INTRODUCTORY ECONOMETRICS

Dr Javier Fernández

etpfemaj@ehu.es

Dpt. of Econometrics & Statistics

UPV—EHU

1 Introduction

1.1 Definitions. Elements of Econometrics

Introduction: Definitions

ECONOMETRICS

- (plz, do not confuse with economic + tricks !!!)
- etymological:
  $o\acute{\iota}\kappa o\varsigma$ [oíkos], 'household',
  and $\nu\acute{o}\mu o\varsigma$ [nómos], 'rules'
  hence economics $\rightsquigarrow$ household management,
  $+\ \mu\varepsilon\tau\rho\acute{\omega}$ [metró], 'measure'.
  Economy + Measurement
- additive:
  Social science which applies
  Economic theory, Mathematics and Statistical inference
  to the analysis of economic phenomena (Goldberger(1964)).
- utilitarian: The art of the econometrician = define appropriate model + find optimal statistical procedure
  $\rightsquigarrow$ econometrician $\neq$ statistician;
  $\cdots +$ sound training in economics (Malinvaud(1963)).

Introduction: Definitions

- plain: application of statistical methods to economic data (Maddala(1977)).
- concise: empirical determination of economic laws (Theil(1971)).
- AFG(2004): Econometrics deals with
  - formulation (or specification),
  - quantification (or estimation),
  - validation (or testing),

  of relationships among economic variables.

Introduction: 3 Elements:

- **ECONOMIC THEORY:**
  in charge of
  - (general:) analysis of the economy
  - (specific:) relationships among economic variables
- **DATA:**
  to quantify is NOT one of the objectives of Economic Theory
- **STATISTICS:**
  provides basic structure of data processing methods for:
  - (estimation:)
    quantify relationships among variables in an appropriate way.
  - (testing:)
    validate results in agreement with certain established standards.

1.2 Concept and example of model: From the economic model to the econometric model.

Element 1: Economic Th: basic model

- **Case:** company manager or sales director,
- **Interest:** to know relationship between their sales and their price.

- **basic economic logic:** sales as a function of price $\rightsquigarrow$ basic economic model:

$$\underset{sales}{V} = f(\underset{\underset{(-)}{price}}{p})$$

$f(\bullet)$ is a generic function
(Ec Th : $f(\bullet) =$ inverse fn $\quad\rightsquigarrow\quad$ sales $\uparrow$ if price $\downarrow$.)

Element 1: Economic Th: additional vars

- **additional economic logic:**
  sales depend on
  - conditions of rival firms (e.g. competition price)
  - market conditions (e.g. economic cycle)
- **complete Model:**

$$\underset{\text{sales}}{V} = f(\underset{\substack{\text{price}\\(-)}}{p}, \underset{\substack{\text{competition price}\\(+)}}{pc}, \underset{\substack{\text{cycle}\\(+)}}{c})$$

- NOTE:
  proposed economic model $\equiv$ summary of ideas,
  but nothing new for manager;
  they need specific model for their company
  $\rightsquigarrow$ how their sales **respond** to **their** price.

Element 2: Data:

- specific Information:
  manager has **information** about:
-
  - their sales and their prices **(quantitative data)**
  - prices of the competition **(quantitative data)**
  - cyclical moment **(qualitative data)**
- *e.g.*:

dates	Sales	price	comp.p.	cycle
jan 80	1725	12.37	11.23	high
feb 80	1314	11.25	10.75	high
apr 95	1234	13.57	14.5	low
⋮	⋮	⋮	⋮	⋮

and all this month after month until December of 2004.

Element 2: Data: specific model

- specific model for available data:

$$V_t = f(p_t, pc_t, c_t), \qquad t = 1980.1, \ldots, 2004.12$$

where subindex t indicates period or moment of relationship.

- up to now:
  - economic model: summary of general ideas about relationship
  - data: or specific information on the different variables

  - **How to put together both elements?...????**

E2: (generic) model + (specific) data?:

- **A:** assumptions about $f(\bullet)$; *e.g.*: linear relationship.
  The model will then be:

$$V_t = \beta_0 + \beta_1 p_t + \beta_2 pc_t + \beta_3 c_t, \qquad t = 1980.1, \ldots, 2004.12$$

- β's $=$ parameters or coefficients :
  e.g. β_1 answers the question:
  how much sales change if price changes in one monetary unit?
  $\rightsquigarrow$ price policies, production decisions etc. for the company.

- **B:** indicators:
  allocate quantitative values to qualitative variables (like Cycle): *e.g.* substitute with indicator such as Industrial Production Index.

E2: Model +data?: random disturbances

- After this the model expresses a quantitative relationship among variables:

$$
\begin{aligned}
1725 &= \beta_0 + 12.37\beta_1 + 11.23\beta_2 + 101.7\beta_3 \qquad \text{(1980.Jan)} \\
1314 &= \beta_0 + 11.25\beta_1 + 10.75\beta_2 + 97.3\beta_3 \qquad \text{(1980.Feb)} \\
\vdots &= \vdots
\end{aligned}
$$

- **NOTE:** . . . different relationship for each month??? . . .
- **C:** disturbance term;

- back to the generic *economic* model:
  - ⇒ stable behaviour among variables
  - ⇒ "average" behaviour reflected in data
  - ⇒ add term u_t to cover up for small discrepancies. . .

E2: Model+data?: interpretation

- The econometric model will finally be:

$$V_t = \underbrace{\beta_0 + \beta_1 p_t + \beta_2 pc_t + \beta_3 c_t}_{\text{(important \& systematic “influences”)}} + \underbrace{u_t}_{\text{(random disturbance term)}}$$

- Interpretation of u_t:
  - $\Rightarrow$ effects that affect sales slightly in every period but not explicitly picked up by the model.
  - $\Rightarrow$ small data discrepancies.
  - $\Rightarrow$ non systematic effects $\equiv$ more erratic.
  - $\Rightarrow$ **random variable** with certain probability law (*e.g.*: Normal dn).

Element 3: Statistics:

- Model contains a random variable

  ⇝ **statistical** procedures that guarantee good results:

⇒ to estimate numeric value of the coefficients,

⇒ to test the validity of the relationship,

- the estimated model
  - won't be a generic model
  - but a specific model for the company
- it will offer the manager

  specific information to make decisions.

1.3 The Econometric Model. The Disturbance or Error term.

Basic Characteristics: data notation

More general econometric model with K variables:

- for time series data:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \cdots + \beta_K X_{Kt} + u_t, \quad t = 1, 2, \ldots, T.$$

- or, for cross-section data:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_K X_{Ki} + u_i, \quad i = 1, 2, \ldots, N.$$

- or, for panel data:

$$Y_{it} = \beta_0 + \beta_1 X_{1it} + \cdots + \beta_K X_{Kit} + u_{it}, \quad \begin{cases} i = 1, 2, \ldots, N; \\ t = 1, 2, \ldots, T. \end{cases}$$

Basic Characteristics: vars notation

- Y: the variable we want *to explain*:
  dependent v, explained v, endogenous v or regressand.
- $X_1, X_2 \ldots X_K$: variables that *explain* the variable Y:
  explanatory v, independent v, exogenous v or regressors.
- $\beta_k, (k = 1 \ldots K)$: *unknown constants* that determine relationship among variables:
  parameters or intercept & coefficients.
  $\widehat{\beta}_k$ is the *estimated* coefficient.
- u: variable that picks up *other non-important effects* present in data: **random disturbance or error term.**

Basic Differences with economic model

Presence of a random disturbance that

- picks up erratic behaviour:

$$Y_t = \underbrace{\beta_0 + \beta_1 X_{1t} + \cdots + \beta_K X_{Kt}}_{\text{systematic part}} + \underbrace{u_t}_{\text{non-systematic or random part}} \qquad t = 1, 2 \ldots T.$$

- has zero mean:

$$E(Y_t) = E(\beta_0 + \beta_1 X_{1t} + \cdots + \beta_K X_{Kt}) + \underbrace{E(u_t)}_{=0} \quad t = 1, 2 \ldots T.$$

- hence systematic part $\equiv$ average behaviour of Y.
- other assumptions on u (basic hypothesis, etc.)
  $\rightsquigarrow$ probabilistic behaviour in different cases
  $\rightsquigarrow$ statistical tools $\rightsquigarrow$ Econometric Methods.

Classification of econometric models

Different approaches:

- looking at type of data:
  - Time series model.
  - Cross-section model.
- looking at period of observation:
  - **static M.:** Vars measured in same moment.
  - **dynamic M.:** Vars referred to different periods:
    e.g. $Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{1,t-1} + \beta_3 X_{2,t-1} + u_t$
- looking at number of relationships:
  - **Single-equation models:**
    a single relationship or equation.
  - **Simultaneous or Multiple-equation models:**
    more than one equation.

etc.

1.4 Stages in the elaboration of the model. Uses of the model.

Stages in the elaboration of the model

0. **Selection.** Outline the theory of interest:
   - select the variable to explain: Y.
   - select the overall relationship: $Y = f(X)$.
1. **Specification.** Outline econometric model coherent with theory:
   - choose the explanatory variables: $X_1 \dots X_K$.
   - choose the functional form: *e.g.* $f(\cdot) \equiv$ lineal.
   - choose the probabilistic behaviour (distribution) of the random disturbance: u, *e.g.* $u_t \sim \text{iid}\,\mathcal{N}(0, \sigma^2)$.

