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 hypotheses: \( H_0 \) vs. \( H_1 \).

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

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

\[
\begin{align*}
T(X) \notin R & \Rightarrow \text{Reject } H_0 \text{ (} & \text{we learn } H_0 \text{! is not true).} \\
T(X) \in R & \Rightarrow \text{Fail to reject } H_0. \text{ (No learning.)}
\end{align*}
\]

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

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

Hypothesis Testing: Brief Review - Fisher

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

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

Example: Suppose \( T(X) \sim \chi^2_2 \). We compute \( T(X) = 7.378 \). Then,

\[
p-value(T(X) = 7.378) = 1 - \text{Prob}[T(X) < 7.378] = 0.025
\]
Hypothesis Testing: Brief Review - Fisher

• Fisher’s Idea.
Steps for testing
1. Form \(H_0\) and set a significance level, \(\alpha\).
2. a Collect a sample of data \(X = \{X_1, \ldots, X_n\}\).
   b Compute the test-statistics \(T(X)\) used to test \(H_0\).
3. Report the \(p\)-value - i.e., the probability, of observing a result at least as extreme as the test statistic, under \(H_0\).
4. Decision rule:
   If the \(p\)-value < \(\alpha\) \(\Rightarrow\) result is significant & \(H_0\) is rejected.
   If the \(p\)-value > \(\alpha\) \(\Rightarrow\) result is “not significant.” No conclusions are reached. Gather more data/modify model.

Note: By setting \(\alpha\), we determine \(R\).

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

1. \(H_0: \mu_{NL} = 15\) & set \(\alpha = .05\).
2a. Get a sample: \(\{X_{1962}, X_{1963}, \ldots, X_{N=2022}\}\), with \(N=61\).
2b. We use \(T(X) = \bar{X}\), which is unbiased, consistent, and, assuming \(X\) is normally distributed, we know its distribution, \(\bar{X} \sim N(\mu, \sigma^2/N)\).
3. Compute \(\bar{X} = 36\) & \(p\)-value(\(\bar{X} = 36\)) = .005.
4. Decision Rule: \(p\)-value < \(\alpha\) \(\Rightarrow\) result is significant & \(H_0\) is rejected.

Instead of using a \(p\)-value, it is common to use a rejection region, \(R\):
\[T(X) = \bar{X} \notin [T_{LB}, T_{UB}] \Rightarrow \text{Reject } H_0; \mu_{NL} = 15.\]
Hypothesis Testing

Example (continuation): That is,

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

- The blue area is the significance level, \( \alpha \).

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, $\theta$:

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

Note: when $\theta \in H_1$ $\Rightarrow \pi(\theta) = 1 - \beta(\theta)$ - the usual application.
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>
<th>H(_0) true</th>
<th>H(_1) true (H(_0) false)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cannot reject ( H_0 )</td>
<td>Correct decision</td>
<td>Type II error</td>
<td></td>
</tr>
<tr>
<td>Reject ( H_0 )</td>
<td>Type I error</td>
<td>Correct decision</td>
<td></td>
</tr>
</tbody>
</table>

Need to control both types of error:

\[
\alpha = P(\text{rejecting } H_0 | H_0) \leq \text{Reject } H_0 \text{ by “accident” or luck (a false positive).}
\]

\[
\beta = P(\text{not rejecting } H_0 | H_1) \leq 1 - \beta = \text{Power of test (under } H_1)\text{.}
\]
Type I and Type II Errors

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

\[ \pi = 1 - \beta = \text{Power of test (under } H_1) \]

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>855</th>
<th>50 (Type II error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cannot reject ( H_0 )</td>
<td>( H_0 ) true</td>
<td>( H_1 ) true (( H_0 ) false)</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

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

<table>
<thead>
<tr>
<th>State of World</th>
<th>H₀ true</th>
<th>H₁ true (H₀ false)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cannot reject $H_0$</td>
<td>855</td>
<td>20 (Type II error)</td>
</tr>
<tr>
<td>Reject $H_0$</td>
<td>45 (Type I error)</td>
<td>80</td>
</tr>
</tbody>
</table>

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

Type I and Type II Errors - Example

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

<table>
<thead>
<tr>
<th>State of World</th>
<th>H₀ true</th>
<th>H₁ true (H₀ false)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cannot reject $H_0$</td>
<td>475</td>
<td>100 (Type II error)</td>
</tr>
<tr>
<td>Reject $H_0$</td>
<td>25 (Type I error)</td>
<td>400</td>
</tr>
</tbody>
</table>

