# Lecture 3
# Specification & Testing in the Classical Linear Model

1

## OLS Estimation - Assumptions

• CLM Assumptions

(**A1**) DGP: $\mathbf{y} = \mathbf{X}\,\beta + \boldsymbol{\varepsilon}$ is correctly specified.

(**A2**) $E[\boldsymbol{\varepsilon}|\mathbf{X}] = 0$

(**A3**) $Var[\boldsymbol{\varepsilon}|\mathbf{X}] = \sigma^2\,\mathbf{I}_T$

(**A4**) $\mathbf{X}$ has full column rank –rank($\mathbf{X}$)=$k$-, where $T \geq k$.

Q: What happens when (**A1**) is not correctly specified?

• In this lecture, we look at (**A1**), always in the context of linearity. Are we omitting a relevant regressor? Are we including an irrelevant variable? What happens when we impose restrictions in the DGP?

## Specification Errors: Omitted Variables

• Omitting relevant variables: Suppose the correct model is

$$\boldsymbol{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} \quad \text{–i.e., with two sets of variables.}$$

But, we compute OLS omitting $\mathbf{X}_2$. That is,

$$\boldsymbol{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon} \qquad \text{<= the "short regression."}$$

Some easily proved results:

(1) $E[\mathbf{b}_1|\mathbf{X}] = E[(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\boldsymbol{y}] = \boldsymbol{\beta}_1 + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\boldsymbol{\beta}_2 \neq \boldsymbol{\beta}_1$. So, unless $\mathbf{X}_1'\mathbf{X}_2 = 0$, $\mathbf{b}_1$ is *biased*. The bias can be huge. It can reverse the sign of a price coefficient in a "demand equation."

(2) $\text{Var}[\mathbf{b}_1|\mathbf{X}] \leq \text{Var}[\mathbf{b}_{1.2}|\mathbf{X}]$. (The latter is the northwest submatrix of the full covariance matrix.) The proof uses **M**, the residual maker. We get a smaller variance when we omit $\mathbf{X}_2$.

## Specification Errors: Omitted Variables

• We get a smaller variance when we omit $\mathbf{X}_2$.

Interpretation: Omitting $\mathbf{X}_2$ amounts to using extra information –i.e., $\boldsymbol{\beta}_2 = \mathbf{0}$. Even if the information is wrong, it reduces the variance.

(3) MSE

$\mathbf{b}_1$ may be more "precise."

Precision = Mean squared error

= variance + squared bias.

Smaller variance but positive bias. If bias is small, may still favor the short regression.

Note: Suppose $\mathbf{X}_1'\mathbf{X}_2 = \mathbf{0}$. Then the bias goes away. Interpretation, the information is not "right," it is irrelevant. $\mathbf{b}_1$ is the same as $\mathbf{b}_{1.2}$.

## Omitted Variables Example: Gasoline Demand

• We have a linear model for the demand for gasoline:

$$\mathbf{G} = \text{PG}\, \boldsymbol{\beta}_1 + \text{Y}\, \boldsymbol{\beta}_2 + \boldsymbol{\varepsilon},$$

Q: What happens when you wrongly exclude Income (Y)?

$$\text{E}[b_1|\mathbf{X}] = \beta_1 + \frac{Cov[Price, Income]}{Var[Price]}\,\beta_2$$

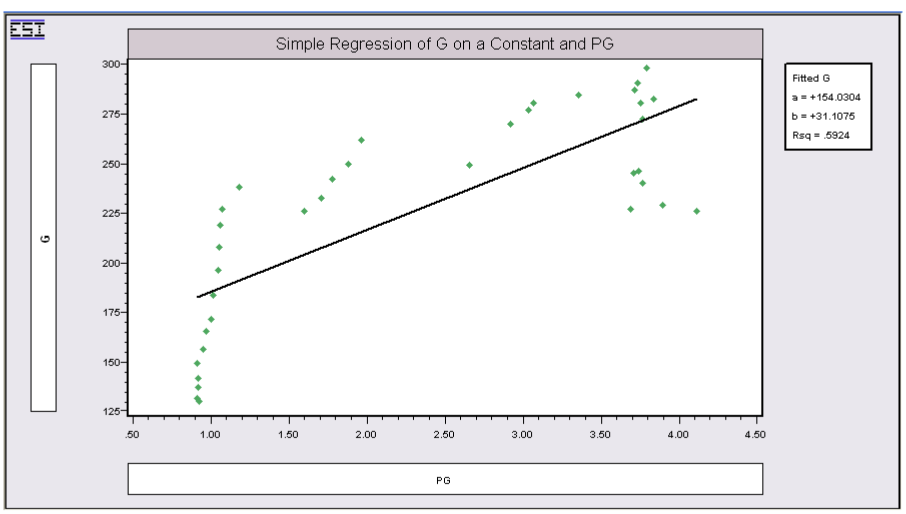
In time series data, $\beta_1 < 0,\ \beta_2 > 0$ (usually)

Cov[*Price*, *Income*] $>$ 0 in time series data.

