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.

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.

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.

<u>Example</u>: H_0:θ=θ_0 against H_1:θ=θ_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.

<u>Example</u>: H_0:θ>θ_0 against H_1:θ<θ_1.

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 can collect a sample of data $X=\{X_1,\ldots X_n\}$ and device a decision rule:

$$\text{if } X \in R, \quad \Rightarrow \text{we reject } \mathrm{H}_0$$
$$\text{if } X \notin R \text{ or } X \in R^C \quad \Rightarrow \text{we fail to reject } \mathrm{H}_0$$

The set R is called the *region of rejection* or the *critical region* of the test.

Hypothesis Testing: Brief Review

• The rejection region is defined in terms of a statistics $T(X)$, called the *test statistic*. Note that like any other statistic, $T(X)$ is a random variable. Given this test statistic, the decision rule can then be written as:

$$T(X) \in R \quad \Rightarrow \text{reject } \mathrm{H}_0$$
$$T(X) \in R^C \quad \Rightarrow \text{fail to reject } \mathrm{H}_0$$

• Remember, we only learn from 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 (1901-1954, Italy)

Hypothesis Testing: Brief Review - Fisher

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

• Fisher's Idea

Form H_0. Collect a sample of data $X=\{X_1,\dots X_n\}$. Compute the test-statistics $T(X)$ used to test H_0. Report the *p-value* -i.e., the probability, of observing a result at least as extreme as the test statistic, under H_0.

If the p-value is smaller than a *significance level,* say 5%, the result is *significant* and H_0 is rejected. If the results are "not significant," no conclusions are reached. Go back gather more data or modify model.

• Fisher used the p-value as a way to determine the faith in H_0.

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.

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.

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).

• Notation: 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:

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

Note: when $\theta \in H_1$ $\Rightarrow \pi(\theta) = 1 - \beta(\theta)$.

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

Type I and Type II Errors

	State of World	
Decision	H_o true	H_1 true (H_o false)
"Accept" (cannot reject) H_o	*Correct decision*	Type II error
Reject H_o	Type I error	*Correct decision*

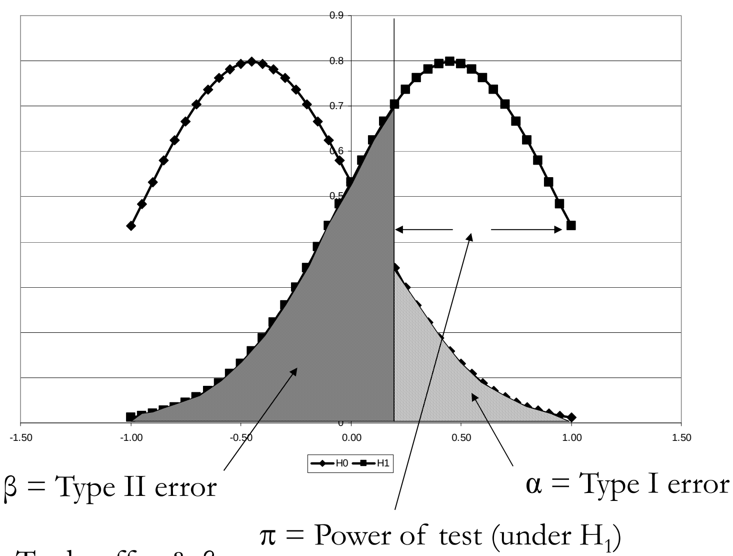
← *Learning*

Need to control both types of error:

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

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

Type I and Type II Errors



Note: Trade-off α & β.

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."

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

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.

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 $\beta(R) \leq \beta(R_1)$.

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:$\theta=\theta_1$.

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.

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 $L(\mathbf{x}|\theta_i)$, i=0,1, using a test with rejection region R that satisfies

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

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

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 $P[X\in A|H_0] = P[X\in A|H_1]=0$

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.

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

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.

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 $\alpha = \Pr(T > t_0 | \theta_0)$.

KR Theorem: Practical Use

<u>Goal</u>: 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 $T < t_0$.

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 $T>t_0$).

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.

<u>Simple intuition</u>: 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.

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.

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.

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 α **=.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: $z \geq z_{\alpha=.05} = 1.645$.

No UMP test

Example (continuation):

Test-statistitc: 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 $\bar{x} \geq .1736$ $[= z_{\alpha/2} * \sigma/\text{sqrt}(n) + \theta_0 = 1.645 * .2/\text{sqrt}(20) + .1]$

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

• Changing $\theta - \theta_0$

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

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

No UMP test

Example (continuation):

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

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

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

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

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

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

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

$\Rightarrow$ Power = P[X ∈ R|H_1] = P[$\bar{x} \geq$ (.12323 – 0.2)/(.2/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 ∈ R|H_1] = .90.

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.

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

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

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.

Likelihood Ratio Tests: Example I

Example: $\lambda(X)$ 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.

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

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

Note: 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: $\lambda(X)$ for a $X \sim$ *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 $P[\lambda(X)<k] = \alpha$. 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)

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) Test H_0. For example, reject H_0: if θ_0 is too far from $\hat{\theta}$ (we would say the hypothesis is inconsistent with the sample evidence.)