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

Conclusion: The proportion of false positives depends on percentage of false $H_0$ and the power of test. Higher power, lower proportion.
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>20%</td>
<td></td>
<td>5.9</td>
</tr>
<tr>
<td>80%</td>
<td></td>
<td>2.4</td>
</tr>
<tr>
<td>50%</td>
<td></td>
<td>1.5</td>
</tr>
<tr>
<td>10%</td>
<td></td>
<td>20.0</td>
</tr>
<tr>
<td>50%</td>
<td></td>
<td>9.1</td>
</tr>
<tr>
<td>80%</td>
<td></td>
<td>5.9</td>
</tr>
<tr>
<td>10%</td>
<td></td>
<td>69.2</td>
</tr>
<tr>
<td>50%</td>
<td></td>
<td>47.4</td>
</tr>
<tr>
<td>80%</td>
<td></td>
<td>36.0</td>
</tr>
<tr>
<td>1%</td>
<td></td>
<td>96.1</td>
</tr>
<tr>
<td>50%</td>
<td></td>
<td>90.8</td>
</tr>
<tr>
<td>80%</td>
<td></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 $R_i$ of size $\alpha$, $\beta(R) \leq \beta(R_i)$.

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 $R_i$ of level $\alpha$ (that is, $\alpha(R_i) \leq \alpha$), if $\beta(R) \leq \beta(R_i)$.

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 $R_i$ of level $\alpha$ (that is, $\alpha(R_i) \leq \alpha$), if $\pi(R) \geq \pi(R_i)$.

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

- 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.
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 $L(x | \theta_1)$ 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:

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

Then,

$$\ln \lambda = \sum_i T(X_i) [A(\theta_1) - A(\theta_0)] + n B(\theta_1) - n B(\theta_0).$$

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

$$\Rightarrow \delta \ln \lambda / \delta u = [A(\theta_1) - A(\theta_0)] > 0,$$

if $A(.)$ is monotonic in $\theta$.

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

• Some distributions with MLRP in $T(X) = \sum_i X_i$: 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 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 $\sup_{\theta \in H_0} \pi(\theta) \leq \alpha$) 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 ($\theta - \theta_0$, $n$, & $\alpha$):
  – **Effect size**: True value ($\theta$) – 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 $\alpha$.

**Example**: We randomly collect 20 stock returns ($n = 20$), 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 $\theta = 0.2$ ($H_1$: $\theta = 0.2$ is true)?

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

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

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

• Changing $\theta - \theta_0$

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

$$\Rightarrow\text{Power} = P[X \in R | H_1] = P[\zeta \geq (.1736 - 0.3)/(.2/sqrt(20))] = P[\zeta \geq - 2.82713] = 0.997652$$
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: $x \geq 0.2042 = 2.33 \cdot \frac{0.2}{\sqrt{20}} + 0.1$

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

= $P[z \geq 0.093915] = 0.46259$

• Changing $n$ ($\theta_1 = 0.2; \alpha = 0.05$)

If $n = 200$, then rejection rule: $x \geq 0.12332 = 1.645 \cdot \frac{0.2}{\sqrt{200}} + 0.1$

$\Rightarrow$ Power = $P[X \in R | H_1] = P[x \geq (0.12323 - 0.2)/(2/sqrt(200))]$

= $P[z \geq -5.4261] = 0.9999999$

Note: We can select $n$ to achieve a given power (for given $\theta_i$ & $\alpha$). 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.

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

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 $\chi_p^2$, thus $-2 \ln k$ needs to be the $1 - \alpha$ percentile of a $\chi_p^2$. We need not solve explicitly for $k$, we just need the rejection rule.

---

Likelihood Ratio Tests: Example II

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

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

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

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) Find an appropriate test and know its distribution under \( H_0 \)
  (5) Decision Rule (Test \( H_0 \)). Reject \( H_0 \) if \( \theta_0 \) is too far from \( \hat{\theta} \) (“the hypothesis is inconsistent with the sample evidence.”)

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

• To determine if the statistic is “large,” we need a null distribution.

• Ideally, we use a test that is most powerful to test \( H_0 \).
Hypothesis Testing: Issues

• Logic of the Neyman-Pearson methodology:
If $H_0$ is true, then $T(X)$ will have a certain distribution (under $H_0$). We call this distribution **null distribution** or **distribution under the null**.