$\Rightarrow$ The short regression will overestimate the price coefficient.

In a simple regression of G (demand) on a constant and PG, the Price Coefficient ($\beta_1$) should be negative.

## Estimation of a 'Demand' Equation (Greene): Shouldn't the Price Coefficient be Negative?



## Estimation of a 'Demand' Equation (Greene): Multiple Regression - Theory Works.

```
Ordinary     least squares regression ............
LHS=G        Mean                 =      226.09444
             Standard deviation   =       50.59182
             Number of observs.   =             36
Model size   Parameters           =              3
             Degrees of freedom   =             33
Residuals    Sum of squares       =     1472.79834
             Standard error of e  =        6.68059
Fit          R-squared            =         .98356
             Adjusted R-squared   =         .98256
Model test   F[  2,    33] (prob) =   987.1(.0000)
--------+--------------------------------------------------------
Variable| Coefficient    Standard Error  t-ratio  P[|T|>t]
--------+--------------------------------------------------------
Constant|   -79.7535***      8.67255       -9.196   .0000
       Y|     .03692***       .00132       28.022   .0000
      PG|   -15.1224***      1.88034       -8.042   .0000
--------+--------------------------------------------------------
```

• Note: Income is helping us to identify a demand equation –i.e., with a negative slope for the price variable.

## Specification Errors: Irrelevant Variables

• Irrelevant variables . Suppose the correct model is

$$\boldsymbol{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon} \qquad \text{–i.e., with one set of variables.}$$

But, we estimate

$$\boldsymbol{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} \qquad \text{<= the "long regression."}$$

Some easily proved results: Including irrelevant variables just reverse the results: It increases variance -the cost of not using information-; but does not create biases.

$\Rightarrow$ Since the variables in $\mathbf{X}_2$ are truly irrelevant, then $\boldsymbol{\beta}_2 = \mathbf{0}$, so $E[\mathbf{b}_{1.2}|\mathbf{X}] = \boldsymbol{\beta}_1$.

## Specification Errors: Irrelevant Variables

• A simple example

Suppose the correct model is: $\mathbf{y} = \beta_1 + \beta_2 \mathbf{X}_2 + \boldsymbol{\varepsilon}$

But, we estimate: $\mathbf{y} = \beta_1 + \beta_2 \mathbf{X}_2 + \beta_3 \mathbf{X}_3 + \boldsymbol{\varepsilon}$

• Unbiased: given that $\beta_3 = 0$ $\Rightarrow$ $E[b_2|X] = \beta_2$

• Efficiency:

$$\sigma_{b_2}^2 = \frac{\sigma^2}{\sum (X_{2i} - \bar{X}_2)^2} \times \frac{1}{1 - r_{X_2,X_3}^2} > \frac{\sigma^2}{\sum (X_{2i} - \bar{X}_2)^2}$$

Note: These are the results in general. Note that if $X_2$ and $X_3$ are uncorrelated, there will be no loss of efficiency after all.

9

## Other Models

• Looking ahead to nonlinear models: neither of the preceding results extend beyond the linear regression model.

> "Omitting relevant variables from a model is always costly. (No exceptions.) The benign result above almost never carries over to more involved nonlinear models." (Greene)

## Specification and Functional Form: Non-linearity

• In the context of OLS estimation, we can introduce some non-linearities: quadratic, cubic and interaction effects can be easily estimated by OLS. For example:

$$\mathbf{y} = \beta_1 + \beta_2\, \mathbf{X}_2 + \beta_3\, \mathbf{X}_2^2 + \beta_4\, \mathbf{X}_2\, \mathbf{X}_3 + \boldsymbol{\varepsilon}$$

• Partial effects , $\partial\mathbf{y}/\partial\mathbf{X}_2$, (and standard errors) can be different. In the above model

$$\partial\mathbf{y}/\partial\mathbf{X}_2 = \beta_2 + 2\,\beta_3\, \mathbf{X}_2 + \beta_4\, \mathbf{X}_3 \neq \beta_2$$

Note: Recall that in a simple linear model:

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

the partial effect is equal to the $\beta_i$ coefficient:

$$\partial\mathbf{y}/\partial\mathbf{X}_2 = \beta_2.$$

## Specification and Functional Form: Non-linearity

• The estimator of partial effects and their variances are different from $b_i$ and $\text{Var}[b_i|\mathbf{X}]$ in the presence of non-linearities

**Example**: Quadratic Effect

| Population | Estimators |
|---|---|
| $y = \beta_1 + \beta_2 x + \beta_3 x^2 + \beta_4 z + \varepsilon$ | $\hat{y} = b_1 + b_2 x + b_3 x^2 + b_4 z$ |
| $\delta_x = \dfrac{\partial E[y \mid x, z]}{\partial x} = \beta_2 + 2\beta_3 x$ | $\hat{\delta}_x = b_2 + 2b_3 x$ |

Estimator of the variance of $\hat{\delta}_x$

$$Est.Var[\hat{\delta}_x] = Var[b_2] + 4x^2 Var[b_3] + 4x Cov[b_2, b_3]$$

Note: Now, the partial effect and the variance are a function of the data! Usually, an average is used in the estimation.