To test H_0 we need a decision rule. This decision rule will be based on a statistic. 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.

Hypothesis Testing: Issues

• Logic of the Neyman-Pearson methodology:
If H_0 is true, then the statistic 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.

• To test H_0 we need a decision rule. This decision rule will be based on a statistic that will tells us what is too far.

$\Rightarrow$ too far: statistic falls in the rejection region, R.

If the observed value falls in R, we conclude that the assumed distribution must be incorrect and H_0 should be rejected.

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.

Hypothesis Testing in the CLM

• The CLM is used to test hypotheses about the underlying DGP, which is assumed to be linear.

Example:
Suppose the model (DGP) we use is $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$
Using OLS, we estimate $\mathbf{b}_1$ and $\mathbf{b}_2$.
We formulate a hypothesis: The variable $\mathbf{X}_2$ should not be in the DGP
This hypothesis is testable: H_0: $\boldsymbol{\beta}_2 = \mathbf{0}$ against H_1: $\boldsymbol{\beta}_2 \neq \mathbf{0}$.
We need a statistic to test H_0: $z_2 = (\mathbf{b}_2 - \mathbf{0})/\text{sqrt}\{\sigma^2(\mathbf{X}'\mathbf{X})_{22}^{-1}\}$
If $\boldsymbol{\varepsilon}|\mathbf{X} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_T)$ and if σ^2 is known, then, under H_0, $z_2 \sim N(0, 1)$.
Decision Rule: We reject H_0, at the 5% level, if $|z_2|>1.96$.

Note: It should be clear that under H_1, z_2 will not follow a $N(0, 1)$.

Hypothesis Testing: Confidence Intervals

• The OLS estimate $\mathbf{b}$ is a point estimate for $\boldsymbol{\beta}$, meaning that $\mathbf{b}$ 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 $\text{Prob}[\theta \in C_n]$.

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 -α)% C.I.

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

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

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

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,

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

• 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.

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

Testing a Hypothesis About a Single Parameter

• We estimate by OLS the linear model $\mathbf{y} = \mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon}$
• We are interested in testing H_0: $\beta_k = \beta^0_k$ against H_1: $\beta_k \neq \beta^0_k$.

• For now, we will rely on assumption (**A5**) $\boldsymbol{\varepsilon}|\mathbf{X} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_T)$

• Let b_k = OLS estimator of β_k
Std Dev $[b_k|\mathbf{X}] = \text{sqrt}\{[\sigma^2(\mathbf{X'X})^{-1}]_{kk}\} = v_k$
From assumption (**A5**), we know that
$b_k|\mathbf{X} \sim N(\beta_k, v_k^2)$ $\Rightarrow$ Under H_0: $b_k|\mathbf{X} \sim N(\beta^0_k, v_k^2)$.
$\Rightarrow$ Under H_0: $(b_k - \beta^0_k)/v_k|\mathbf{X} \sim N(0,1)$.

• Q: How far is b_k from β^0_k? If it is too far, H_0 is inconsistent with the sample evidence. We measure distance in standard error units:
$z_b = (b_k - \beta^0_k)/v_k$

Testing a Hypothesis About a Single Parameter

• We measure distance in standard error units:
$z_b = (b_k - \beta^0_k)/v_k$
Note: z_b is an example of the *Wald* (normalized) *distance measure*. Most tests in econometrics will use this measure.

Decision rule: If z_b is large (larger than a critical value), reject H_0.

• If σ^2 is known, $v_k^2 = [\sigma^2(\mathbf{X'X})^{-1}]_{kk}$ is known $\Rightarrow z_b|\mathbf{X} \sim N(0,1)$.
• If σ^2 is unknown, $v_k^2 = [\sigma^2(\mathbf{X'X})^{-1}]_{kk}$ is not known because σ^2 must be estimated. We use s^2 instead of σ^2. Then,

$$t_b = (b_k - \beta^0_k)/\text{Est.}(v_k) \sim t_{T-k}.$$

• Rule for H_0: $\beta_k = \beta^0_k$ against H_1: $\beta_k \neq \beta^0_k$:
If $|t_b| > t_{T-k}(\alpha/2)$, reject H_0 at the α significance level.

Recall: A *t*-distributed variable

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

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,

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

OLS Estimation – 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)-1X' 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  (see Later)
Sigma2 <- as.numeric(RSS/(T-k))                    # Estimated σ2 = s2
SE_reg <- sqrt(Sigma2)                             # Estimated σ – Regression stand error
Var_b <- Sigma2*solve(t(x)%*% x)                   # Estimated Var[b|X] = s2 (X'X)-1
SE_b <- sqrt(diag(Var_b))                          # SE[b|X]
t_b <- b/SE_b                                      # t-stats (See Chapter 4)
```

OLS Estimation – Testing Example in R