• It tells us how likely certain values are, if $H_0$ is true. Thus, we expect ‘large values’ for $\theta_0$ to be unlikely.

• **Decision rule.**
If the observed value for $T(X)$ falls in rejection region $R$
$\Rightarrow$ Assumed distribution must be incorrect: $H_0$ should be rejected.

---

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.
We test a hypothesis about a single parameter, say \( \beta_k \), of the DGP.

**Example:** The linear model (DGP): \( y = X\beta + \epsilon \)

1. Formulate \( H_0: X_k \) should not be in the DGP \( \Rightarrow H_0: \beta_k = \beta_k^0 \)
   \( H_1: \beta_k \neq \beta_k^0 \).

2. Construct \( T(X) \) test \( H_0: t_k = \frac{(b_k - \beta_k^0)}{\sqrt{s^2(X'X)_{kk}^{-1}}} \)
   Distribution of \( T(X) \) under \( H_0 \) with \( s^2 \) estimating \( \sigma^2 \) (unknown):
   If (A5) \( \epsilon | X \sim N(0, \sigma^2I_T) \), \( \Rightarrow t_k \sim t_{T-k} \).
   If (A5) not true, asymptotic results: \( \Rightarrow t_k \rightarrow \text{N}(0, 1) \).

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

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

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

- Special case: \( H_0: \beta_k = 0 \)
  \( H_1: \beta_k \neq 0 \).

Then,

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

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

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

- Usually, \( \alpha = 5\% \), then if \( |t_k| > 1.96 \approx 2 \), we say the coefficient \( b_k \) is “significant.”
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 pre-specified target, usually 90% or 95%. $C_n$ is called a $(1 - \alpha)%$ C.I.

• When we know the distribution for the point estimate, it is easy to construct a C.I. For example, under (A5), the distribution of $b$ is normal, then a 95% C.I. is given by:
  \[ C_n = [b_k - z_{0.025} \times \text{Estimated SE}(b_k), \ b_k + z_{0.025} \times \text{Estimated SE}(b_k)] \]

• 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, 
  \[ C_n = \{ b_k : | z_n(b_k) | \leq z_{\alpha/2} \} \]  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.

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 \( \chi^2 \_\nu \) RV divided by \( \nu \).

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

Let 
\[
U = \frac{(n-1)s^2}{\nu} \sim \chi^2_{\nu-1}
\]

Assume that \( Z \) and \( U \) are independent (check the middle matrices in the quadratic forms!). Then,
\[
t = \frac{(\bar{x} - \mu)}{\sqrt{n} \frac{s}{\sqrt{n}}} = \frac{(\bar{x} - \mu)}{s} \sim t_{\nu-1}
\]
Hypothesis Testing: Testing Example in R

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

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

b <- solve(t(x)%*% x)%*% t(x)%*%y # b = (X'X)^(-1)X'y (OLS regression)

c <- y - x%*%b # regression residuals, e

RSS <- as.numeric(t(e)%*%e) # RSS

R2 <- 1 - as.numeric(RSS)/as.numeric(t(y)%*%y) # R-squared

Sigma2 <- as.numeric(RSS/(T-k)) # Estimated \sigma^2 = \hat{s}^2

SE_reg <- sqrt(Sigma2) # Estimated \sigma – Regression stand error

Var_b <- Sigma2*solve(t(x)%*% x) # Estimated \text{Var}[b|X] = \hat{s}^2 (X'X)^{-1}

SE_b <- sqrt(diag(Var_b)) # SE[b|X]

t_b <- b/SE_b # t-stats (See Chapter 4)
```

```r
> t(b)

     Mkt_RF    SMB   HML
[1,] -0.005088944 0.9082989 -0.2124596 -0.1715002

> t(SE_b)

     Mkt_RF    SMB   HML
[1,] 0.002487509 0.05672206 0.08411188 0.08468165

> t(t_b)

     Mkt_RF    SMB   HML
[1,]  -2.045799 16.01315 -2.525917 -2.025235
```

▶ all coefficients are significant (|t|>2).

• Q: Is the market beta (\beta_1) equal to 1? That is,

\[ H_0: \beta_1 = 1 \] vs. \[ H_1: \beta_1 \neq 1 \]

\[ t_k = (b_k - \bar{\beta}_k)/\text{Est. SE}(b_k) \]

\[ t_1 = (0.9082989 - 1)/0.05672206 = -1.616674 \]

\[ |t_1| < 1.96 \implies \text{Cannot reject } H_0 \text{ at 5% level} \]

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