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_K X_K + u.$$

Stages in the elaboration of the model

2. **Estimation.** Quantify unknown parameters according to the available information:
   - find data for variables: $Y_t, X_{1t}, \ldots, \ldots, X_{Kt}$ for $t = 1, \ldots, T$.
   - choose the appropriate statistical method, *e.g.* OLS:

$$\widehat{Y}_t = \widehat{\beta}_0 + \widehat{\beta}_1 X_{1t} + \cdots + \widehat{\beta}_K X_{Kt} + \widehat{u}_t, \quad t = 1, 2 \ldots T.$$

3. **Validation.** Evaluate whether the model represents the initial problem correctly:
   - statistical inference on hypotheses.
   - model not adequate $\rightsquigarrow$ back to specification phase.

Using the econometric model

The model that has gone thru all the previous stages can then be used for:

- **economic analysis:**
  - interpretation of coefficients,
  - hypothesis testing,
  - etc.
- **prediction:**
  - **time series forecasting:**
    to forecast (predict) future values of Y.
  - **in general:**
    to respond to questions of the type,
    what would happen if...?

2 The Linear Regression Model (I). Specification and Estimation.

2.1 Specification of the General Linear Regression Model (GLRM).

Specification of the GLRM (1)

- **Objective:** Quantifying the relationship between:
  - a variable Y and
  - a set of K explanatory variables $X_1, X_2, \ldots, X_K$,
  - by means of a linear model.
- **Starting point:**
  - a linear model: $Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_K X_K + u$,
  - a data sample of size T: $Y_t, X_{1t}, X_{2t}, \ldots, X_{Kt},\ t = 1 \ldots T$, where

$$Y_t = t\text{-th obs of } Y,$$

$$X_{kt} = t\text{-th obs of } X_k,\ k = 1, 2 \ldots K.$$

Specification of the GLRM (2)

- GLRM:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \cdots + \beta_K X_{Kt} + u_t, \; t = 1, 2 \ldots T,$$

whose elements are (recall):

- Y: dependent variable,
- $X_k, \; k = 1 \ldots K$: explanatory variables,
- β_0: intercept,
  $\beta_k, \; k = 1 \ldots K$: coefficients (parameters to be estimated),
- u: (non-observable random) error or disturbance,
  that allows for:
  - variables not included in the model,
  - random behaviour of economic agents,
  - measurement errors, etc.

The GLRM in observation form

The model

$$Y_t = \beta_0 + \beta_1 X_{1t} + \cdots + \beta_K X_{Kt} + u_t, \; t = 1, 2 \ldots T,$$

implies for each observation:

$$Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{21} + \cdots + \beta_K X_{K1} + u_1$$
$$Y_2 = \beta_0 + \beta_1 X_{12} + \beta_2 X_{22} + \cdots + \beta_K X_{K2} + u_2$$
$$\ldots\ldots$$
$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_K X_{Kt} + u_t$$
$$\ldots\ldots$$
$$Y_T = \beta_0 + \beta_1 X_{1T} + \beta_2 X_{2T} + \cdots + \beta_K X_{KT} + u_T$$

The GLRM in matrix form (1)

or else in matrix form:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_t \\ \dots \\ Y_T \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 X_{11} + \beta_2 X_{21} + \cdots + \beta_K X_{K1} \\ \beta_0 + \beta_1 X_{12} + \beta_2 X_{22} + \cdots + \beta_K X_{K2} \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_K X_{Kt} \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \beta_0 + \beta_1 X_{1T} + \beta_2 X_{2T} + \cdots + \beta_K X_{KT} \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_t \\ \dots \\ u_T \end{bmatrix}$$

The GLRM in matrix form (2)

- that is:

$$\underset{(T\times 1)}{\underset{Y}{\begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_t \\ \dots \\ Y_T \end{bmatrix}}} = \underset{(T\times K{+}1)}{\underset{X}{\begin{bmatrix} 1 & X_{11} & X_{21} & \dots & X_{K1} \\ 1 & X_{12} & X_{22} & \dots & X_{K2} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & X_{1t} & X_{2t} & \dots & X_{Kt} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & X_{1T} & X_{2T} & \dots & X_{KT} \end{bmatrix}}} \underset{(K{+}1\times 1)}{\underset{\beta}{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{pmatrix}}} + \underset{(T\times 1)}{\underset{u}{\begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_t \\ \dots \\ u_T \end{bmatrix}}}$$

-

$$Y = X\beta + u.$$

↰

2.2 Basic (Classical) Assumptions. Interpretation.

Basic Assumptions of the GLRM (1)

1. Assumptions about the relationship:
   - Model is correctly specified:

$$X_k \text{ explains } Y \Leftrightarrow X_k \in \text{model}.$$

2. Assumptions about the parameters:
   - they are constant throughout the sample,
   - they appear linearly (*i.e.* a constant plus coefficients)
     - $Y_t = \beta_0 + \beta_1 X_t + u_t$
   - Note: but vars $Y, X_1, X_2, \ldots$ may be transformations:
     - $Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_t^2 + \beta_3 \frac{1}{X_t} + u_t$
     - $Y_t = A X_{1t}^{\beta_1} X_{2t}^{\beta_2} e^{u_t}$ (Why?)
     - and this? $Y_t = \beta_0 + \beta_1 \frac{1}{X_t - \beta_2} + u_t$
     - and these other? $\ln Y_t = \beta_0 X_t^{\beta_1} u_t$; $Y_t = \beta_0 X_t^{\beta_1} + u_t$
       $Y_t = \beta_1 X_{1t} + \beta_2 X_{1t} X_{2t} + u_t$; $Y_t = \beta_0 + \beta_1 X_{1t}^{X_{2t}} + u_t$

Basic Assumptions of the GLRM (2)

3. Assumptions about the explanatory variables:
   (a) $X_1, \ldots, X_K$, are quantitative and fixed (*i.e.* not random).
   (b) $X_1, \ldots, X_K$, are linearly independent:
   - $\nexists X_k | X_k =$ lin. comb. of others (Why?)

- Examples of not valid cases:
  - $Y_t = \beta_0 + \beta_1 X_t + \beta_2 \left(2X_t + 3\right) + u_t$
  - $Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 \left(X_{1t} + X_{2t}\right) + u_t$
- Examples of valid cases:
  - $Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_t^2 + u_t$
  - $Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{1t} X_{2t} + u_t$

Basic Assumptions of the GLRM (3)

4. Assumptions about the disturbance term:
   (a) Zero mean:
   $\mathsf{E}(u_t) = 0 \quad \forall t$ (isn't obvious?).
   (b) Homoscedastic:
   $\mathsf{Var}(u_t) = \mathsf{E}(u_t^2) = \sigma_u^2 \; (= \sigma^2)$ const ($\forall t$).
   (c) Serially uncorrelated:
   $\mathsf{Cov}(u_t, u_s) = \mathsf{E}(u_t u_s) = 0 \quad \forall t \neq s$.
   (d) Normally distributed$^{(*)}$:
   $u_t \sim \mathcal{N} \quad \forall t$. (* added)

- Assumptions 4a–4d jointly:

$$u_t \sim \text{iid}\,\mathcal{N}(0, \sigma_u^2)$$

Basic Assumptions in matrix form (1)

- from 4a: Mean Vector:

$$\underset{(T\times 1)}{\mathsf{E}(u)} = \begin{bmatrix} \mathsf{E}(u_1) \\ \mathsf{E}(u_2) \\ \vdots \\ \mathsf{E}(u_T) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = 0_T$$

- from 4b and 4c: Covariance Matrix:

$$\underset{(T\times T)}{\mathsf{E}(uu')} = \begin{bmatrix} \mathsf{E}(u_1^2) & \mathsf{E}(u_1u_2) & \dots & \mathsf{E}(u_1u_T) \\ \mathsf{E}(u_2u_1) & \mathsf{E}(u_2^2) & \dots & \mathsf{E}(u_2u_T) \\ \dots & \dots & \dots & \dots \\ \mathsf{E}(u_Tu_1) & \mathsf{E}(u_Tu_2) & \dots & \mathsf{E}(u_T^2) \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_u^2 & 0 & \dots & 0 \\ 0 & \sigma_u^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_u^2 \end{bmatrix} = \sigma_u^2 I_T$$

Basic Assumptions in matrix form (2)

- more compactly:

$$\underset{(T\times 1)}{u} \sim \left(\underset{(T\times 1)}{0}, \underset{(T\times T)}{\sigma_u^2 I_T} \right)$$

- plus 4d:

$$\underset{(T\times 1)}{u} \sim \mathcal{N}\left(\underset{(T\times 1)}{0}, \underset{(T\times T)}{\sigma_u^2 I_T} \right)$$

2.3a Ordinary Least Squares (OLS) in a Single Linear Regression Model (SLRM).

SLRM: the PRF

- With $K = 1 \quad \rightsquigarrow \quad Y_t = \beta_0 + \beta_1 X_{1t} + u_t\,,$

$$\text{(SLRM):} \quad Y_t = \alpha + \beta X_t + u_t\,. \qquad (1)$$

- Population Regression Function (PRF):
  $\mathsf{E}\big(u_t\big) = 0 \quad \rightsquigarrow \quad$ *systematic part* or PRF:

$$\mathsf{E}\big(Y_t\big) = \alpha + \beta X_t$$

- Interpretation of the parameters:
  - $\alpha = \mathsf{E}\big(Y_t | X_t = 0\big)$: Expected value of Y_t when the explanatory variable is zero.
  - $\beta = \dfrac{\partial \mathsf{E}\big(Y_t\big)}{\partial X_t} \simeq \dfrac{\Delta \mathsf{E}\big(Y_t\big)}{\Delta X_t}$: Increase in (expected) value of Y_t when $X \uparrow$ one unit (*c.p.*).