## Application (Greene): Log Income Equation

```
----------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LOGY     Mean                    =       -1.15746
             Standard deviation      =         .49149
             Number of observs.      =          27322
Model size   Parameters              =              7
             Degrees of freedom      =          27315
Residuals    Sum of squares          =     5462.03686
             Standard error of e     =         .44717
Fit          R-squared               =         .17237
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+-------------------------------------------------------------
     AGE|     .06225***       .00213       29.189   .0000      43.5272
   AGESQ|    -.00074***     .242482D-04   -30.576   .0000      2022.99
Constant|   -3.19130***       .04567      -69.884   .0000
 MARRIED|     .32153***       .00703       45.767   .0000       .75869
  HHKIDS|    -.11134***       .00655      -17.002   .0000       .40272
  FEMALE|    -.00491          .00552        -.889   .3739       .47881
    EDUC|     .05542***       .00120       46.050   .0000      11.3202
--------+-------------------------------------------------------------
At Average Age = x = 43.5272.
    Estimated Partial effect = .066225 - 2(.00074) x 43.5272 = .00018.
    Estimated Variance 4.54799e-6 + 4(43.5272)²*(5.87973e-10) + 4(43.5272)*
    (-5.1285e-8) = 7.4755086e-08.
    Estimated standard error = .00027341.
```

Estimated Cov[b1,b2]

| | 1 | 2 |
|---|---|---|
| 1 | 4.54799e-006 | -5.1285e-008 |
| 2 | -5.1285e-008 | 5.87973e-010 |
| 3 | -9.0034e-005 | 9.91407e-007 |

## Specification and Functional Form: Non-linearity

**Example**: Interactive Effect

Population

$$y = \beta_1 + \beta_2 x + \beta_3 z + \beta_4 xz + \varepsilon$$

$$\delta_x = \frac{\partial E[y \mid x, z]}{\partial x} = \beta_2 + \beta_4 z$$

Estimators

$$\hat{y} = b_1 + b_2 x + b_3 z + b_4 xz$$

$$\hat{\delta}_x = b_2 + b_4 z$$

Estimator of the variance of $\hat{\delta}_x$

$$Est.Var[\hat{\delta}_x] = Var[b_2] + z^2 Var[b_4] + 2zCov[b_2, b_4]$$

## Application (Greene): Interaction Effect

```
--------------------------------------------------------------------
Ordinary     least squares regression ...........
LHS=LOGY     Mean                  =      -1.15746
             Standard deviation    =        .49149
             Number of observs.    =         27322
Model size   Parameters            =             4
             Degrees of freedom    =         27318
Residuals    Sum of squares        =    6540.45988
             Standard error of e   =        .48931
Fit          R-squared             =        .00896
             Adjusted R-squared    =        .00885
Model test   F[  3, 27318] (prob) =    82.4(.0000)
--------+-----------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+-----------------------------------------------------------
Constant|   -1.22592***       .01605      -76.376   .0000
     AGE|     .00227***       .00036        6.240   .0000      43.5272
  FEMALE|     .21239***       .02363        8.987   .0000       .47881
 AGE_FEM|    -.00620***       .00052      -11.819   .0000      21.2960
--------+-----------------------------------------------------------
Do women earn more than men (in this sample?)  The +.21239 coefficient on
FEMALE would suggest so.
But, the female "difference" -i.e., partial effect- is: +.21239 - .00620*Age.

At average Age, the effect is:    .21239 - .00620 * (43.5272) = -.05748.
```

## OLS Subject to Restrictions

• Restrictions: Theory imposes certain restrictions on parameters.

**Examples**:

(1) Dropping variables from the equation. That is, certain coefficients in **b** forced to equal 0. (Is variable $\mathbf{x}_3$=*size* significant?")

(2) Adding up conditions: Sums of certain coefficients must equal fixed values. Adding up conditions in demand systems. Constant returns to scale in production functions ($\alpha + \beta = 1$ in a Cobb-Douglas production function).

(3) Equality restrictions: Certain coefficients must equal other coefficients. Using real vs. nominal variables in equations.

• Usual formulation with J linear restrictions (**R** is J$\times k$ and **q** is J$\times$1):

$$\text{Min}_{\mathbf{b}} \{S(x_i, \theta) = \sum_{i=1}^{T} e_i^2 = \mathbf{e'e} = (\mathbf{y} - \mathbf{Xb})' (\mathbf{y} - \mathbf{Xb})\} \quad \text{s.t. } \mathbf{Rb} = \mathbf{q}$$

## Restricted Least Squares

• In practice, restrictions can usually be imposed by solving them out.

