Lecture 4
Testing in the Classical Linear Model

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.

  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 $H_1: \theta < \theta_1$. 
Hypothesis Testing: Brief Review

• We compare two competing hypothesis:
  \( H_0 \) vs. \( H_1 \).

• Suppose the two hypothesis partition the universe:
  \( H_1 = \text{Not } H_0 \).

• Then, we can collect a sample of data \( X = \{X_1, \ldots, X_n\} \) and device a decision rule:
  
  \[
  \begin{align*}
  \text{if } X \in R, & \quad \Rightarrow \text{ we reject } H_0 \\
  \text{if } X \notin R \text{ or } X \in R^C, & \quad \Rightarrow \text{ we fail to reject } H_0
  \end{align*}
  \]

  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:
  \[
  \begin{align*}
  T(X) \in R & \quad \Rightarrow \text{ reject } H_0 \\
  T(X) \in R^C & \quad \Rightarrow \text{ fail to reject } H_0
  \end{align*}
  \]

• 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 $X = \{X_1, \ldots, X_n\}$. 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: $\alpha = P[X \in R \mid H_0]$  
  Probability of Type II error: $\beta = P[X \in R^c \mid H_1]$

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, $\theta$:

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

Note: when $\theta \in H_1$  
  => $\pi(\theta) = 1 - \beta(\theta)$. 
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 $\theta$. The maximum value of this when $\theta \in H_0$ is called *level of significance* of a test, denoted by $\alpha$. Thus,

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

Define a *level $\alpha$ test* to be a test with $\sup_{\theta \in H_0} \pi(\theta) \leq \alpha$.

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 $\alpha$* and the actual *rejection probability (empirical size)*. They may differ greatly.

---

**Type I and Type II Errors**

<table>
<thead>
<tr>
<th>State of World</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$ true</td>
<td>“Accept” (cannot reject) $H_0$</td>
</tr>
<tr>
<td>$H_1$ true (H_0 false)</td>
<td>Correct decision</td>
</tr>
<tr>
<td>$H_1$ true (H_0 false)</td>
<td>Type II error</td>
</tr>
<tr>
<td>$H_0$</td>
<td>Reject $H_0$</td>
</tr>
<tr>
<td>$H_1$ true (H_0 false)</td>
<td>Correct decision</td>
</tr>
</tbody>
</table>

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 $H_1$).
Type I and Type II Errors

\[ \beta = \text{Type II error} \quad \alpha = \text{Type I error} \]

\[ \pi = \text{Power of test (under } H_1) \]

Note: Trade-off \( \alpha \) & \( \beta \).

Type I and Type II Errors - Example

- We conduct a 1,000 studies of some hypothesis (say, \( H_0: \mu = 0 \))
  - 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)

<table>
<thead>
<tr>
<th>State of World</th>
<th>Decision</th>
<th>( H_0 ) true</th>
<th>( H_1 ) true (( H_0 ) false)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cannot reject ( H_0 )</td>
<td>855</td>
<td>50 (Type II error)</td>
<td></td>
</tr>
<tr>
<td>Reject ( H_0 )</td>
<td>45 (Type I error)</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td></td>
<td>900</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

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 $\alpha$ ($P$), higher power, lower % of false-positives –i.e., more true learning.

<table>
<thead>
<tr>
<th>Proportion of ideas that are correct (null hypothesis false)</th>
<th>Power of study</th>
<th>Percentage of “significant” results that are false-positives</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$P=0.05$</td>
</tr>
<tr>
<td>80%</td>
<td>20%</td>
<td>5.9</td>
</tr>
<tr>
<td></td>
<td>50%</td>
<td>2.4</td>
</tr>
<tr>
<td></td>
<td>80%</td>
<td>1.5</td>
</tr>
<tr>
<td>50%</td>
<td>20%</td>
<td>20.0</td>
</tr>
<tr>
<td></td>
<td>50%</td>
<td>9.1</td>
</tr>
<tr>
<td></td>
<td>80%</td>
<td>5.9</td>
</tr>
<tr>
<td>10%</td>
<td>20%</td>
<td>69.2</td>
</tr>
<tr>
<td></td>
<td>50%</td>
<td>47.4</td>
</tr>
<tr>
<td></td>
<td>80%</td>
<td>36.0</td>
</tr>
<tr>
<td>1%</td>
<td>20%</td>
<td>96.1</td>
</tr>
<tr>
<td></td>
<td>50%</td>
<td>90.8</td>
</tr>
<tr>
<td></td>
<td>80%</td>
<td>86.1</td>
</tr>
</tbody>
</table>