```r
> t(b)

     Mkt_RF    SMB   HML
[1,] -0.005088944 0.9082989 -0.2124596 -0.1715002

> t(SE_b)

     Mkt_RF    SMB   HML
[1,] 0.002487509 0.05672206 0.08411188 0.08468165

> t(t_b)

     Mkt_RF    SMB   HML
[1,]  -2.045799 16.01315 -2.525917 -2.025235 \implies \text{all coefficients are significant (}|t|>2)\]"
Testing: The Expectation Hypothesis (EH)

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

Implication of EH: \( S_{t+T} - F_{t,T} = \text{unpredictable} \).

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

Empirical tests of the EH are based on a regression:

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

where \( Z_t \) represents any economic variable that might have power to explain \( S_t \), for example, \((i_t - i)\).

Then, under EH, \( H_0: \alpha = 0 \text{ and } \beta = 0 \).

vs \( H_1: \alpha \neq 0 \text{ and/or } \beta \neq 0 \).

Testing: The Expectation Hypothesis (EH)

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

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

53
### Testing: The Expectation Hypothesis (EH)

**Example (continuation):** We do two individual t-tests on $\alpha$ & $\beta$.

```r
> summary(fit)
Call:
  lm(formula = y ~ x)
Residuals:
   Min      1Q  Median      3Q     Max
-0.125672 -0.014576 -0.000439  0.017356  0.094283

Coefficients:  
Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.0001854 0.0016219  -0.114  0.90906
x           -0.2157540  0.0731553 -2.949  0.00339 **
```

- $\alpha$: Constant not significant ($|t|<2$).
- $\beta$: Slope is significant ($|t|>2$).

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

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

![Image](image.png)

- 95% C.I. for $b$:

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

Then,

\[
C_n = [-0.215754 - 1.96 \times 0.0731553, -0.215754 + 1.96 \times 0.0731553] = [-0.3591384, -0.07236961]
\]

Since $\beta = 0$ is not in $C_n$ with 95% confidence $\Rightarrow$ Reject $H_0$: $\beta_1 = 0$

at 5% level.

**Note:** The EH is a joint hypothesis, it should be tested with a joint test!
Testing a Hypothesis: Wald Statistic

- Most of our test statistics, including joint tests, are Wald statistics.
  Wald = normalized distance measure:
  One parameter: \( t_b = (b_k - \beta_0^k) / s_{b_k} = \text{distance/unit} \)
  More than one parameter.

Let \( z = (\text{random vector} - \text{hypothesized value}) \) be the distance

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

- Distribution of \( W \)? We have a quadratic form.
  - If \( z \) is normal and \( \sigma^2 \) known, \( W \sim \chi^2_{\text{rank}(\text{Var}(z))} \)
  - If \( z \) is normal and \( \sigma^2 \) unknown, \( W \sim F \)
  - If \( z \) is not normal and \( \sigma^2 \) unknown, we rely on asymptotic theory, \( W \xrightarrow{d} \chi^2_{\text{rank}(\text{Var}(z))} \)

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,

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

\( \Rightarrow \) If \( z \sim N(0, \text{Var}(z)) \) \( W \) 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 \( \text{Var}(z) \)

\( \Rightarrow \) \( W \sim F \) 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 $F = \frac{\chi^2_J / J}{\chi^2_T / T} \sim F_{J,T}$

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

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

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

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

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: $r \text{ Var}[b] r$.
  - The distance measure is $t = m / \text{Est. SE}(m) \sim t_{T-k}$.
  - This $t$-ratio is the $\sqrt{t}$ of $F$-ratio.

- $t$-ratios are used for individual restrictions, while $F$-ratios are used for joint tests of several restrictions.
The General Linear Hypothesis: $H_0: R\beta - q = 0$

- Suppose we are interested in testing $J$ joint hypotheses.

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

We can write linear restrictions as $H_0: R\beta - q = 0$, where $R$ is a $J \times k$ matrix and $q$ a $J \times 1$ vector.

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

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

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

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

$$m = Rb - q.$$  
Then, we construct a Wald statistic:

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

to test if $m$ is different from 0.