- Objective: To obtain estimates $\widehat{\alpha}$, $\widehat{\beta}$ of the unknown parameters α, β in (1).

The Sample Regression Function (SRF)

- $\widehat{\alpha}, \widehat{\beta} \quad \rightsquigarrow$ estimated model or SRF:

$$\widehat{Y}_t = \widehat{\alpha} + \widehat{\beta} X_t$$

- Interpretation of the estimates:
  - $\widehat{\alpha} = (\widehat{Y}_t | X_t = 0)$: Estimated value of Y_t when the explanatory variable is zero.
  - $\widehat{\beta} = \frac{\partial \widehat{Y}_t}{\partial X_t} \simeq \frac{\Delta \widehat{Y}_t}{\Delta X_t}$: Estimated increase in Y_t when $X \uparrow$ one unit (*c.p.*).
- Note difference: an estimator (a formula) vs. an estimate (a number).

Disturbances vs. Residuals

- Disturbances in PRF:

$$u_t = Y_t - \mathsf{E}\left(Y_t\right) = Y_t - \alpha - \beta X_t$$

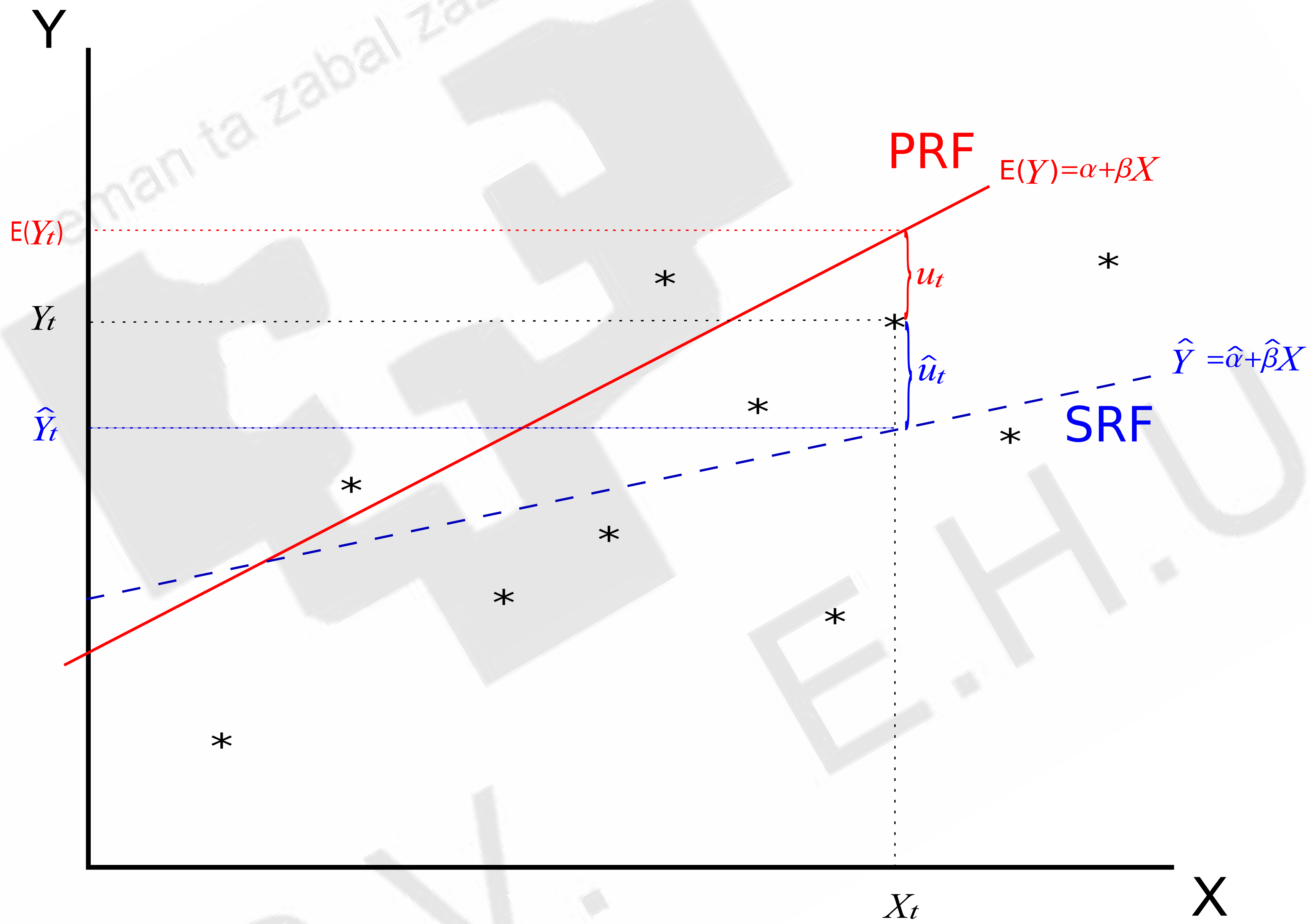
- Residuals in SRF:

$$\widehat{u}_t = Y_t - \widehat{Y}_t = Y_t - \widehat{\alpha} - \widehat{\beta} X_t$$

- Residuals are to the SRF

  what disturbances are to the PRF.

SLRM: PRF and SRF

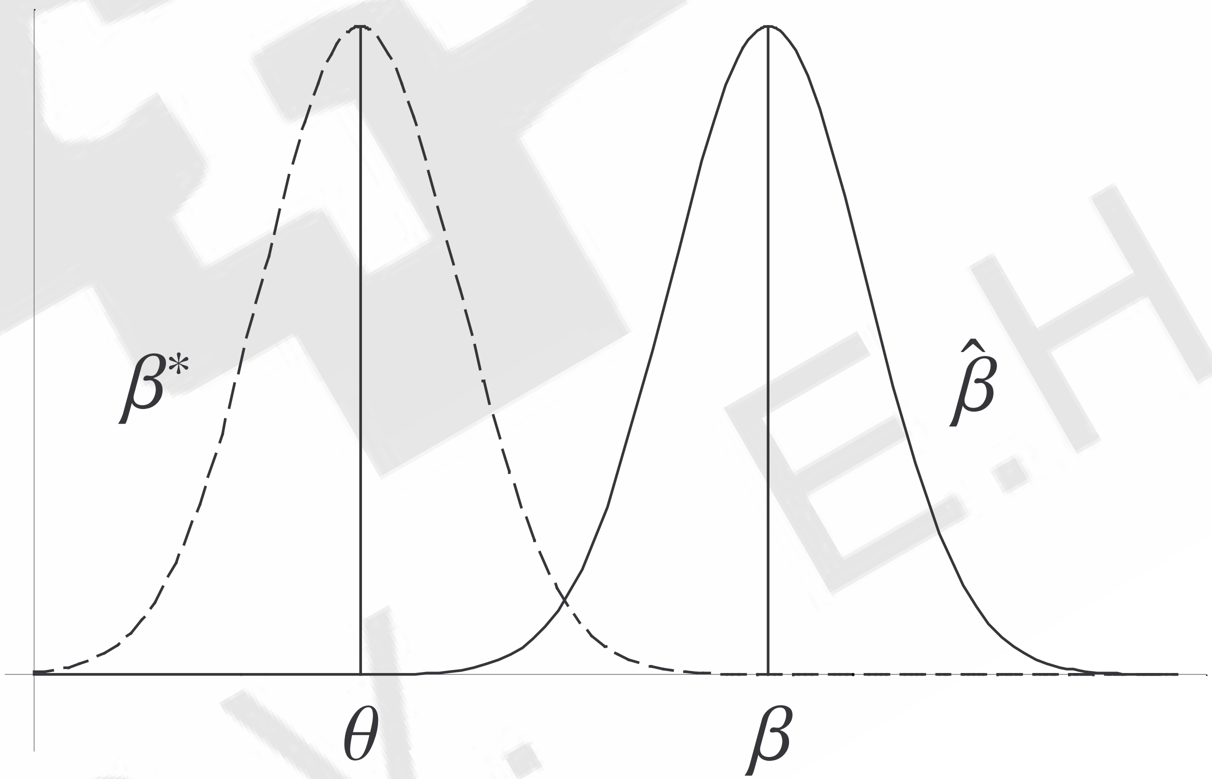


Estimation: Desired Properties (1)

Let $\widehat{\beta}$ be an estimator of $\beta \ldots$

Unbiasedness:

$$\mathrm{E}\left(\widehat{\beta}\right) = \beta \quad \Leftrightarrow \quad \widehat{\beta} \text{ unbiased}$$