**More Powerful Test**

**Definition:** More Powerful Test

Let $(\alpha_1, \beta_1)$ and $(\alpha_2, \beta_2)$ 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 $\alpha$

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

**Definition:** Most powerful test of level $\alpha$

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

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

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

- This idea is the basis of the Neyman-Pearson Lemma to construct a test of a hypothesis about $\theta$: $H_0; \theta = \theta_0$ against $H_1; \theta = \theta_1$. 
• Neyman-Pearson Lemma provides a procedure for selecting the best test of a simple hypothesis about $\theta$: $H_0: \theta = \theta_0$ against $H_1: \theta = \theta_1$.

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

$$\begin{align*}
(1) \quad x \in R & \text{ if } L(x|\theta_1) > k L(x|\theta_0) \\
& \quad x \in R' \text{ if } L(x|\theta_1) < k L(x|\theta_0),
\end{align*}$$

for some $k \geq 0$, and

$$\begin{align*}
(2) \quad \alpha = P[X \in R|H_0]
\end{align*}$$

Then,

(a) Any test that satisfies (1) and (2) is a UMP level $\alpha$ test.

(b) If there exists a test satisfying (1) and (2) with $k > 0$, then every UMP level $\alpha$ test satisfies (2) and every UMP level $\alpha$ 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 $\theta_i$. 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
\[
\lambda = \frac{L(x|\theta_1)}{L(x|\theta_0)}
\]
is a non-decreasing function of $u(X)$ for each choice of $\theta_1$ and $\theta_0$, with $\theta_1 > \theta_0$.

If $L(x|\theta_i)$ 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 $L(x|\theta_i)$ 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,
\[
\ln \lambda = \Sigma_i T(X_i) [A(\theta_1)-A(\theta_0)] + nB(\theta_1) - nB(\theta_0).
\]
Let $u(X)=\Sigma_i T(X_i)$.

\[=\] \[\delta \ln \lambda / \delta u = [A(\theta_1)-A(\theta_0)] >0, \text{ if } A(.) \text{ is monotonic in } \theta.
\]
In addition, $u(X)$ is a sufficient statistic.

• Some distributions with MLRP in $T(X)=\Sigma_i x_i$: normal (with $\sigma$ 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 $\alpha$, where $\alpha = \Pr(T > t_0 | \theta_0)$.

**KR Theorem: Practical Use**

**Goal:** Find the UMP level $\alpha$ 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 $\alpha$ 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 $\theta$. Choose the direction to reject ($T > t_0$ or $T < t_0$) based on whether $E[T]$ is higher or lower for $\theta$ 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 $\alpha$ percentile (if you reject for $T < t_0$) or the $1 - \alpha$ 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.

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

$$\pi(\theta) \geq \alpha \quad \text{for all } \theta \in H_1$$

and

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

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

Most two-sided tests we use are UMP level $\alpha$ unbiased (UMPU) tests.
Some problems left for students

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

- 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 \( \alpha \) tests (all tests with \( \sup_{\theta \in H_0} \pi(\theta) \leq \alpha \)) and find the MP one available.

- In some cases there is a UMP level \( \alpha \) 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:
  - Effect size: True value - Hypothesized value. (Say, $\theta - \theta_0$). Bigger deviations from $H_0$ are easier to detect.
  - Sample size. Higher sample size, smaller sampling error. Sampling distributions are more concentrated!
  - Statistical significance –i.e., the $\alpha$.

**Example:** We randomly collect 20 stock returns, which are assumed $N(\theta, 0.2^2)$ (known $\sigma^2$ 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 $\theta$ is 20% ($H_1: \theta=0.2$ is true)?

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

Rejection rule: $z \geq z_{\alpha/2} = 1.645$.

$=>$ Power = $P[X \in R | H_1] = P[z \geq 1.645 | \theta > 0.2]$

$= P[z \geq (0.1736 - 0.2)/(0.2/sqrt(20))]$

$= P[z \geq -0.591]

$= 1 - P[z < -0.591] = 0.722760$

**No UMP test (continuation):**

Test statistic: $z-statistic = (\bar{X} - \theta_0)/[\sigma/sqrt(n)] = (\bar{X} - 0.1)/(0.2/sqrt(20))$.