(2) We base the answer on a model loss of fit when restrictions are imposed: RSS must increase and $R^2$ must go down. Then, we construct an F test to check if the unrestricted RSS ($RSS_U$) is different from the restricted RSS ($RSS_R$).
The General Linear Hypothesis: \( H_0: R\beta - q = 0 \)

- Q: Is \( Rb - q \) close to 0? There are two different approaches to this question. Both have in common the property of unbiasedness for \( b \).

1. We base the answer on the discrepancy vector:

\[ m = Rb - q. \]

Then, we construct a Wald statistic:

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

to test if \( m \) is different from 0.

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

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

- To test \( H_0 \), we calculate the discrepancy vector:

\[ m = Rb - q. \]

Then, we compute the Wald statistic:

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

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

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

Under \( H_0 \), assuming \((A5)\) & estimating \( \sigma^2 \) with \( s^2 = e'e/(T - k) \):

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

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

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

If \((A5)\) is not assumed, the results are only asymptotic: \( F^d \rightarrow \chi_j^2 \).
Wald Test Statistic for $H_0$: $R\beta - q = 0$

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

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

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

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

2. Test-statistic: $F = \frac{(R\beta - q)'(R\beta - q)}{J} \cdot \frac{N(0, \sigma^2 I)}{\sum \frac{\mu^2}{\sigma^2}} 
\Rightarrow F_{J=2, T-2}$

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

3. Get OLS results, compute F.

**Example (continuation):**

4. **Decision Rule**: $\alpha = 0.05$ level. We reject $H_0$ if $p$-value($F$) < .05.

Or, reject $H_0$, if $F > F_{J=2,T-2,0.05}$

```r
J <- 2 # number of restrictions
R <- matrix(c(0,0,0,0,1,0,0,1), nrow=2) # matrix of restrictions
q <- c(0.2,1) # hypothesized values
m <- R%*%b - q # m = Estimated R*Beta - q
Var_m <- R%*%Var_b%*%t(R) # Variance of m
det(Var_m) # check for non-singularity
W <- t(m)%*%solve(Var_m)%*%m # F-test statistic
F_t <- as.numeric(W/J)
qf(.95, df1=J, df2=(T - k)) # exact distribution (F-dist) if errors normal
p_val <- 1 - pf(F_t, df1=J, df2=(T - k)) # p-value(F_t) under errors normal
p_val
```

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

**Example (continuation):**

```r
> F_t
[1] 49.21676
>
> qf(.95, df1=J, df2=(T - k)) # exact distribution (F-dist) if errors normal
[1] 3.011672
F_t > 3.011672 \Rightarrow reject H_0 at 5% level

> p_val <- 1 - pf(F_t, df1=J, df2=(T - k)) # p-value(F_t) under errors normal
> p_val
[1] 0.00984081
very low chance H_0 is true.
```

Wald Test Statistic for $H_0$: Does EH hold?

**Example:** Now, we do a joint test of the EH. $H_0$: $\alpha = 0$ and $\beta = 0$.

Using the previous program but with:

```r
J <- 2 # number of restriction
R <- matrix(c(1,0,0,1), nrow=2) # matrix of restrictions
q <- c(0,0) # hypothesized values
> F_t
[1] 4.1024
>
> qf(.95, df1=J, df2=(T - k)) # exact distribution (F-dist) if errors normal
[1] 3.020661
F_t > 3.020661 \Rightarrow reject H_0 at 5% level

> p_val <- 1 - pf(F_t, df1=J, df2=(T - k)) # p-value(F_t) under errors normal
> p_val
[1] 0.01731
very low chance H_0 is true.
```
The F Test: $H_0: R\beta - q = 0$

(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) $e^* = y - Xb^* = e - X(b^* - b)$

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

$\Rightarrow e^*e^* = e'e + (b^* - b)'X'X(b^* - b)$

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

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

Recall

$W = (Rb - q)'[R[\sigma^2(X'X)^{-1}]R']^{-1}(Rb - q) \sim \chi^2_j$ (if $\sigma^2$ is known)

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

Then,

$F = (e^*e^* - e'e)/J / [e'e/(T - k)] \sim F_{J,T-k}$. 69

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

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

Let $R^2 = \text{unrestricted model} = 1 - \text{RSS}/TSS$

$R^{*2} = \text{restricted model fit} = 1 - \text{RSS}^*/TSS$

Then, dividing and multiplying $F$ by TSS we get

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

or

$F = \left\{ (R^2 - R^{*2})/J \right\} / \left[(1 - R^2) / (T - k)\right] \sim F_{J,T-k}$. 70
The F Test: F-test of goodness of fit