(1) Dropping variables –i.e., force a coefficient to equal zero.

Problem: $\min_b \sum_{i=1}^{n}(y_i - b_1x_{i1} - b_2x_{i2} - b_3x_{i3})^2 \quad s.t. \quad b_3 = 0$

$\min_b \sum_{i=1}^{n}(y_i - b_1x_{i1} - b_2x_{i2})^2$

(2) Adding up. Do least squares subject to $b_1+b_2+b_3=1$. Then, $b_3 = 1-b_1-b_2$. Make the substitution so $(\mathbf{y} - \mathbf{x}_3) = b_1(\mathbf{x}_1 - \mathbf{x}_3) + b_2(\mathbf{x}_2 - \mathbf{x}_3) + \mathbf{e}$.

Problem: $\text{Min}_b \quad \sum_{i=1}^{n}((y_i - x_{i3}) - b_1(x_{i1} - x_{i3}) - b_2(x_{i2} - x_{i3}))^2$

(3) Equality. If $b_3 = b_2$, then $\mathbf{y} = b_1\mathbf{x}_1 + b_2\mathbf{x}_2 + b_2\mathbf{x}_3 + \mathbf{e}$
$= b_1\mathbf{x}_1 + b_2(\mathbf{x}_2+\mathbf{x}_3) + \mathbf{e}$

Problem: $\text{Min}_b \quad \sum_{i=1}^{n}(y_i - b_1x_{i1} - b_2(x_{i2} + x_{i3}))^2$

## Restricted Least Squares

• Theoretical results provide insights and the foundation of several tests.

• Programming problem with $J$ restrictions ($\mathbf{R}$ is $J$x$k$ and $\mathbf{q}$ is $k$x1):
Minimize wrt $\mathbf{b}$ $\quad S = (\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta})' (\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta}) \quad$ s.t. $\mathbf{R}\,\boldsymbol{\beta} = \mathbf{q}$

• Quadratic programming problem
⇒ Minimize a quadratic criterion s.t. a set of linear restrictions.
- Concave programming problem, all binding constraints. No need for Kuhn-Tucker
- Solve using a Lagrangean formulation.

• The Lagrangean approach (the 2 is for convenience with is $\boldsymbol{\lambda}$ $J$x1 ).

$$\text{Min}_{\mathbf{b},\lambda} \quad L^* = (\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta})' (\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta}) + 2\, \boldsymbol{\lambda}' (\mathbf{R}\,\boldsymbol{\beta} - \mathbf{q})$$
$$= (\boldsymbol{y}'\boldsymbol{y} - \boldsymbol{\beta}'\mathbf{X}'\boldsymbol{y} - \boldsymbol{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}) + 2\, \boldsymbol{\lambda}'(\mathbf{R}\,\boldsymbol{\beta} - \mathbf{q})$$

## Restricted Least Squares

• The Lagrangean approach

Min $_{\mathbf{b},\lambda}$ $\quad L^* = = (\boldsymbol{y}'\boldsymbol{y} - 2\,\boldsymbol{\beta}'\mathbf{X}'\boldsymbol{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}) + 2\,\boldsymbol{\lambda}'(\mathbf{R}\,\boldsymbol{\beta} - \mathbf{q})$

f.o.c:

$$\partial L^*/\partial \mathbf{b}' = -2\mathbf{X}'(\boldsymbol{y} - \mathbf{X}\mathbf{b}^*) + 2\mathbf{R}'\boldsymbol{\lambda} = \mathbf{0} \;\Rightarrow\; -\mathbf{X}'(\boldsymbol{y} - \mathbf{X}\mathbf{b}^*) + \mathbf{R}'\boldsymbol{\lambda} = \mathbf{0}$$

$$\partial L^*/\partial \boldsymbol{\lambda}' = 2\,(\mathbf{R}\mathbf{b}^* - \mathbf{q}) = \mathbf{0}. \qquad \Rightarrow (\mathbf{R}\mathbf{b}^* - \mathbf{q}) = \mathbf{0}$$

Then, from the 1st equation (and assuming full rank for $\mathbf{X}$):

$$-\mathbf{X}'\boldsymbol{y} + \mathbf{X}'\mathbf{X}\mathbf{b}^* + \mathbf{R}'\boldsymbol{\lambda} = \mathbf{0} \Rightarrow \quad \mathbf{b}^* = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\boldsymbol{\lambda}$$
$$= \mathbf{b} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\boldsymbol{\lambda}$$

Premultiply both sides by $\mathbf{R}$ and then subtract $\mathbf{q}$

$$\mathbf{R}\mathbf{b}^* - \mathbf{q} = \mathbf{R}\mathbf{b} - \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\boldsymbol{\lambda} - \mathbf{q}$$
$$\mathbf{0} = -\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\boldsymbol{\lambda} + (\mathbf{R}\mathbf{b} - \mathbf{q})$$

Solving for $\boldsymbol{\lambda}$ $\quad\Rightarrow\quad \boldsymbol{\lambda} = [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b} - \mathbf{q})$

Substituting in $\mathbf{b}^*$ $\quad\Rightarrow\quad \mathbf{b}^* = \mathbf{b} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\,\mathbf{b} - \mathbf{q})$

## Linear Restrictions

• Q: How do linear restrictions affect the properties of the least squares estimator?