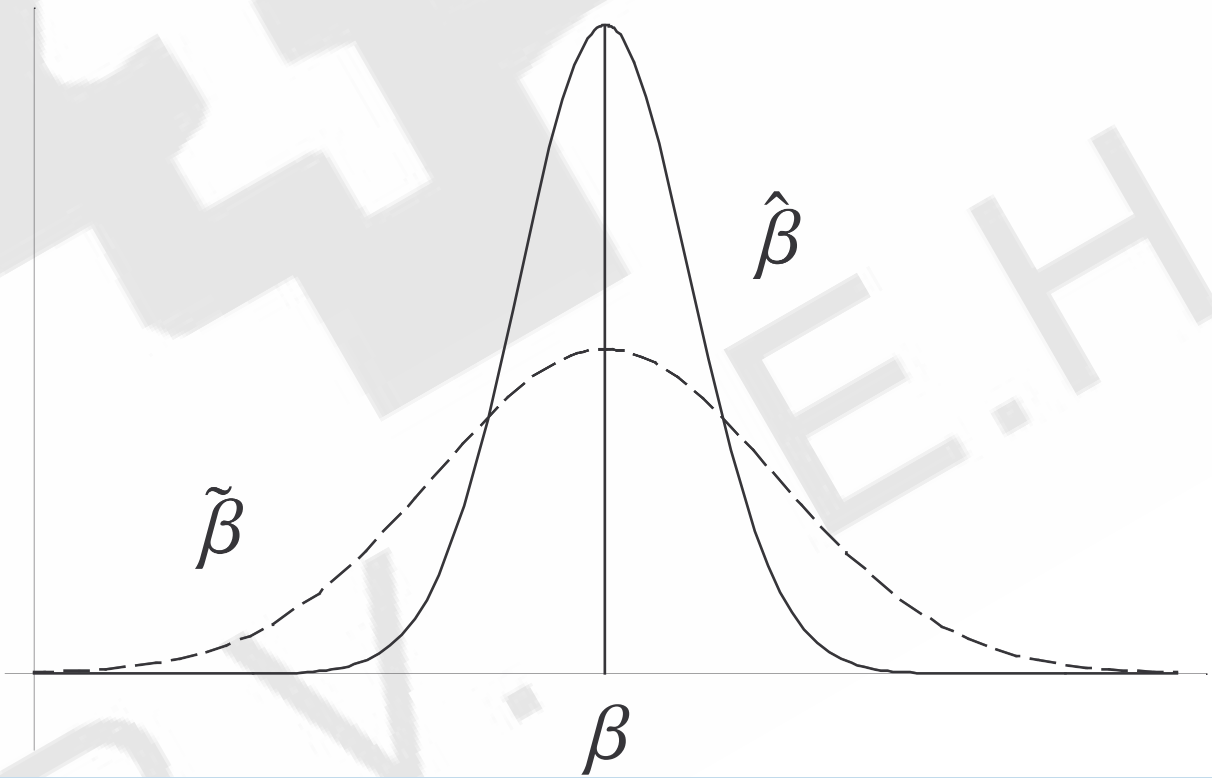


Estimation: Desired Properties (2)

Let $\widehat{\beta}$ and $\widetilde{\beta}$ be unbiased estimators of β...

Relative efficiency:

$$\text{Var}\left(\widehat{\beta}\right) \leq \text{Var}\left(\widetilde{\beta}\right) \quad \Leftrightarrow \quad \widehat{\beta} \text{ relatively efficient}$$



Estimation: OLS criteria

SLRM: $Y_t = \alpha + \beta X_t + u_t$,

- apply Least-Squares fit:

$$\min_{\alpha,\beta} \sum_{t=1}^{T} u_t^2 \quad \text{where} \quad u_t = Y_t - \alpha - \beta X_t :$$

- First derivatives:
  - $$\frac{\partial \sum u_t^2}{\partial \alpha} = 2 \sum u_t \frac{\partial u_t}{\partial \alpha} = 2 \sum u_t (-1)$$
  - $$\frac{\partial \sum u_t^2}{\partial \beta} = 2 \sum u_t \frac{\partial u_t}{\partial \beta} = 2 \sum u_t (-X_t)$$
- 1st.o.c. (minimum) $\Rightarrow$ first derivatives are zero:
  - $$\sum \widehat{u}_t = \sum (Y_t - \widehat{\alpha} - \widehat{\beta} X_t) = 0$$
  - $$\sum \widehat{u}_t X_t = \sum (Y_t X_t - \widehat{\alpha} X_t - \widehat{\beta} X_t^2) = 0$$

Estimation: Normal equations & LSE of α

- From the above 1st.o.c's:

$$\sum (Y_t - \widehat{\alpha} - \widehat{\beta} X_t) = 0$$

$$\sum (Y_t X_t - \widehat{\alpha} X_t - \widehat{\beta} X_t^2) = 0$$

- we obtain the Normal Equations:

$$\left.\begin{array}{r} \sum Y_t = T\widehat{\alpha} + \widehat{\beta} \sum X_t \\ \sum Y_t X_t = \widehat{\alpha} \sum X_t + \widehat{\beta} \sum X_t^2 \end{array}\right\}$$

2 equation system with 2 unknowns!!

- Dividing the 1st. normal eq. by T:

$$\frac{1}{T} \sum Y_t = \frac{1}{T} T \widehat{\alpha} + \widehat{\beta} \frac{1}{T} \sum X_t$$

- That is:

$$\boxed{\widehat{\alpha}_{OLS} = \overline{Y} - \widehat{\beta}\, \overline{X}}$$

Estimation: Normal equations & LSE of β

- Substituting $\widehat{\alpha}$ in the 2nd. normal eq.:

$$\sum Y_t X_t = (\overline{Y} - \widehat{\beta}\overline{X}) \sum X_t + \widehat{\beta} \sum X_t^2$$

- ...dividing by T and gathering terms together:

$$\frac{1}{T} \sum Y_t X_t = (\overline{Y} - \widehat{\beta}\overline{X}) \tfrac{1}{T} \sum X_t + \widehat{\beta} \frac{1}{T} \sum X_t^2$$

$$\frac{1}{T} \sum Y_t X_t - \overline{Y}\overline{X} = \widehat{\beta} \left(\frac{1}{T} \sum X_t^2 - \overline{X}^2 \right)$$

- ...and solving for the unknown:

$$\widehat{\beta} = \frac{\frac{1}{T} \sum Y_t X_t - \overline{Y}\,\overline{X}}{\frac{1}{T} \sum X_t^2 - \overline{X}^2} = \frac{\frac{1}{T} \sum y_t x_t}{\frac{1}{T} \sum x_t^2} \quad \left[\frac{\text{Why?}}{\text{Why?}} \right] \longrightarrow$$

- That is:

$$\boxed{\widehat{\beta}_{OLS} = \frac{\sum y_t x_t}{\sum x_t^2} = \frac{\text{Cov}(Y, X)}{\text{Var}(X)}}$$

Recall: variances and covariances?

- variance from original (uncentred) data?

$$\text{Var}(X) = \frac{1}{T}\sum x_t^2 = \frac{1}{T}\sum (X_t - \overline{X})^2$$
$$= \frac{1}{T}\sum X_t^2 + \frac{1}{T}\sum \overline{X}^2 - \frac{2}{T}\overline{X}\sum X_t$$

$$\boxed{\frac{1}{T}\sum x_t^2 = \frac{1}{T}\sum X_t^2 - \overline{X}^2}$$

- covariance from original (uncentred) data?

$$\text{Cov}(Y, X) = \frac{1}{T}\sum x_t y_t = \frac{1}{T}\sum (X_t - \overline{X})(Y_t - \overline{Y})$$
$$= \frac{1}{T}\sum X_t Y_t + \frac{1}{T}\sum \overline{X}\overline{Y} - \frac{1}{T}\overline{Y}\sum X_t - \frac{1}{T}\overline{X}\sum Y_t$$

$$\boxed{\frac{1}{T}\sum x_t y_t = \frac{1}{T}\sum X_t Y_t - \overline{X}\overline{Y}}$$

Numerical example: strawberry prod data

- Data...
- Centred data or "in deviation form"
  (deviations from respective means)...
- Squares and products...