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

  • We want to test if the slopes \( X_2, \ldots, X_k \) are equal to zero. That is,
    \[ H_0: \beta_2 = \ldots = \beta_k = 0 \]
    \[ H_1: \text{at least one } \beta \neq 0 \]
    \[ \Rightarrow J = k - 1 \]

  • We can write \( H_0: R\beta - q = 0 \)
    \[ \Rightarrow \begin{bmatrix} 0 & 1 & \ldots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \]

  • We have \( J = k - 1 \). Then,
    \[ F = \left\{ \frac{R^2 - R^*2}{(k - 1)} \right\} / \left\{ \frac{(1 - R^2)/(T - k)}{[(1 - R^2)/(T - k)]} \right\} \sim F_{k-1,T-k} \]
    \[ \Rightarrow \frac{F(k - 1, T - k)}{F_{k-1,T-k}} \sim \frac{ESS / (k - 1)}{RSS / (T - k)} \]

• For the restricted model, \( R^*2 = 0 \).

The F Test: F-test of goodness of fit

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 \( (1 - R^2) \).

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

• This test statistic is called the F-test of goodness of fit.
The F Test: F-test of goodness of fit

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

\[ y <- ibm_x \]
\[ T <- \text{length}(x) \]
\[ x0 <- \text{matrix}(1,T,1) \]
\[ x <- \text{cbind}(x0,Mkt_RF, SMB, HML) \]
\[ k <- \text{ncol}(x) \]
\[ b <- \text{solve}(t(x)%*%x)%*%t(x)%*%y \] #OLS regression
\[ e <- y - x%*%b \]
\[ \text{RSS} <- \text{as.numeric}(t(e)%*%e) \]
\[ R2 <- 1 - \text{as.numeric(RSS)/as.numeric(t(y)%*%y)} \] #R-squared
\[ > R2 \]
\[ [1] 0.338985 \]
\[ \text{F_goodfit} <- (R2/(k-1))/((1-R2)/(T-k)) \] #F-test of goodness of fit.
\[ > \text{F_goodfit} \]
\[ [1] 96.58204 \]
\[ \Rightarrow \text{F_goodfit} > F_{2,565,.05} = 2.387708 \Rightarrow \text{Reject } H_0. \]

The F Test: General Case - Example

- In the linear model
  \[ y = X \beta + \epsilon = \beta_1 + X_2 \beta_2 + X_3 \beta_3 + X_4 \beta_4 + \epsilon \]

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

Define
\[ Y = \beta_1 + \beta_2 X_2 + \epsilon \]
\[ RSS_R \]
\[ Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon \]
\[ RSS_U \]
\[ F = \frac{RSS_R - RSS_U / k_U - k_R}{RSS_U / (T-k_U)} \] 74

37
### The F Test: General Case - Example

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

Unrestricted Model:

(U) \[ \text{IBM}_t - r_t = \beta_0 + \beta_1 (\text{Mkt}_t - r_t) + \beta_2 \text{SMB} + \beta_3 \text{HML} + \epsilon \]

Hypothesis:  

- \( H_0: \beta_2 = \beta_3 = 0 \)  
- \( H_1: \beta_2 \neq 0 \) and/or \( \beta_3 \neq 0 \)

Then, the Restricted Model:

(R) \[ \text{IBM}_t - r_t = \beta_0 + \beta_1 (\text{Mkt}_t - r_t) + \epsilon \]

Test: \[ F = \frac{(\text{RSS}_R - \text{RSS}_U)/J}{\text{RSS}_U/(T-k_R)} \sim F_{J,T-K} \] with \( J = k_U - k_R = 4 - 2 = 2 \)

---

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

```r
y <- ibm_x
x0 <- matrix(1,T,1)
x_r <- cbind(x0,Mkt_RF)  # Restricted X vector
T <- nrow(x)
k2 <- ncol(x)
b2 <- solve(t(x_r)%*% x_r)%*% t(x_r)%*%y  # Restricted OLS regression
e2 <- y - x_r%*%b2
RSS2 <- as.numeric(t(e2)%*%e2)
> RSS = 1.932442 # RSS_U
> RSS2 = 1.964844 # RSS_R
J <- k - k2  # J = degrees of freedom of numerator
F_test <- ((RSS2 - RSS)/J)/(RSS/(T-k))
```