Model ( DGP): $\quad \boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

Theory (information): $\quad \mathbf{R}\boldsymbol{\beta} - \mathbf{q} = \mathbf{0}$

Restricted LS estimator: $\quad \mathbf{b}^* = \mathbf{b} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b} - \mathbf{q})$

1. Unbiased?

$$E[\mathbf{b}^*|\mathbf{X}] = \boldsymbol{\beta} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\,E[(\mathbf{R}\mathbf{b} - \mathbf{q})|\mathbf{X}] = \boldsymbol{\beta}$$

2. Efficiency?

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

$\text{Var}[\mathbf{b}^*|\mathbf{X}] = \text{Var}[\mathbf{b}|\mathbf{X}]$ – a nonnegative definite matrix $< \text{Var}[\mathbf{b}|\mathbf{X}]$

3. $\mathbf{b}^*$ may be more "precise."

Precision = Mean squared error = variance + squared bias.

## Linear Restrictions

1. $\mathbf{b}^* = \mathbf{b} - \mathbf{Cm}$, $\mathbf{m}$ = the "discrepancy vector" $\mathbf{Rb} - \mathbf{q}$.

Note: If $\mathbf{m} = \mathbf{0} \Rightarrow \mathbf{b}^* = \mathbf{b}$. (Q: What does $\mathbf{m} = \mathbf{0}$ mean?)

2. $\boldsymbol{\lambda} = [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{Rb} - \mathbf{q}) = [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\mathbf{m}$

When does $\boldsymbol{\lambda} = \mathbf{0}$? What does this mean?

3. Combining results: $\mathbf{b}^* = \mathbf{b} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\boldsymbol{\lambda}$

4. Recall: $\mathbf{e}'\mathbf{e} = (\mathbf{y} - \mathbf{Xb})'(\mathbf{y} - \mathbf{Xb}) \leq \mathbf{e}^{*\prime}\mathbf{e}^* = (\boldsymbol{y} - \mathbf{Xb}^*)'(\boldsymbol{y} - \mathbf{Xb}^*)$

$\Rightarrow$ Restrictions cannot increase $R^2$ $\Rightarrow R^2 \geq R^{2*}$

## Linear Restrictions – Interpretation

- Two cases
  - Case 1: Theory is correct: $\mathbf{R}\boldsymbol{\beta} - \mathbf{q} = \mathbf{0}$ (restrictions hold).
    $\mathbf{b}^*$ is unbiased & $\text{Var}[\mathbf{b}^*|\mathbf{X}] \leq \text{Var}[\mathbf{b}|\mathbf{X}]$
  - Case 2: Theory is incorrect: $\mathbf{R}\boldsymbol{\beta} - \mathbf{q} \neq \mathbf{0}$ (restrictions do not hold).
    $\mathbf{b}^*$ is biased & $\text{Var}[\mathbf{b}^*|\mathbf{X}] \leq \text{Var}[\mathbf{b}|\mathbf{X}]$.

- Interpretation

- The theory gives us information.

Bad information produces bias (away from "the truth.")

Any information, good or bad, makes us more certain of our answer. In this context, *any* information reduces variance.

## Linear Restrictions - Interpretation

- What about ignoring information (theory)?
  Not using the correct information does not produce bias.
  Not using information foregoes the variance reduction.

## 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)

24

## Hypothesis Testing

- Testing involves the comparison between two competing hypothesis:
  - $H_0$: The maintained hypothesis.
  - $H_1$: The hypothesis considered if $H_0$.

- Idea: We collect a sample, $X = \{X_1, X_2, \dots, X_n\}$. We construct a statistic $T(X) = f(X)$, called the *test statistic*. Now we have a decision rule:
  - If $T(X)$ is contained in space $R$, we reject $H_0$ (& we learn).
  - If $T(X)$ is in the complement of $R$ ($R^C$), we fail to reject $H_0$.

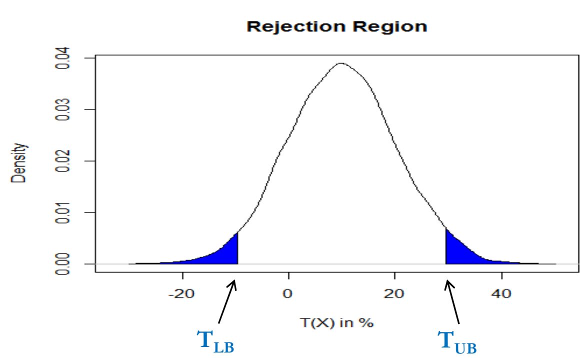
Note: $T(X)$, like any other statistic, is a RV. It has a distribution. We use the distribution of $T(X)$ to determine $R$, the *rejection region* (& we associate a probability to $R$).