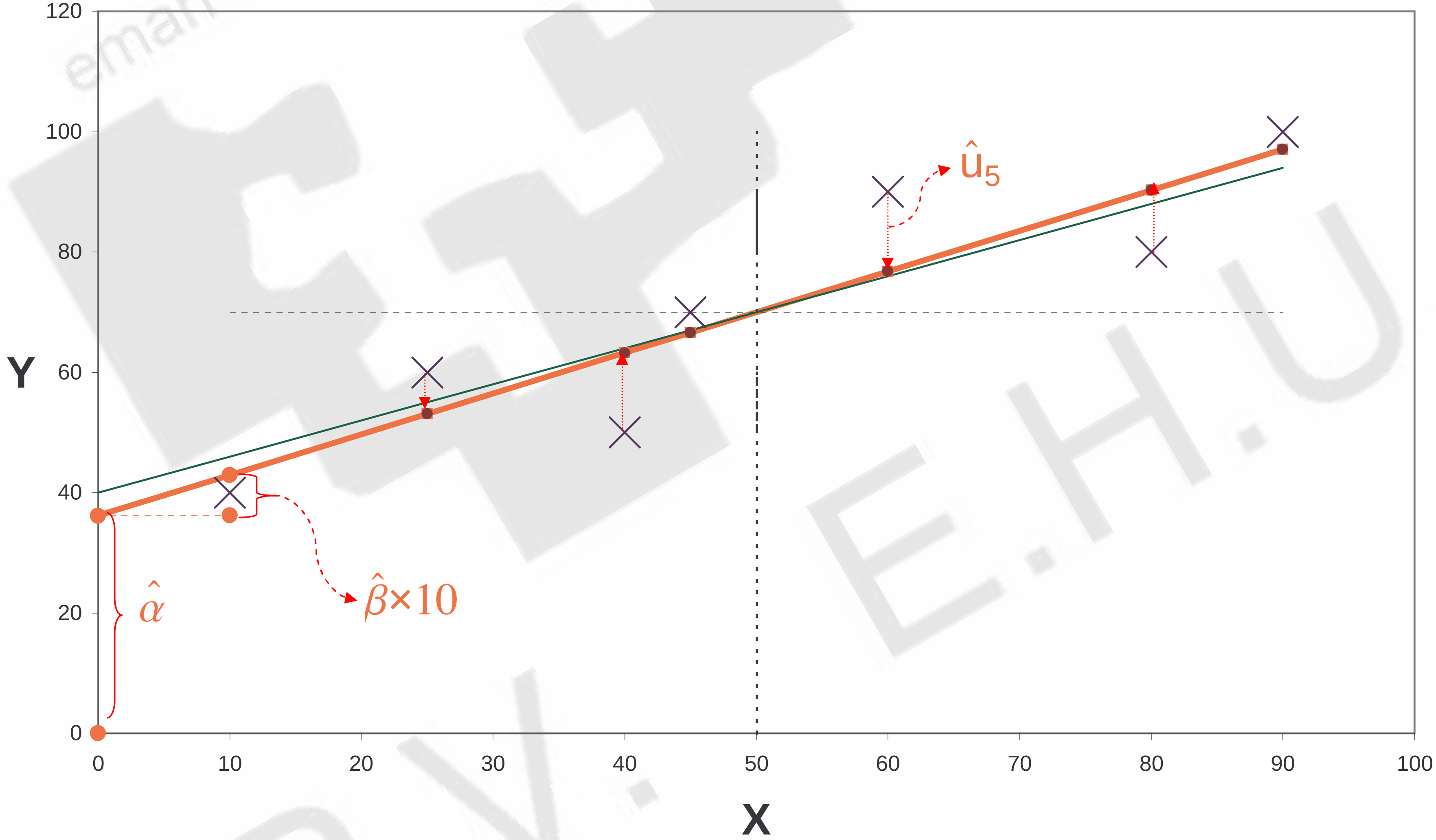
	Y	X	y	x	y^2	x^2	yx
	40	10	-30	-40	900	1600	1200
	60	25	-10	-25	100	625	250
	50	40	-20	-10	400	100	200
	70	45	0	-5	0	25	0
	90	60	20	10	400	100	200
	80	80	10	30	100	900	300
	100	90	30	40	900	1600	1200
Sum					2800	4950	3350
Average	70	50	0	0	400	707.14	478.57

$\widehat{\alpha} = 36.162 \; (= \overline{Y} - \widehat{\beta}\overline{X})$

$\widehat{\beta} = 0.677 \left(= \frac{\text{Cov}(Y,X)}{\text{Var}(X)} \right)$

Can also use formulae based on original data... (Exercise: Try it!!)

Numerical example: strawberry regres plot



2.4a Properties of the Sample Regression Function.

Properties of residuals and SRF (1)

$$\widehat{\beta}_{\text{OLS}} \rightsquigarrow \widehat{\alpha}_{\text{OLS}} \rightsquigarrow \widehat{Y}_t = \widehat{\alpha} + \widehat{\beta} X_t \rightsquigarrow \widehat{u}_t = Y_t - \widehat{Y}_t$$

1. residuals add up to zero: $\sum \widehat{u}_t = 0$
   Demo: directly from 1st.o.c. □
2. $\overline{\widehat{Y}} = \overline{Y}$
   Demo: by def.: $\widehat{u}_t = Y_t - \widehat{Y}_t \rightsquigarrow \overline{\widehat{Y}} = \overline{Y} - \overline{\widehat{u}}$,
   but $\overline{\widehat{u}} = \frac{1}{T} \sum \widehat{u}_t = 0$ (from prop 1) $\rightsquigarrow \overline{\widehat{Y}} = \overline{Y}$. □
3. the SRF passes thru the pair of means $(\overline{X}, \overline{Y})$:
   $\overline{Y} = \widehat{\alpha} + \widehat{\beta} \overline{X}$
   Demo: from $\widehat{\alpha} = \overline{Y} - \widehat{\beta} \overline{X}$ (1st. normal eq.) □

Properties of residuals and SRF (2)

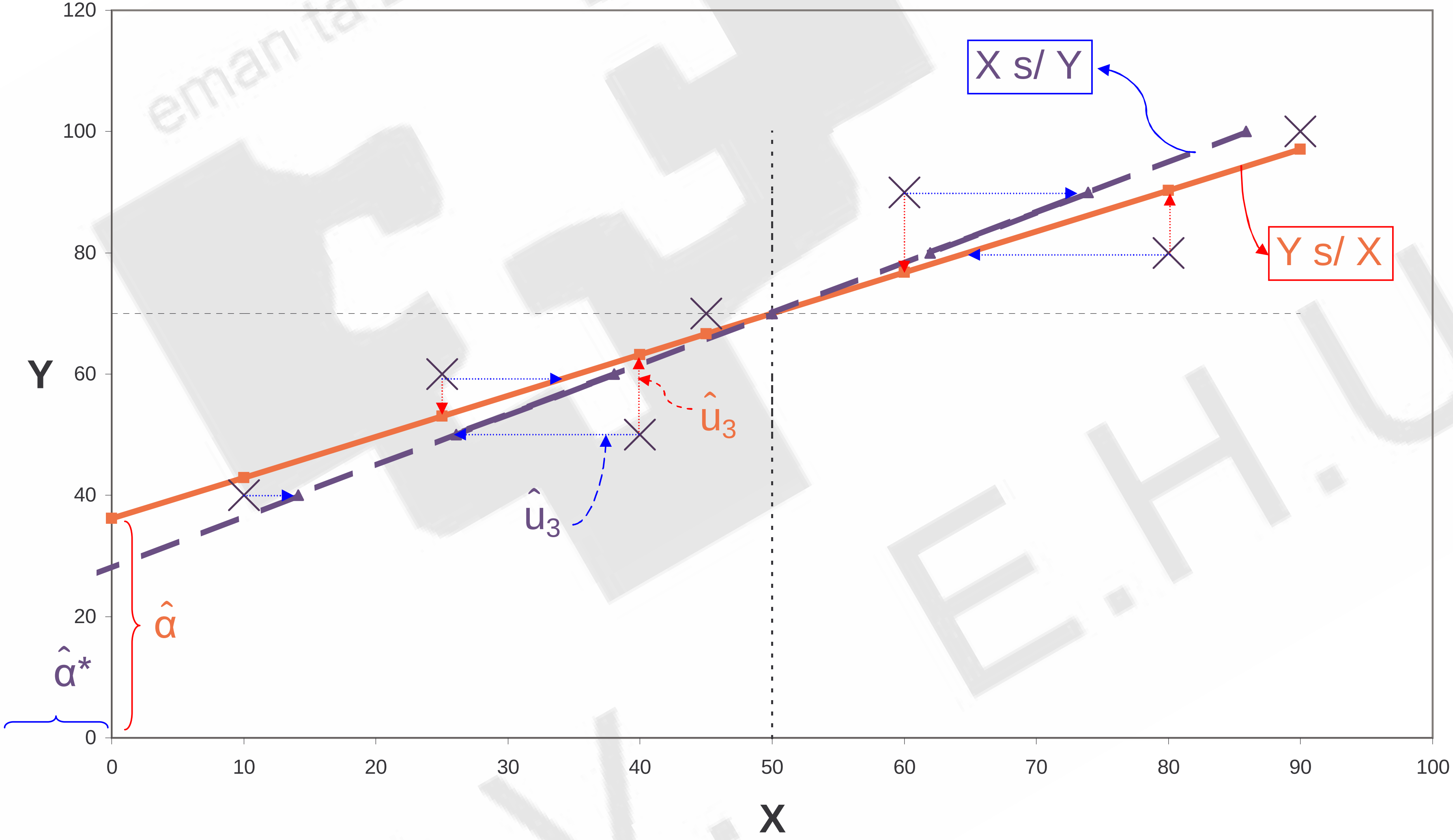
4. residuals orthogonal to expl. v. X: $\sum X_t \widehat{u}_t = 0$
   Demo: directly from 1st.o.c. □

5. residuals orthogonal to the explained part of Y: $\sum \widehat{Y}_t \widehat{u}_t = 0$
   Demo: $\sum (\widehat{\alpha} + \widehat{\beta} X_t) \widehat{u}_t =$

$$\widehat{\alpha} \underbrace{\sum \widehat{u}_t}_{=0 \text{ (from prop 1)}} + \widehat{\beta} \underbrace{\sum X_t \widehat{u}_t}_{=0 \text{ (from prop 4)}} = 0$$

□

Causality: Y on X vs X on Y



Properties of residuals and SRF (5)

8. $\widehat{\alpha}_{\text{OLS}}$ and $\widehat{\beta}_{\text{OLS}}$ unbiased $\rightsquigarrow$ expected value $=$ true value!