```
> t(b)
                x1        x2         x3
[1,] -0.2258839 1.061934 0.1343667 -0.3574959
> SE_b
                x1         x2         x3
0.01095196 0.26363344 0.35518792 0.37631714
> t(t_b)
                x1        x2         x3
[1,] -20.62498 4.028071 0.3782976 -0.9499857
```

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

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

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

$t_1 = (\mathbf{1.061934} - \mathbf{1})/\ \mathbf{0.263634} = 0.23492$

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

Testing a Hypothesis: Wald Statistic

• Most of our test statistics are Wald statistics.

Wald = normalized distance measure:

W = (random vector – hypothesized value)′ * [Variance] $^{-1}$ *

* (random vector – hypothesized value)

$= \mathbf{z}' \,\text{Var}(\mathbf{z})^{-1}\mathbf{z}$

• 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, $W \sim F$

Abraham Wald (1902–1950, Hungary)

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,

$(y-\mu_y)' \, \Sigma_y^{-1} \, (y-\mu_y) \sim \chi_n^2$. –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.

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

If Z and U are independent, then

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

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.

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

• Now, we have J joint hypotheses. Let $\mathbf{R}$ be a Jxk matrix and $\mathbf{q}$ be a Jx1 vector.

• Two approaches to testing (unifying point: OLS is unbiased):

(1) Is $\mathbf{Rb} - \mathbf{q}$ close to $\mathbf{0}$? Basing the test on the discrepancy vector: $\mathbf{m} = \mathbf{Rb} - \mathbf{q}$. Using the Wald statistic:

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

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

Under the usual assumption and assuming σ^2 is known, $W \sim \chi_J^2$

In general, σ^2 is unknown, we use $s^2 = \mathbf{e}'\mathbf{e}/(T-k)$

$$W^* = (\mathbf{Rb} - \mathbf{q})'\{\mathbf{R}[s^2(\mathbf{X'X})^{-1}]\mathbf{R}\}^{-1}(\mathbf{Rb} - \mathbf{q})$$
$$= (\mathbf{Rb} - \mathbf{q})'\{\mathbf{R}[\sigma^2(\mathbf{X'X})^{-1}]\mathbf{R}\}^{-1}(\mathbf{Rb} - \mathbf{q})/(s^2/\sigma^2)$$
$$F = W/J \;/\; [(T-k)\,(s^2/\sigma^2)/(T-k)] = W^*/J \sim F_{J,T-k}.$$

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

(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

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

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

Then,

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

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

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

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

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

Example I: Testing H_0: $R\beta - q = 0$

- 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,

$$H_0: \beta_2 = \ldots = \beta_k = 0$$
$$H_1: \text{at least one } \beta \neq 0$$

- We can write H_0: $\mathbf{R\beta - q = 0}$

$$\begin{bmatrix} 0 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \ldots \\ \beta_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \ldots \\ 0 \end{bmatrix}$$

- 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, $R^{*2} = 0$.

10

Example I: Testing H_0: $R\beta - q = 0$

- 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 $(1 - R^2)$.

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

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

10

Example I: Testing H_0: $R\beta - q = 0$

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

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

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

$$= \frac{ESS/(k-1)}{RSS/(T-k)}$$

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

10

Example II: Testing H_0: $R\beta - q = 0$

- 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,

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

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

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

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

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

32

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.

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

where

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

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

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

Then,

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

Application: 3 Factor Fama-French Model

• We estimate the 3 factor F-F model for IBM returns, using monthly data Jan 1990 – Aug 2016 (T=320)

(U) $\quad IBM_{Ret} - r_f = \beta_0 + \beta_1\,(Mkt_{Ret} - r_f) + \beta_2\,SMB + \beta_4\,HML + \boldsymbol{\varepsilon}$

• We test if the additional F-F factors (SMB, HML) are significant:

H_0: $\beta_2 = \beta_3 = 0$ vs. H_1: $\beta_2 \neq 0$ and/or $\beta_3 \neq 0$

We estimate: (R) $\quad IBM_{Ret} - r_f = \beta_0 + \beta_1\,(Mkt_{Ret} - r_f) + \boldsymbol{\varepsilon}$

Then,

$$F = \frac{(RSS_R - RSS_U)/(k_U - k_R)}{RSS_U/(T - k_U)} =$$
$$= \frac{(12.5477 - 12.4882)/(4-2)}{12.4882/(320-4)} = 0.7532$$

$\Rightarrow$ cannot reject H_0 at 5% level ($F_{2,316,.05} \approx 3$).

Application (Greene): Gasoline Demand

• Time series regression,

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

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.

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
--------+-------------------------------------------------------------
```

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 H_0.

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

• Do the three aggregate price elasticities sum to zero?

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

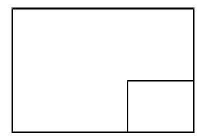
$\mathbf{R} = [0, 0, 0, 0, 0, 0, 1, 1, 1]$, $\quad \mathbf{q} = 0$

```
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

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.

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

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 logG = β_1 + logY – logPg + β_4 logPNC + ...

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 ⇒ H_0 is rejected.

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

Gasoline Demand - Using the Wald Statistic

```
--> Matrix ; R = [0,1,0,0,0,0,0,0,0 /
                  0,0,1,0,0,0,0,0,0]$
--> 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
        +-------------+
```

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.

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?