Rejection rule: $z \geq z_{\alpha/2} = 1.645$, or, equivalently, when the observed $\bar{X} \geq .1736 \ [= z_{\alpha/2} * \sigma/sqrt(n) + \theta_0 = 1.645 * 0.2/sqrt(20) + 0.1]$

$=>$ Power = $P[X \in R | H_1] = P[z \geq 1.645 | \theta = 0.3]$

$= P[z \geq (0.1736 - 0.3)/(0.2/sqrt(20))]$

$= P[z \geq -2.82713] = 0.997652$

- Changing $\theta_1$ If ($H_1: \theta=0.3$ is true), then the power of the test (under $H_1$):

$=>$ Power = $P[X \in R | H_1] = P[z \geq (.1736 - 0.3)/(0.2/sqrt(20))]$

$= P[z \geq 2.82713] = 0.002348$
No UMP test

Example (continuation):

- Changing \( \alpha (\theta_1 = 0.2; n=20) \)
  
  If \( \alpha = 0.01 \), then rejection rule: \( Z \geq z_{\alpha/2} = 2.33 \).
  
  Or equivalently: \( X \bar{\theta} \geq 0.2042 \) \( = 2.33 \cdot \frac{0.2}{\sqrt{20}} + 0.1 \)
  
  \[ \Rightarrow \text{Power} = P[X \in R | H_1] = P[X \bar{\theta} \geq (0.2042 - 0.2)/(0.2/\sqrt{20})] \]
  
  \[ = P[Z \geq 0.093915] = 0.46259 \]

- Changing \( n (\theta_1 = 0.2; \alpha = 0.05) \)
  
  If \( n = 200 \), then rejection rule: \( X \bar{\theta} \geq 0.12332 \) \( = 1.645 \cdot \frac{0.2}{\sqrt{200}} + 0.1 \)
  
  \[ \Rightarrow \text{Power} = P[X \in R | H_1] = P[X \bar{\theta} \geq (0.12323 - 0.2)/(0.2/\sqrt{200})] \]
  
  \[ = P[Z \geq -5.4261] = 0.9999999 \]

**Note:** We can select \( n \) to achieve a given power (for given \( \theta_1 \) & \( \alpha \)). Say, set \( n = 34 \) to set \( P[X \in R | H_1] = 0.90 \).

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
  \[ \lambda(X) = \sup_{\theta \in H_0} L(X|\theta) / \sup_{\theta} L(X|\theta) \]

  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 \( \lambda(X) = \sup_{\theta \in H_0} L(X|\theta) / \sup_{\theta} L(X|\theta) \)
  (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, \( -2 \ln \lambda(X) \sim \chi^2_p \), 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 \( X \sim N(\theta, \sigma^2) \) for \( H_0: \theta = \theta_0 \) vs. \( H_1: \theta \neq \theta_0 \). Assume \( \sigma^2 \) is known.

\[
\lambda(x) = \frac{L(\theta_0 | x)}{L(x | x)} = \frac{(2\pi)^{n/2} e^{-\frac{1}{2} \sum_{i=1}^{n} (x_i - \theta_0)^2 / 2\sigma^2}}{(2\pi)^{n/2} e^{-\frac{1}{2} \sum_{i=1}^{n} (x_i - \bar{x})^2 / 2\sigma^2}} = e^{\frac{-n(x-\theta_0)^2}{2\sigma^2}}
\]

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

Note: Finding \( k \) is not needed. Why? We know the left hand side is distributed as a \( \chi^2_p \), thus \(-2 \ln k\) needs to be the \( 1-\alpha \) percentile of a \( \chi^2_p \). We need not solve explicitly for \( k \), we just need the rejection rule.

Likelihood Ratio Tests: Example II

Example: \( \lambda(X) \) for a \( X \sim \text{exponential}(\lambda) \) for \( H_0: \lambda = \lambda_0 \) vs. \( H_1: \lambda \neq \lambda_0 \).

\[
L(X | \theta) = \lambda^n \exp(-\lambda \sum x_i) = \lambda^n \exp(-\lambda n \bar{x}) \quad \Rightarrow \lambda_{\text{MLE}} = 1/\bar{x}
\]

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

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

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\( (n; \lambda / n) \), but we cannot get further.

It is, however, possible to determine the cutoff point, \( k \), by simulation (set \( n, \lambda_0 \)).
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, \( y = f(X, \theta) + \varepsilon \)
  (2) We gather data \((y, X)\) 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 \(\theta_0\) 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 $\theta_0$ 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.