Demo:

- $$\widehat{\beta} = \frac{\sum y_t x_t}{\sum x_t^2}$$

$$\mathsf{E}\left(\widehat{\beta}\right) = \frac{1}{\sum x_t^2} \sum \underbrace{\mathsf{E}\left(y_t\right)}_{\beta x_t} x_t = \frac{1}{\sum x_t^2} \; \beta \sum x_t^2$$

$$\boxed{\mathsf{E}\left(\widehat{\beta}\right) = \beta}$$

- $$\widehat{\alpha} = \overline{Y} - \widehat{\beta}\,\overline{X}$$

$$\mathsf{E}\left(\widehat{\alpha}\right) = \frac{1}{T} \sum \mathsf{E}\left(Y_t\right) - \mathsf{E}\left(\widehat{\beta}\right)\overline{X}$$

$$= \frac{1}{T} \sum \left(\alpha + \beta X_t\right) - \beta \overline{X} = \alpha + \beta \overline{X} - \beta \overline{X}$$

$$\boxed{\mathsf{E}\left(\widehat{\alpha}\right) = \alpha}$$

2.5a Goodness of Fit: the Coefficient of Determination (R^2).

Goodness of fit: Coefficient of determination

- Sum-of-Squares decomposition:

$$\sum Y_t^2 = \sum(\widehat{Y}_t^2 + \widehat{u}_t^2 + 2\widehat{Y}_t\widehat{u}_t) = \sum \widehat{Y}_t^2 + \sum \widehat{u}_t^2 \qquad \text{(from prop 5)}$$

- $$\sum Y_t^2 - T\overline{Y}^2 = \sum \widehat{Y}_t^2 - T\overline{\widehat{Y}}^2 + \sum \widehat{u}_t^2 \qquad \text{(from prop 2)}$$

- $$\underset{(TSS)}{\sum y_t^2} = \underset{(ESS)}{\sum \widehat{y}_t^2} + \underset{(RSS)}{\sum \widehat{u}_t^2}$$

- Definition of R^2:

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

$0 \leq R^2 \leq 1$ (Interpretation in terms of total variance??)

No intercept $\rightsquigarrow$ invalid R^2

SLRM: $Y_t = \beta X_t + u_t$,

- apply Least-Squares fit:

$$\min_{\beta} \sum_{t=1}^{T} u_t^2 \quad \text{where} \quad u_t = Y_t - \beta X_t :$$

- First derivatives:

$$\frac{\partial \sum u_t^2}{\partial \beta} = 2 \sum u_t \frac{\partial u_t}{\partial \beta} = 2 \sum u_t (-X_t)$$

- 1st.o.c. (minimum) $\Rightarrow$ first derivative $=$ zero:

$$\sum \widehat{u}_t X_t = \sum (Y_t X_t - \widehat{\beta} X_t^2) = 0$$

-

$$\nexists \text{ 1st equation!!} \rightsquigarrow \begin{cases} \sum \widehat{u}_t \neq 0, \\ \overline{\widehat{Y}} \neq \overline{Y}, \end{cases} \rightsquigarrow \text{invalid } R^2 \quad \text{(Why?)}$$

Relationship of R^2 with correlation coef

$$R^2 = \frac{\frac{1}{T}\sum \widehat{y}_t^2}{\frac{1}{T}\sum y_t^2} = \frac{\frac{1}{T}\sum(\widehat{\beta} x_t)^2}{\frac{1}{T}\sum y_t^2} = \frac{\widehat{\beta}^2 \frac{1}{T}\sum x_t^2}{\frac{1}{T}\sum y_t^2}$$

$$= \widehat{\beta}^2 \frac{\text{Var}(X)}{\text{Var}(Y)} = \frac{\text{Cov}(Y,X)^2}{\text{Var}(X)^2} \frac{\text{Var}(X)}{\text{Var}(Y)}$$

$$= \frac{\text{Cov}(Y,X)^2}{\text{Var}(X)\text{Var}(Y)}$$

$$R^2 = r_{X,Y}^2$$

Numerical example: strawberry prod data (cont)

- recall data & previous calculations...
- do the same for fitted values...
- now calculate R^2...

	y^2	$\widehat{Y}$	$\widehat{y}$	$\widehat{y}^2$	$\widehat{u}$	$\widehat{u}^2$
	900	42.92	-27.07	732.82	-2.92	8.58
	100	53.08	-16.91	286.25	6.91	47.87
	400	63.23	-6.76	45.80	-13.23	175.09
	0	66.61	-3.38	11.45	3.38	11.45
	400	76.76	6.76	45.80	13.23	175.09
	100	90.30	20.30	412.21	-10.30	106.15
	900	97.07	27.07	732.82	2.92	8.58
Average	400	70	0	323.88		
Sum	2800			2267.17		532.82
	TSS			ESS		RSS

$$R^2 = 0.8097 \left(= \frac{\text{ESS}}{\text{TSS}} = 1 - \frac{\text{RSS}}{\text{TSS}}\right)$$

(Exercise: How does this compare with Corr(X,Y)? ... Try it!!)

2.3b OLS in the GLRM.

GLRM: the PRF

- Recall: model with K explanatory variables:

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 X_{1t} + \cdots + \beta_K X_{Kt} + u_t\,, \\ Y &= X\beta + u \end{aligned} \qquad (2)$$

  is called GLRM.

- Population Regression Function (PRF):
  $E(u) = 0 \rightsquigarrow$ *systematic part* or PRF:

$$\begin{aligned} E(Y_t) &= \beta_0 + \beta_1 X_{1t} + \cdots + \beta_K X_{Kt} \\ E(Y) &= X\beta \end{aligned}$$

- Interpretation of the coefficients:
  - $\beta_0 = E(Y_t | X_{1t} = X_{2t} = \cdots = X_{Kt} = 0)$: Expected value of Y_t when all explanatory variables are equal to zero.
  - $\beta_k = \dfrac{\partial E(Y_t)}{\partial X_{kt}} \simeq \dfrac{\Delta E(Y_t)}{\Delta X_{kt}}, \quad k = 1 \ldots K$: Increase in (expected) value Y_t when $X_k \uparrow$ one unit (c.p.).

The Sample Regression Function (SRF)

- Objective of GLRM: To obtain estimator $\widehat{\beta} = (\widehat{\beta}_0, \widehat{\beta}_1 \ldots, \widehat{\beta}_K)'$ of unknown parameter vector in (2).
  $\widehat{\beta} \rightsquigarrow$ estimated model, fit or SRF:

$$\widehat{Y}_t = \widehat{\beta}_0 + \widehat{\beta}_1 X_{1t} + \cdots + \widehat{\beta}_K X_{Kt}$$

$$\widehat{Y} = X\widehat{\beta}$$

- Notes:
  - Disturbances in PRF:

$$u_t = Y_t - E(Y_t) = Y_t - \beta_0 - \beta_1 X_{1t} - \cdots - \beta_K X_{Kt}$$

$$u = Y - E(Y) = Y - X\beta$$

  - Residuals in SRF:

$$\widehat{u}_t = Y_t - \widehat{Y}_t = Y_t - \widehat{\beta}_0 - \widehat{\beta}_1 X_{1t} - \cdots - \widehat{\beta}_K X_{Kt}$$

$$\widehat{u} = Y - \widehat{Y} = Y - X\widehat{\beta}$$

- Residuals are to the SRF what disturbances are to the PRF.

Estimation: OLS

- apply Least-Squares fit to GLRM: $Y = X\beta + u$,
- either in observation form:

$$\min_{\beta_0 \ldots \beta_K} \sum_{t=1}^{T} u_t^2 \text{ where } u_t = Y_t - \beta_0 - \beta_1 X_{1t} - \cdots - \beta_K X_{Kt}$$

- or in matrix form:

[recall:

$$u' = (u_1, u_2, \ldots, u_T) \qquad u = \begin{pmatrix} u_1 \\ u_2 \\ \ldots \\ u_T \end{pmatrix}$$

so $u'u = u_1^2 + u_2^2 + \cdots + u_T^2 = \sum_{t=1}^{T} u_t^2$]

- that is

$$\boxed{\min_{\beta} u'u \quad \text{where} \quad u = Y - X\beta}$$

Note: vector derivatives

- Let $u = u(\beta)$: derivs of cu and cu^2 with respect to β:

$$\frac{\partial}{\partial \beta}(cu) = c\frac{\partial u}{\partial \beta} \qquad \text{and} \qquad \frac{\partial}{\partial \beta}u^2 = 2u\frac{\partial u}{\partial \beta}$$

- With vectors and matrices this is quite similar:

- The derivative of the linear combination $u'c$

$$\underset{(1\times n)}{u'}\ \underset{(n\times 1)}{c} \qquad (=\textstyle\sum_{i=1}^n c_i u_i,\ \textit{i.e.}\ \text{scalar!!})$$

with respect to $\underset{(k\times 1)}{\beta}$ is: $\frac{\partial(u'c)}{\partial \beta} = \frac{\partial u'}{\partial \beta}c$

- The derivative of the sum of squares $u'u$

$$\underset{(1\times n)}{u'}\ \underset{(n\times 1)}{u} \qquad (=\textstyle\sum_{i=1}^n u_i^2,\ \textit{i.e.}\ \text{scalar!!})$$

with respect to $\underset{(k\times 1)}{\beta}$ is: $\frac{\partial(u'u)}{\partial \beta} = 2\frac{\partial u'}{\partial \beta}u$