25

## Hypothesis Testing: Rejection Region

**Example:** Suppose $T(X) = \bar{X}$. If data is normal, the distribution of $\bar{X}$ is also normal. Then, under $H_0$, we build a Rejection Region, $R$:

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




Note: The blue area ("*significance level*") represents the $P[R|H_0]$. For example, if the blue area is 5%, then, $T_{LB}$ = **-1.96** & $T_{UB}$ = **1.96**.

26

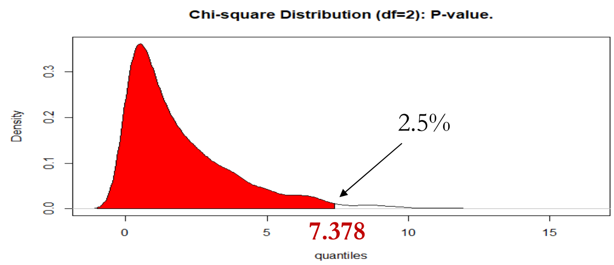
## Hypothesis Testing: *p-value*

• The *classical approach*, also known as *significance testing*, relies on *p-values:*

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

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

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




27

## Hypothesis Testing: Steps

• Steps for the *classical approach*, also known as *significance testing*:

1. Identify $H_0$ & set a *significance level* (α%).

2. Determine the appropriate test statistic T(*X*) and its distribution under the assumption that $H_0$ is true.

3. Calculate T(*X*) from the data.

4. Rule: If *p-value* of T(*X*) < α ⇒ Reject $H_0$ (& we learn $H_0$! is not true).
   If *p-value* of T(*X*) > α ⇒ Fail to reject $H_0$. (No learning.)

Note: In Step 4, setting α% is equivalent to setting *R*. Thus, instead of looking at *p-value*, we can look if T(*X*) falls in *R* (in the blue area). We do this by constructing a (1 - α)% C.I.

• Mistakes are made. We want to quantify these mistakes.

28

## Hypothesis Testing: Error Types

• 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: $\beta = P[X \in R^C | H_1]$

**Example:** From the U.S. Jury System

*Type I error* is the error of finding an innocent defendant guilty.

*Type II error* is the error of finding a guilty defendant not guilty.

• There is a trade-off between both errors.

## Hypothesis Testing: Type I and Type II Errors

• Traditional view: Set *Type I error* equal to a small number & find a test that minimizes *Type II error*.

The usual tests (t-tests, F-tests, Likelihood Ratio tests) incorporate this traditional view.

**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.

30

## Hypothesis Testing: Summary

• Hypothesis testing in Econometrics:

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

(2) We gather data ($\boldsymbol{y}$, $\mathbf{X}$) and estimate the model ⇒ 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 $H_0$.

31

## 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

⇒ Assumed distribution must be incorrect: $H_0$ should be rejected.

32

## 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.

33

## Testing in the CLM: Single Parameter

• We test a hypotheses about a single parameter, say $\beta_k$, of the DGP.

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

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

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

3. Using OLS, we estimate $b_1, b_2, \ldots, b_k, \ldots,$ & estimate $t_k$ ⇒ $\hat{t}$.
4. Decision Rule: Set α level. We reject $H_0$ if p-value($\hat{t}$) < α.
   Or, reject $H_0$, if $|\hat{t}| > t_{T-k,1-\alpha/2}$.

34

## Testing in the CLM: *t-value*

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

Then,

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

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 $|\hat{t}| >$ **1.96** ≈ 2, we say the coefficient $b_k$ is "*significant*."

35

## Hypothesis Testing: Confidence Intervals

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

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

• When the parameter is real-valued, it is common to focus on intervals $C_n = [L_n; U_n]$, called an *interval estimate* for $\theta$. 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 Prob[$\theta \in C_n$].

36

## 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:

$C_n = [b_k + z_{.025} \times \text{Estimated SE}(b_k),\ b_k + z_{.975} \times \text{Estimated SE}(b_k)]$

(Note: The Normal distribution is symmetric $\Rightarrow -z_{.025} = z_{..975} = \mathbf{1.96}$).

• This C.I. is symmetric around $b_k$, with length proportional to its SE.

37

## 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,

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

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

38

## Recall: A *t*-distributed variable

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

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

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

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

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

39

## Hypothesis Testing: Testing Example in R

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

$\mathbf{IBM_{Ret}} - \mathbf{r_f} = \beta_1 + \beta_{Mkt}\,(\mathbf{Mkt_{Ret}} - \mathbf{r_f}) + \beta_{SMB}\,\mathbf{SMB} + \beta_{HML}\,\mathbf{HML} + \varepsilon$

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

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

40

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

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

⇒ all coefficients are *significant* (|t|>2).