  => 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 \( y = X_1\beta_1 + X_2\beta_2 + \epsilon \)
Using OLS, we estimate \( b_1 \) and \( b_2 \).
We formulate a hypothesis: The variable \( X_2 \) should not be in the DGP
This hypothesis is testable: \( H_0: \beta_2 = 0 \) against \( H_1: \beta_2 \neq 0 \).
We need a statistic to test \( H_0: z_2 = (b_2 - 0)/\sqrt{\sigma^2(X'X)^{-1}} \)
If \( \epsilon | X \sim N(0, \sigma^2I_T) \) and if \( \sigma^2 \) 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 \( b \) is a point estimate for \( \beta \), 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 \( \theta \). 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 \( \theta \) 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 prespecified 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 straightforward to construct a C.I. For example, if the distribution of \( b \) is normal, then a 95% C.I. is given by:

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

• 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 \( \frac{z_{\alpha/2}}{2} \). That is,

\[
C_n = \{ b_k : | z_n(b_k) | \leq \frac{z_{\alpha/2}}{2} \} \quad \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.

• Summary: C.I.’s are a simple but effective tool to assess estimation uncertainty.
• We estimate by OLS the linear model $y = X \beta + \epsilon$
• 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 $(A5)$ $\epsilon | X \sim N(0, \sigma^2 I_T)$

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

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

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

  \textbf{Decision rule:} If $z_b$ is large (larger than a critical value), reject $H_0$.

• If $\sigma^2$ is known, $v_k^2 = [\sigma^2 (X'X)^{-1}]_{kk}$ is known $\Rightarrow z_b | X \sim N(0,1)$.

• If $\sigma^2$ is unknown, $v_k^2 = [\sigma^2 (X'X)^{-1}]_{kk}$ is not known because $\sigma^2$ must be estimated. We use $s^2$ instead of $\sigma^2$. 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 $\alpha$ significance level.
Recall: A $t$-distributed variable

- Recall a $t_{\nu}$-distributed variable is a ratio of two independent RV: a $N(0,1)$ RV and the square root of a $\chi_{\nu}^2$ RV divided by $\nu$.

Let

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

Let

$$U = \frac{(n-1)s^2}{\sigma^2} ~ \chi_{n-1}^2$$

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

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

Testing a Hypothesis: Wald Statistic

- Most of our test statistics are Wald statistics.

Wald = normalized distance measure:

$$W = \text{(random vector - hypothesized value)}' \times \text{[Variance]}^{-1} \times \text{(random vector - hypothesized value)}$$

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

- Distribution of $W$? We have a quadratic form.
  - If $z$ is normal and $\sigma^2$ known, $W \sim \chi^2_{\text{rank(Var(z))}}$
  - If $z$ is normal and $\sigma^2$ 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,

\[
(y - \mu_y)' \Sigma_y^{-1} (y - \mu_y) \sim \chi^2_n. \quad \text{--note: } n=\text{rank}(\Sigma_y).
\]

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

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

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

=> $W^* \sim F$ distribution.

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

Recall: An $F$-distributed variable

Let $F = \frac{\chi^2_j}{\chi^2_T} \sim F_{j, T}$

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

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

If $Z$ and $U$ are independent, then

\[
F = \frac{\left[ \sqrt{n} \frac{(x - \mu)}{\sigma} \right]^2}{(n - 1)s^2 / \sigma^2 / (n - 1)} = \frac{(x - \mu)^2}{s^2 / n} \sim F_{1, n-1}
\]
Recall: An $F$-distributed variable

- There is a relationship between $t$ and $F$ when testing one restriction.
  - For a single restriction, $m = rb - q$. The variance of $m$ is: $r \text{Var}[b] r$.
  - The distance measure is $t = m / \text{Est. SE}(m) \sim t_{T-k}$.
  - 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: $H_0: R\beta - q = 0$

- Now, we have $J$ joint hypotheses. Let $R$ be a $J \times k$ matrix and $q$ be a $J \times 1$ vector.

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

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

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

  Under the usual assumption and assuming $\sigma^2$ is known, $W^* \sim \chi^2_j$.