1st.o.c. in matrix form

$$\min_{\beta}(u'u) \quad \text{where} \quad u = Y - X\beta$$

First derivatives of SS $u'u$ with respect to β:

$$\begin{aligned}\frac{\partial u'u}{\partial \beta} &= 2\,\frac{\partial u'}{\partial \beta}u \\ &= 2\,\frac{\partial(Y' - \beta' X')}{\partial \beta}u \\ &= -2X'u\end{aligned}$$

in the minimum:

$$\text{1st.o.c.:} \quad \underset{(K+1 \times T)}{X'} \; \underset{(T \times 1)}{\widehat{u}} = 0_{K+1}$$

Estimation: Normal equations & LSE of β

Solving the 1st.o.c. we obtain the normal equations: $X'(Y - X\widehat{\beta}) = 0 \;\Rightarrow$

$$\underset{(K+1 \times 1)}{X'Y} = \underset{(K+1 \times K+1)}{X'X} \; \underset{(K+1 \times 1)}{\widehat{\beta}} \tag{3}$$

Whence premultiplying by $(X'X)^{-1}$ we obtain the OLS estimator:

$$\widehat{\beta}_{\text{OLS}} = (X'X)^{-1}X'Y$$

Estimation: LSE of β (cont)

- where $X'X$ is a $[K{+}1 \times K{+}1]$ matrix: [recall X & Y? $\longrightarrow$]

- $$\underset{(K+1\times K+1)}{X'X} = \begin{bmatrix} T & \sum X_{1t} & \sum X_{2t} & \dots & \sum X_{Kt} \\ \sum X_{1t} & \sum X_{1t}^2 & \sum X_{1t}X_{2t} & \dots & \sum X_{1t}X_{Kt} \\ \dots & \dots & \dots & \dots & \dots \\ \sum X_{Kt} & \sum X_{Kt}X_{1t} & \sum X_{Kt}X_{2t} & \dots & \sum X_{Kt}^2 \end{bmatrix}$$

- and $X'Y$ and $\widehat{\beta}$ are $[K{+}1 \times 1]$ vectors:

$$\underset{(K+1\times 1)}{X'Y} = \begin{bmatrix} \sum Y_t \\ \sum X_{1t}Y_t \\ \dots \\ \sum X_{Kt}Y_t \end{bmatrix} \qquad \underset{(K+1\times 1)}{\widehat{\beta}} = \begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \dots \\ \widehat{\beta}_K \end{bmatrix}$$

OLS estimator with centred (demeaned) data

An alternative way to obtain the OLS estimator is

$$\widehat{\beta}^{\star}_{\mathrm{OLS}} = (x'x)^{-1}x'y$$

for the model coefficients.

. . . together with the estimated intercept obtained from the first normal equation

$$\widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1\overline{X}_1 - \cdots - \widehat{\beta}_K\overline{X}_K$$

Note: special case with $K = 1 \rightsquigarrow$ identical formulae as in SLRM!! (Prove it!!)

2.4b Properties of the SRF.

Properties of residuals and SRF (1)

$$\left.\begin{array}{c}\widehat{\beta} \\ \widehat{\beta}^{\star} \rightsquigarrow \widehat{\beta}_0\end{array}\right\} \rightsquigarrow \widehat{Y} = X\widehat{\beta} \rightsquigarrow \widehat{u} = Y - \widehat{Y}$$

1. residuals add up to zero: $\sum \widehat{u}_t = 0$

   Demo: directly from 1st.o.c.:

$$X'\widehat{u} = 0 \Rightarrow \begin{bmatrix} \sum_1^T \widehat{u}_t \\ \sum_1^T X_{1t}\widehat{u}_t \\ \sum_1^T X_{2t}\widehat{u}_t \\ \dots \\ \sum_1^T X_{Kt}\widehat{u}_t \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \dots \\ 0 \end{bmatrix}$$

□

2. $\overline{\widehat{Y}} = \overline{Y}$
3. the SRF passes thru vector $(\overline{X}_1, \dots \overline{X}_K, \overline{Y})$: $\overline{Y} = \widehat{\beta}_0 + \widehat{\beta}_1\overline{X}_1 + \cdots + \widehat{\beta}_K\overline{X}_K$

Note: These properties 1 thru 3 are fulfilled if the regression has an intercept; that is, if X has a column of "ones".

Properties of residuals and SRF (2)

4. residuals orthogonal to explanatory v.: $X'\widehat{u} = 0$

   Demo: directly from 1st.o.c. (see 1) or, alternatively:

$$X'\widehat{u} = X'(Y - X\widehat{\beta}) = X'Y - X'X\widehat{\beta}$$
$$= X'Y - \underbrace{X'X(X'X)^{-1}}_{=I_{K+1}} X'Y = 0$$

□

5. residuals orthogonal to explained part of Y: $\widehat{Y}'\widehat{u} = 0$

   Demo: $\widehat{Y}'\widehat{u} = (X\widehat{\beta})'\widehat{u} = \widehat{\beta}' \underbrace{X'\widehat{u}}_{=0} = 0$ □

2.5b Goodness of Fit: Coefficient of Determination (R^2) & Estimation of the Error Variance.

Goodness of fit: R^2 Revisited

Recall (same as before but now we'll do it with vectors):

$$
\begin{aligned}
Y'Y &= (\widehat{Y}' + \widehat{u}')(\widehat{Y} + \widehat{u}) \\
&= \widehat{Y}'\widehat{Y} + \widehat{u}'\widehat{u} + 2\widehat{Y}'\widehat{u} \\
&= \widehat{Y}'\widehat{Y} + \widehat{u}'\widehat{u} \quad \text{(from prop 5)}
\end{aligned}
$$

$$
Y'Y - T\overline{Y}^2 = \widehat{Y}'\widehat{Y} - T\overline{\widehat{Y}}^2 + \widehat{u}'\widehat{u} \qquad \text{(from prop 2)}
$$

$$
\boxed{\underset{(TSS)}{\underset{\downarrow}{y'y}} = \underset{(ESS)}{\underset{\downarrow}{\widehat{y}'\widehat{y}}} + \underset{(RSS)}{\underset{\downarrow}{u'u}}}
$$

$$
R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}
$$

$$
0 \leq R^2 \leq 1
$$

Goodness of fit: R^2 Revisited (cont)

Note 1: R^2 measures the proportion of the dependent variable variation explained by the variation of (a linear combination of) the explanatory variables.

Note 2:

$$\text{no intercept} \Rightarrow \begin{cases} \nexists \text{1st row of 1st.o.c.} \rightsquigarrow \begin{cases} \sum \widehat{u}_t \neq 0, \\ \overline{\widehat{Y}} \neq \overline{Y}, \end{cases} \\ \text{not valid} R^2 \text{ (Remember!)} \end{cases}$$

Estimation of $\text{Var}(u_t)$

$$\sigma^2 = \text{Var}(u_t) = \text{E}(u_t^2) \simeq \frac{1}{T}\sum_{t=1}^{T} u_t^2$$

but with residuals, they must satisfy $K+1$ linear relationships in $X'\widehat{u} = 0$ so we loose $K+1$ degrees of freedom:

$$\widehat{\sigma}^2 = \frac{1}{T-K-1}\sum_{t=1}^{T} \widehat{u}_t^2$$

Therefore we propose the following estimator:

$$\boxed{\widehat{\sigma}^2 = \frac{\text{RSS}}{T-K-1}}$$

which clearly is an unbiased estimator:

Demo:

$$\text{E}(\widehat{\sigma}^2) = \frac{\text{E}(\text{RSS})\,(^*)}{T-K-1} = \frac{T-K-1}{T-K-1} = \sigma^2$$

(* see textbook)

2.6 Finite-sample Properties of the Least-Squares Estimator.
The Gauss-Markov Theorem.

Properties of the OLS Estimator (1)

The estimator $\widehat{\beta}_{\text{OLS}} = (X'X)^{-1}X'Y$ has the following properties:

- **Linear:** $\widehat{\beta}_{\text{OLS}}$ is a linear combination of disturbances:

$$\begin{aligned}\widehat{\beta} &= (X'X)^{-1}X'(X\beta + u) \\ &= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u \\ &= \beta + (X'X)^{-1}X'u \\ &= \beta + \Gamma'u\end{aligned}$$

- **Unbiased:** Since $\mathsf{E}(u) = 0$, $\widehat{\beta}_{\text{OLS}}$ is unbiased:

$$\begin{aligned}\mathsf{E}(\widehat{\beta}) &= \mathsf{E}(\beta + \Gamma'u) \\ &= \beta + \Gamma'\mathsf{E}(u) \\ &= \beta\end{aligned}$$

Properties of the OLS Estimator (2)

