
LHS=LG       Mean                 =   5.392989
             Standard deviation   =   .2487794
             Number of observs.   =         36
Model size   Parameters           =          8  <*** 9 – 1 restriction
             Degrees of freedom   =         28
Residuals    Sum of squares       =   .0112599  <*** With the restriction
Residuals    Sum of squares       =   .0085531  <*** Without the
restriction
Fit          R-squared            =   .9948020
Restrictns.  F[  1,     27] (prob) =  8.5(.01)
Not using OLS or no constant.R2 & F may be < 0
```

Variable	Coefficient	Standard Error	t-ratio	P[|T|>t]	Mean of X
Constant	-10.1507***	.78756	-12.889	.0000	
LY	1.71582***	.08839	19.412	.0000	9.11093
LPG	-.45826***	.06741	-6.798	.0000	.67409
LPNC	.46945***	.12439	3.774	.0008	.44320
LPUC	-.01566	.06122	-.256	.8000	.66361
LPPT	.24223***	.07391	3.277	.0029	.77208
LPN	1.39620***	.28022	4.983	.0000	.60539
LPD	.23885	.15395	1.551	.1324	.43343
LPS	-1.63505***	.27700	-5.903	.0000	.68105

84

F = [(.0112599 - .0085531)/1] / [.0085531/(36 – 9)] = 8.544691

Gasoline Demand (Greene)- Joint Hypotheses

• Joint hypothesis: Income elasticity = +1, Own price elasticity = -1.
The hypothesis implies that $\log G = \beta_1 + \log Y - \log Pg + \beta_4 \log PNC + ...$

Strategy: Regress logG – logY + logPg on the other variables and

• Compare the sums of squares
With two restrictions imposed
Residuals Sum of squares = .0286877
Fit R-squared = .9979006
Unrestricted
Residuals Sum of squares = .0085531
Fit R-squared = .9960515

F = ((.0286877 - .0085531)/2) / (.0085531/(36-9)) = 31.779951
The critical F for 95% with 2,27 degrees of freedom is 3.354 $\Rightarrow H_0$ is rejected.

• Q: Are the results consistent? Does the R^2 really go up when the restrictions are imposed?

85

Gasoline Demand - Using the Wald Statistic

```
--> Matrix ; R = [0,1,0,0,0,0,0,0,0,0 /
                  0,0,1,0,0,0,0,0,0,0]$
--> Matrix ; q = [1/-1]$
--> Matrix ; list ; m = R*b - q $
Matrix M        has  2 rows and  1 columns.
               1
        +-------------+
       1|     .35721
       2|     .49421
        +-------------+
--> Matrix ; list ; vm = R*varb*R' $
Matrix VM       has  2 rows and  2 columns.
               1             2
        +-------------+-------------+
       1|     .02120       .00291
       2|     .00291       .00384
        +-------------+-------------+
--> Matrix ; list ; w = 1/2 * m'<vm>m $
Matrix W        has  1 rows and  1 columns.
               1
        +-------------+
       1|   31.77981
        +-------------+
```

86

Gasoline Demand (Greene) – Testing Details

• Q: Which restriction is the problem? We can look at the *J*x1 estimated LM, λ, for clues:

$$\lambda = [\mathbf{R}(\mathbf{X}'\mathbf{X})\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b} - \mathbf{q})$$

• Recall that under H_0, λ should be 0.

Matrix Result has 2 rows and 1 columns.

```
              1
  +--------------+
 1|     -.88491      Income elasticity
 2|   129.24760      Price elasticity
  +--------------+
```

Results *suggest* that the constraint on the price elasticity is having a greater effect on the sum of squares.

87

Gasoline Demand (Greene) - Basing the Test on R^2

• After building the restrictions into the model and computing restricted and unrestricted regressions: Based on R^2s,

F = [(.9960515 – .997096)/2]/[(1 –.9960515)/(36-9)]
 = -3.571166 (!)

• Q: What's wrong?

88