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

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

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

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

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

41

## Testing: The Expectation Hypothesis (EH)

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

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

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

Empirical tests of the EH are based on a regression:

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

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

Then, under EH, $H_0$: $\alpha = 0$ and $\beta = 0$.

vs $H_1$: $\alpha \neq 0$ and/or $\beta \neq 0$.

42

## Testing: The Expectation Hypothesis (EH)

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

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

43

## Testing: The Expectation Hypothesis (EH)

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

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

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

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

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

44

## Testing: The Expectation Hypothesis (EH)

- 95% C.I. for b:

  $C_n = [\, b_k \pm t_{k,1-.05/2}$ * Estimated SE($b_k$)]

Then,

$C_n$ = [**-0.215754** – **1.96** * **0.0731553**, **-0.215754** + **1.96** * **0.0731553**]

= [**-0.3591384**, -0.07236961]

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

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

45

## Testing a Hypothesis: Wald Statistic

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

  Wald = normalized distance measure:

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

More than one parameter.

Let **z** = (random vector – hypothesized value) be the distance

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

- Distribution of $W$? We have a quadratic form.
  - If **z** is normal and $\sigma^2$ known, $W \sim \chi^2_{v=Rank(Var[z])}$
  - 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, $W \xrightarrow{d} \chi^2_{v=Rank(Var[z])}$

Abraham Wald (1902–1950, Hungary)

46

## Testing a Hypothesis: Wald Statistic

• Distribution of *W*? We have a quadratic form.

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

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

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

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

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

47

## Recall: An *F*-distributed variable

Let $F = \dfrac{\chi^2_J / J}{\chi^2_T / T} \sim F_{J,T}$

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

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

If Z and U are independent, then

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

48

## Recall: An *F*-distributed variable

• There is a relationship between $t$ and $F$ when testing one restriction.

- For a single restriction, $m = r'b - q$. The variance of $m$ is: r Var[b] r.
- The distance measure is $t = m$ / Est. SE($m$) ~ $t_{T-k}$.
- This $t$-ratio is the sqrt{$F$-ratio}.

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

49

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

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

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

We can write linear restrictions as $H_0$: $\mathbf{R\beta - q = 0}$,
where **R** is a $J$x$k$ matrix and **q** a $J$x1 vector.

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

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

50

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

• 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:

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

Then, we construct a Wald statistic:

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

to test if **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 ($RSS_R$).

51

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

• 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:

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

Then, we construct a Wald statistic:

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

to test if **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 ($RSS_R$).

52

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

• To test $H_0$, we calculate the discrepancy vector:

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

Then, we compute the Wald statistic:

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

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

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

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

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

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

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

53

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

**Example:** In the 3 FF factor model for IBM excess returns ($T$=569)

$\mathbf{IBM_{Ret}} - \mathbf{r_f} = \beta_1 + \beta_{Mkt}\,(\mathbf{Mkt_{Ret}} - \mathbf{r_f}) + \beta_{SMB}\,\mathbf{SMB} + \beta_{HML}\,\mathbf{HML} + \varepsilon$

we want to test if $\beta_{SMB} = 0.2$ and $\beta_{HML} = 0.6$.

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

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

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

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

54

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

**Example (continuation):**

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

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

3. Get OLS results, compute F.

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

   Or, reject $H_0$, if $F > F_{J=2,T-2,.05}$.

```
J <- 2                                 # number of restriction
R <- matrix(c(0,0,0,0,1,0,0,1), nrow=2)  # matrix of restrictions
q <- c(.2,1)                           # hypothesized values
m <- R%*%b - q                         # m = Estimated R*Beta - q
```

55

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

**Example (continuation):**

```
Var_m <- R %*% Var_b %*% t(R)          # Variance of m
det(Var_m)                             # check for non-singularity
W <- t(m)%*%solve(Var_m)%*%m
F_t <- as.numeric(W/J)                 # F-test statistic

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

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

56

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

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

Using the previous program but with:

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

57

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

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

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

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

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

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

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

Recall

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

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

Then,

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

58

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

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

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

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

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

or

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

59

## 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 $X_2, ..., X_k$ are equal to zero. That is,

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

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

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

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

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

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

60

10

## The F Test: F-test of goodness of fit

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

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

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

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

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

61

10

## The F Test: F-test of goodness of fit

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

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

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

62

10

## The F Test: General Case – Example

- In the linear model

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

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

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

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

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

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

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

63

32

## The F Test: General Case – Example

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

Unrestricted Model:

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

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

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

Then, the Restricted Model:

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

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

64

## The F Test: General Case – Example

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

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

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

65

## The F Test: General Case – Example

**Example (continuation):**