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

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

  \[
  F = W* / [(T-k) \hat{\sigma}^2/(T-k)] = W* / \hat{\sigma}^2 \sim F_{J,T-k'}
  \]
The General Linear Hypothesis: \( H_0: \mathbf{R}\beta - \mathbf{q} = 0 \)

(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) \( e^* = y - Xb^* = e - X(b^* - b) \)
(ii) \( b^* = b - (X'X)^{-1}R'(X'X)^{-1}R^{-1}(Rb - q) \)

\[ e^*e^* = e'e + (b^* - b)'X'X(b^* - b) \]
\[ e^*e^* - e'e = (Rb - q)'[R(X'X)^{-1}R']^{-1}(Rb - q) \]

Recall
- \( \mathbf{W} = (Rb - q)'[R(\sigma^2(X'X)^{-1})'R']^{-1}(Rb - q) \sim \chi^2_j \) (if \( \sigma^2 \) is known)
- \( e'e/\sigma^2 \sim \chi^2_{J-k} \).

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

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

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

Let \( \mathbf{R}^2 = \text{unrestricted model} = 1 - \text{RSS/TSS} \)
\( \mathbf{R}^*^2 = \text{restricted model fit} = 1 - \text{RSS*}/\text{TSS} \)

Then, dividing and multiplying \( F \) by TSS we get
\[ F = ((1 - \mathbf{R}^*^2) - (1 - \mathbf{R}^2))/J / [(1-\mathbf{R}^2)/(T-k)] \sim F_{J,T-K} \]
or
\[ F = \{ (\mathbf{R}^2 - \mathbf{R}^*^2)/J \} / [(1-\mathbf{R}^2)/(T-k)] \sim F_{J,T,K}. \]
Example I: Testing  $H_0: R\beta - q = 0$

- In the linear model  
  \[ y = X \beta + \varepsilon = X_1 \beta_1 + X_2 \beta_2 + \ldots + X_k \beta_k + \varepsilon \]

- We want to test if the slopes $X_2, \ldots, X_k$ 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 $H_0: R\beta - q = 0$
  \[
  \begin{bmatrix}
  0 & 0 & \ldots & 0 \\
  0 & 1 & \ldots & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  0 & 0 & \ldots & 1
  \end{bmatrix}
  \begin{bmatrix}
  \beta_1 \\
  \beta_2 \\
  \vdots \\
  \beta_k
  \end{bmatrix}
  =
  \begin{bmatrix}
  0 \\
  0 \\
  \vdots \\
  0
  \end{bmatrix}
  
  \]

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

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

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

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

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

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

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

- This test statistic is called the \textit{F-test of goodness of fit}. 

\[\text{---}\]
Example I: Testing $H_0: R\beta - q = 0$

- Then, 
  $$ F = \{ \frac{R^2 / (k-1)}{(1 - R^2) / (T-k)} \} \sim F_{k-1,T-k} $$
- Recall $ESS/TSS$ is the definition of $R^2$. $RSS/TSS$ is equal to $(1 - R^2)$.
  $$ F(k-1,n-k) = \frac{R^2 / (k-1)}{(1 - R^2) / (T-k)} = \frac{ESS / (k-1)}{RSS / (T-k)} $$
  $$ = \frac{ESS / (k-1)}{RSS / (T-k)} $$

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

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

- In the linear model 
  $$ y = X \beta + \epsilon = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon $$
- We want to test if the slopes $X_3, X_4$ are equal to zero. That is,
  $$ H_0 : \beta_3 = \beta_4 = 0 $$
  $$ H_1 : \beta_3 \neq 0 \text{ or } \beta_4 \neq 0 \text{ or both } \beta_3 \text{ and } \beta_4 \neq 0 $$
- We can use, 
  $$ F = (e^* e^* - e' e) / J / [e' e / (T-k)] \sim F_{j,T-k} $$
  Define 
  $$ Y = \beta_1 + \beta_2 X_2 + \epsilon \quad RSS_R $$
  $$ Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon \quad RSS_U $$
  $$ F(\text{cost in } df, \text{unconstr } df) = \frac{RSS_R - RSS_U}{RSS_U} \cdot \frac{k_i - k_R}{T - k_U} $$
Lagrange Multiplier Statistics

- Specific to the classical model.

Recall the Lagrange multipliers:
\[ \lambda = [R(X'X)^{-1}R']^{-1} m \]

Suppose we just test \( H_0: \lambda = 0 \), using the Wald criterion.
\[ W' = \lambda'(\text{Var}[\lambda | X])^{-1} \lambda \]

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

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

Application (Greene): Gasoline Demand

- Time series regression,
\[ \log G = \beta_1 + \beta_2 \log Y + \beta_3 \log PG + \beta_4 \log PNC + \beta_5 \log PUC + \beta_6 \log PPT + \beta_7 \log PN + \beta_8 \log PD + \beta_9 \log PS + \epsilon \]


- 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.
Ordinary least squares regression ............