---

...continued...
The F Test: General Case - Example

Example (continuation):

\[ F_{\text{test}} = \frac{(RSS_2 - RSS)/J}{RSS/(T-k)} \]

\[ F_{\text{test}} \]

\[ > F_{\text{test}} \]

\[ [1] 4.736834 \]

\[ > qf(.95, df1=J, df2=(T-k)) \] # F 2,565,.05 value (≈ 3)


\[ p_{\text{val}} = 1 - pf(F_{\text{test}}, df1=J, df2=(T-k)) \] # p-value of F_{\text{test}}

\[ > p_{\text{val}} \]

\[ [1] 0.009117494 \] ⇒ p-value is small ⇒ Reject H₀.

Lagrange Multiplier Statistics

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

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

Suppose we just test H₀: \( \lambda = 0 \), using the Wald criterion.

\[ l' = \lambda' (\text{Var}[\lambda | X])^{-1} \lambda \]

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,

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

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

---

### Application (Greene): Gasoline Demand

<table>
<thead>
<tr>
<th>Ordinary least squares regression ............</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>LHS=LG</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>5.39299</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>.24878</td>
</tr>
<tr>
<td>Number of observs.</td>
<td>36</td>
</tr>
<tr>
<td>Model size</td>
<td>9</td>
</tr>
<tr>
<td>Degrees of freedom</td>
<td>27</td>
</tr>
<tr>
<td>Residuals</td>
<td>0.00855 &lt;*******</td>
</tr>
<tr>
<td>Standard error of e</td>
<td>.01780 &lt;*******</td>
</tr>
<tr>
<td>Fit</td>
<td>.99605 &lt;*******</td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>.99488 &lt;*******</td>
</tr>
</tbody>
</table>

| 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***  | .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    | .44320    |
| 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: $t_6 = (b_6 - 0) / s_{b_6} = (.11571 - 0) / .07859$
  
  \[ = 1.472 < 2.052 \Rightarrow \text{cannot reject } H_0. \]

  (2) Confidence interval: $b_6 \pm t_{(0.95,27)} \times \text{Standard error}$
  
  \[ = .11571 \pm 2.052 \times (.07859) \]
  
  \[ = .11571 \pm .16127 = (-.045557 ,.27698) \Rightarrow \text{C.I. contains 0} \Rightarrow \text{cannot reject } H_0. \]

  (3) Regression fit if $X_6$ drop? Original $R^2 = .99605,$
  
  Without LPPT, $R^*2 = .99573$
  
  \[ F(1,27) = \frac{(.99605 - .99573)/1}{(1 - .99605)/(36 - 9)} = 2.187 \]
  
  \[ = 1.472^2 \text{ (with some rounding)} \Rightarrow \text{cannot reject } H_0. \]

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

- Do the three aggregate price elasticities sum to zero?
  
  $H_0 : \beta_7 + \beta_8 + \beta_9 = 0$
  
  $R = [0, 0, 0, 0, 0, 0, 1, 1, 1], \quad q = 0$

| Variable | Coefficient | Standard Error | t-ratio | P[|T|>t] |
|----------|-------------|----------------|---------|----------|
| LPN      | 1.10125***  | .26840         | 4.103   | .0003    | .60539   |
| LPD      | .92018***    | .27018         | 3.406   | .0021    | .43343   |
| LPS      | -1.09213***  | .30812         | -3.544  | .0015    | .68105   |
Gasoline Demand (Greene) - Hypothesis Test: Sum of Coefficients – Wald Test

\[ m' [\text{Var}(m)]^{-1} m = \mathbf{8.5446} \]

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

| 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      | .68103    |

\[ F = \frac{(.0112599 - .0085531)}{1} / \frac{.0085531/(36 - 9)} = 8.544691 \]
Gasoline Demand (Greene)- Joint Hypotheses

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

**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

\[ F = \frac{(0.0286877 - 0.0085531)/2}{0.0085531/(36-9)} = 31.779951 \]
  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  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 $J \times 1$ 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.

<table>
<thead>
<tr>
<th></th>
<th>Income elasticity</th>
<th>Price elasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-.88491</td>
<td>129.24760</td>
</tr>
</tbody>
</table>

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?