```
F_test <- ((RSS2 - RSS)/J)/(RSS/(T-k))
> F_test
[1] 4.736834
> qf(.95, df1=J, df2=(T-k))                      # F_{2,565,.05} value (≈ 3)
[1] 3.011672                                  ⇒ Reject H0.
p_val <- 1 - pf(F_test, df1=J, df2=(T-k))        # p-value of F_test
> p_val
[1] 0.009117494                               ⇒ p-value is small ⇒ Reject H0.
```

66

## Lagrange Multiplier Statistics

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

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

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

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

where

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

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

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

Then,

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

67

## Application (Greene): Gasoline Demand

• Time series regression,

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

Period = 1960 - 1995.

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

Note: All coefficients in the model are elasticities.

68

## Application (Greene): Gasoline Demand

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

69

## Application (Greene): Gasoline Demand

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

(1) Distance measure: $t_6 = (b_6 - 0) / s_{b6} = (.11571 - 0) / .07859$
$= 1.472 < 2.052$ ⇒ cannot reject $H_0$.

(2) Confidence interval: $b_6 \pm t_{(.95,27)} \times$ Standard error
$= .11571 \pm 2.052 \times (.07859)$
$= .11571 \pm .16127 = (-.045557 ,.27698)$
⇒ C.I. contains 0 ⇒ cannot reject $H_0$.

(3) Regression fit if $\mathbf{X}_6$ drop? Original $R^2 = .99605$,
Without LPPT, $R^{*2} = .99573$

$F(1,27) = [(.99605 - .99573)/1]/[(1 - .99605)/(36 - 9)] = 2.187$
$= 1.472^2$ (with some rounding) ⇒ cannot reject $H_0$.

70

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

• Do the three aggregate price elasticities sum to zero?

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

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

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

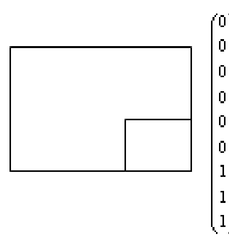
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.6851 | -0.189024 | -0.0256198 | -0.218091 | -0.0240267 | -0.0295907 | -0.0261772 | 0.197857 | 0.176068 |
| 2 | -0.189024 | 0.0212045 | 0.00290895 | 0.0243971 | 0.00269963 | 0.0032894 | 0.00280174 | -0.0222154 |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

71

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

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

Var[m] = R × Var[b] × R′ = [0 0 0 0 0 0 1 1 1]



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

$m' \, [Var(m)]^{-1} \, m = 8.5446$

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

72

## Gasoline Demand (Greene) - Imposing Restrictions

```
Linearly restricted regression
LHS=LG       Mean                 =   5.392989
             Standard deviation   =   .2487794
             Number of observs.   =         36
Model size   Parameters           =          8  <*** 9 – 1 restriction
             Degrees of freedom   =         28
Residuals    Sum of squares       =   .0112599  <*** With the restriction
Residuals    Sum of squares       =   .0085531  <*** Without the
restriction
Fit          R-squared            =   .9948020
Restrictns.  F[  1,    27] (prob) =   8.5(.01)
Not using OLS or no constant.R2 & F may be < 0
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  t-ratio  P[|T|>t]  Mean of X
--------+-------------------------------------------------------------
Constant|   -10.1507***       .78756      -12.889   .0000
      LY|    1.71582***       .08839       19.412   .0000      9.11093
     LPG|    -.45826***       .06741       -6.798   .0000       .67409
    LPNC|     .46945***       .12439        3.774   .0008       .44320
    LPUC|    -.01566          .06122        -.256   .8000       .66361
    LPPT|     .24223***       .07391        3.277   .0029       .77208
     LPN|    1.39620***       .28022        4.983   .0000       .60539
     LPD|     .23885          .15395        1.551   .1324       .43343
     LPS|   -1.63505***       .27700       -5.903   .0000       .68105
--------+-------------------------------------------------------------
F = [(.0112599 - .0085531)/1] / [.0085531/(36 – 9)]  =  8.544691
```

73

## Gasoline Demand (Greene) - Joint Hypotheses

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

Strategy: Regress $\log G - \log Y + \log Pg$ on the other variables and

• Compare the sums of squares

With two restrictions imposed

Residuals Sum of squares = .0286877

Fit R-squared = .9979006

Unrestricted

Residuals Sum of squares = .0085531

Fit R-squared = .9960515

$F = ((.0286877 - .0085531)/2) / (.0085531/(36-9)) = 31.779951$

The critical F for 95% with 2,27 degrees of freedom is 3.354 $\Rightarrow H_0$ is rejected.

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

74

## Gasoline Demand - Using the Wald Statistic

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

75

## Gasoline Demand (Greene) – Testing Details

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

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

• Recall that under $H_0$, λ should be 0.

```
Matrix Result   has  2 rows and  1 columns.
                 1
        +-------------+
       1|    -.88491      Income elasticity
       2|  129.24760      Price elasticity
        +-------------+
```

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

76

## 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,

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

• Q: What's wrong?

77