LHS=LG
Mean = 5.39299
Standard deviation = 0.24878
Number of observs. = 36
Model size
Parameters = 9
Degrees of freedom = 27
Residuals
Sum of squares = 0.00855 <*******
Standard error of e = 0.01780 <*******
Fit
R-squared = 0.99605 <*******
Adjusted R-squared = 0.99488 <*******

| Variable | Coefficient | Standard Error | t-ratio | P[|T|>t] | Mean of X |
|----------|-------------|----------------|---------|---------|-----------|
| Constant| -6.95326*** | 1.29811        | -5.356  | .0000   | 9.11093   |
| LY      | 1.35721***  | 0.14562        | 9.320   | .0000   | 9.11093   |
| LPG     | -0.50579*** | 0.06200        | -8.158  | .0000   | 0.67409   |
| LPNC    | -0.01654    | 0.19957        | -0.083  | .9346   | 0.44320   |
| LPUC    | -1.12354*   | 0.06568        | -1.881  | .0708   | 0.66361   |
| LPPT    | 1.10125***  | 0.26840        | 4.103   | .0003   | 0.60539   |
| LPN     | 0.92018***  | 0.27018        | 3.406   | .0021   | 0.43343   |
| LPD     | -1.09213*** | 0.30812        | -3.544  | .0015   | 0.68105   |
| LPS     |             |                |         |         |           |

Application (Greene): Gasoline Demand

• Q: Is the price of public transportation really relevant? H₀ : β₆ = 0.

(1) Distance measure: t₆ = (b₆ - 0) / s₆ = (.11571 - 0) / .07859 = 1.472 < 2.052 => cannot reject H₀.

(2) Confidence interval: b₆ ± t(.95,27) × Standard error
    = .11571 ± 2.052 × (.07859)
    = .11571 ± .16127 = (-.04557 ,.27698)
    => C.I. contains 0 => cannot reject H₀.

(3) Regression fit if X₆ drop? Original R² = .99605,

Without LPPT, R² = .99573
F(1,27) = [(.99605 - .99573)/1]/[(1-.99605)/(36-9)] = 2.187
    = 1.472² (with some rounding) => cannot reject H₀.
Gasoline Demand (Greene) - Hypothesis Test: Sum of Coefficients

- Do the three aggregate price elasticities sum to zero?

$$\text{H}_0 : \beta_7 + \beta_8 + \beta_9 = 0$$

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

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

**Wald Test**

$$\text{Var}(b) = \sum_{i=1}^{9} \sum_{j=1}^{9} \text{Cov}(b_i, b_j) = 0.10107$$

$$m' [\text{Var}(m)]^{-1} m = 6.5446$$

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    |

\[ F = \frac{(.0112599 - .0085531)/1}{.0085531/(36 - 9)} = 8.544691 \]

Gasoline Demand (Greene)- Joint Hypotheses

• Joint hypothesis: Income elasticity = +1, Own price elasticity = -1.
The hypothesis implies that \( \log G = \beta_1 + \log Y - \log P_g + \beta_4 \log P_{NC} + ... \)

Strategy: Regress \( \log G - \log Y + \log P_g \) 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

\[ F = \frac{(.0286877 - .0085531)/2}{.0085531/(36-9)} = 31.779951 \]
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?
### 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|     .35721  |
2|     .49421  |
+-------------+-------+

--> Matrix ; list ; vm = R*varb*R'$

Matrix VM       has  2 rows and  2 columns.

+-----------------+-------+-------+
1|   .02120       | .00291|
2|   .00291       | .00384|
+-----------------+-------+-------+

--> Matrix ; list ; w = 1/2 * m'$vm)m$

Matrix W        has  1 rows and  1 columns.

+-------------+-------+
1|   31.77981  |
+-------------+-------+
```

### Gasoline Demand (Greene)– Testing Details

- Q: Which restriction is the problem? We can look at the Jx1 estimated LM, $\lambda$, for clues:

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

- Recall that under $H_0$, $\lambda$ should be 0.

Matrix Result  has  2 rows and  1 columns.

+-----------------+-------+
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 $R^2$s,

\[
F = \frac{(.9960515 - .997096)/2}{((1-.9960515)/(36-9))} = -3.571166 \, (!)
\]

• Q: What's wrong?