- **Variance:** Recall:

$$\text{Var}(u) = \sigma^2 I_T,$$

$$\widehat{\beta} = \beta + (X'X)^{-1}X'u,$$

$$\begin{aligned}\text{Var}(\widehat{\beta}) &= \text{E}\big((\widehat{\beta}-\beta)(\widehat{\beta}-\beta)'\big)\\ &= \text{E}\big((X'X)^{-1}X'uu'X(X'X)^{-1}\big)\\ &= (X'X)^{-1}X'\,\text{E}(uu')\,X(X'X)^{-1}\\ &= (X'X)^{-1}X'\,\sigma^2 I_T\,X(X'X)^{-1}\\ &= \sigma^2(X'X)^{-1}X'X(X'X)^{-1}\end{aligned}$$

$$\boxed{\text{Var}(\widehat{\beta}) = \sigma^2(X'X)^{-1}}$$

Properties of the OLS Estimator (2cont)

$$\text{Var}(\widehat{\beta}) = \begin{bmatrix} \text{Var}(\widehat{\beta}_0) & \text{Cov}(\widehat{\beta}_0, \widehat{\beta}_1) & \ldots & \text{Cov}(\widehat{\beta}_0, \widehat{\beta}_K) \\ \text{Cov}(\widehat{\beta}_1, \widehat{\beta}_0) & \text{Var}(\widehat{\beta}_1) & \ldots & \text{Cov}(\widehat{\beta}_1, \widehat{\beta}_K) \\ \ldots & \ldots & \ldots & \ldots \\ \text{Cov}(\widehat{\beta}_K, \widehat{\beta}_0) & \text{Cov}(\widehat{\beta}_K, \widehat{\beta}_1) & \ldots & \text{Var}(\widehat{\beta}_K) \end{bmatrix}$$

$$\sigma^2 (X'X)^{-1} = \sigma^2 \begin{bmatrix} a_{00} & a_{00} & a_{01} & \ldots & a_{0K} \\ a_{10} & a_{11} & a_{12} & \ldots & a_{1K} \\ a_{20} & a_{21} & a_{22} & \ldots & a_{2K} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{K0} & a_{K1} & a_{K2} & \ldots & a_{KK} \end{bmatrix}$$

i.e. a_{kk} is the $(k+1, k+1)$-element of matrix $(X'X)^{-1}$:

$$\text{Var}(\widehat{\beta}_k) = \sigma^2 a_{kk}$$

$$\text{Cov}(\widehat{\beta}_k, \widehat{\beta}_i) = \sigma^2 a_{ki}$$

The Gauss-Markov Theorem

"Given the basic assumptions of GLRM, the OLS estimator is that of minimum variance (best) among all the linear and unbiased estimators"

$$\widehat{\beta}_{\text{OLS}} = \textbf{BLUE} = \textbf{B}\text{est}\,\textbf{L}\text{inear}\,\textbf{U}\text{nbiased}\,\textbf{E}\text{stimator}$$

Demo:

Let $\widetilde{\beta}$ be some other linear and unbiased estimator:

$$\widetilde{\beta} = D'Y = D'(X\beta + u) = D'X\beta + D'u$$

$$\mathsf{E}(\widetilde{\beta}) = D'X\beta + D'\mathsf{E}(u) = D'X\beta = \beta \Rightarrow \boxed{D'X = I_K}$$

then $\widetilde{\beta} = \beta + D'u \quad \rightsquigarrow \quad \widetilde{\beta} - \beta = D'u$
and its variance:

$$\text{Var}(\widetilde{\beta}) = E\left[(\widetilde{\beta} - \beta)(\widetilde{\beta} - \beta)'\right] = \mathsf{E}(D'uu'D)$$

$$= D'\,\mathsf{E}(uu')\,D = D'\,\sigma^2 I_T\,D = \sigma^2 D'D$$

The Gauss-Markov Theorem (cont)

. . . The difference between both covariance matrices is a positive definite matrix:

$$\begin{aligned}
\text{Var}\left(\widetilde{\beta}\right) - \text{Var}\left(\widehat{\beta}\right) &= \sigma^2 D'D - \sigma^2 (X'X)^{-1} \\
&= \sigma^2 \left[D'D - (X'X)^{-1}\right] \\
&= \sigma^2 \left[D'D - D'X (X'X)^{-1} X'D\right] \\
&= \sigma^2 D' \underbrace{\left[I_T - X (X'X)^{-1} X'\right]}_{M} D \\
&= \sigma^2 D' (MM) D \\
&= \sigma^2 (D'M)(M'D) = D^{\star\prime} D^{\star} \\
&> 0
\end{aligned}$$

I.e. in particular **all** individual variances will be bigger than their OLS counterpart.

2.3c OLS: Useful expressions & Timeline.

Useful expressions for SS

- $$TSS = \sum (Y_t - \overline{Y})^2 = \sum Y_t^2 - T\overline{Y}^2 = Y'Y - T\overline{Y}^2$$

- $$ESS = \sum (\widehat{Y}_t - \overline{\widehat{Y}})^2 = \sum \widehat{Y}_t^2 - T\overline{\widehat{Y}}^2 = \sum \widehat{Y}_t^2 - T\overline{Y}^2 = \widehat{Y}'\widehat{Y} - T\overline{Y}^2$$
$$= (X\widehat{\beta})'(X\widehat{\beta}) - T\overline{Y}^2 = \widehat{\beta}' \underbrace{X'X\widehat{\beta}}_{X'Y} - T\overline{Y}^2 = \widehat{\beta}'X'Y - T\overline{Y}^2$$

- $$RSS = \sum \widehat{u}_t^2 = \widehat{u}'\widehat{u} \qquad = \sum Y_t^2 - \sum \widehat{Y}_t^2 = Y'Y - \widehat{\beta}'X'Y$$

Main expressions & Timeline

- $Y = X\beta + u$
- $(X'X)^{-1} \qquad X'Y$
- $\widehat{\beta} = (X'X)^{-1}X'Y$
- $ESS = \widehat{\beta}'X'Y - T\overline{Y}^2$ (needs $\overline{Y}$!)
- $TSS = Y'Y - T\overline{Y}^2$
- $RSS = Y'Y - \widehat{\beta}'X'Y$ (no $\overline{Y}$!)
- $R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$
- $\widehat{\sigma}^2 = \frac{RSS}{T-K-1}$
- $\widehat{\text{Var}(\widehat{\beta})} = \widehat{\sigma}^2(X'X)^{-1}$

2.7a Omission of relevant variables.

Omission of relevant variables

- true relationship:

$$Y = X\beta + u = \left[X_I \;\middle|\; X_{II} \right] \begin{pmatrix} \beta_I \\ \hline \beta_{II} \end{pmatrix} + u$$

$$X = \left[\begin{array}{cccc|ccc} 1 & X_{11} & \cdots & X_{K_1,1} & X_{K_1+1,1} & \cdots & X_{K1} \\ 1 & X_{12} & \cdots & X_{K_1,2} & X_{K_1+1,2} & \cdots & X_{K2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & X_{1T} & \cdots & X_{K_1,T} & X_{K_1+1,T} & \cdots & X_{KT} \end{array} \right], \; \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{K_1} \\ \hline \beta_{K_1+1} \\ \vdots \\ \beta_K \end{pmatrix}.$$

$$Y = X_I\beta_I + X_{II}\beta_{II} + u$$

- estimated relationship:

$$Y = X_I\beta_I + v \qquad \text{where} \quad v = X_{II}\beta_{II} + u,$$

$$\text{then} \quad \mathrm{E}(v) \neq 0 \quad \rightsquigarrow \quad \mathrm{E}(\widehat{\beta}) \neq \beta.$$

i.e. $\widehat{\beta}$ is biased.

Omission of relevant variables: consequences

Summary:

- OLS estimator of coefficients is *biased* (except if $x_I' x_{II} = 0$).
- OLS estimator of intercept is *always biased*.
- Estimator of Error variance is *always biased*.

2.7b Multicollinearity

Perfect Multicollinearity

Extreme case:

- exact linear combination:
  - $\sum_{k=0}^{K} \lambda_k X_{kt} = 0, \quad \lambda \neq 0, \quad X_{0t} = 1,$
  - $\exists X_i \mid X_i = \lambda_0^* + \sum_{\substack{k=1 \\ k \neq i}}^{K} \lambda_k^* X_{kt},$
  - $\exists X_i, X_j \mid \text{Corr}\left(X_i, X_j\right) = 1,$
  - $\exists X_i \mid$ aux regres X_i on $\{X_k\}_{\substack{k=1 \\ k \neq i}}^{K} \quad \rightsquigarrow \quad \mathrm{R}_i^2 = 1.$
- Problem:
  - $\text{rk}\, X < K+1$, (X isn't of full rank)
  - $\rightsquigarrow \quad \det(X) = 0$
  - $\rightsquigarrow \quad \nexists (X'X)^{-1}$
  - $\rightsquigarrow$

$\widehat{\beta}$?

Perfect Multicollinearity: example

- Let $X_{4t} = 2X_{1t} \quad \forall t$:

$$X_{4t} = 0 + 2X_{1t} + 0 \cdot X_{2t} + 0 \cdot X_{3t} + 0 \cdot X_{5t} + \cdots + 0 \cdot X_{Kt},$$

- no error? $\Rightarrow$ aux regres X_4 on $\{X_k\}_{\substack{k=1 \\ k \neq 4}}^{K} \quad \rightsquigarrow \mathrm{R}_4^2 = 1!!$

- Model specification:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \cdots + u_t, t = 1, 2 \ldots, T,$$
$$X_{4t} = 2X_{1t},$$

- and substituting in model:

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 (2X_{1t}) + \cdots + u_t, \\ &= \beta_0 + \underbrace{(\beta_1 + 2\beta_4)}_{\beta_1^\star} X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \cdots + u_t \end{aligned}$$

- now we have one less parameter to estimate.

Multicollinearity: counterexample

$$Y_t = \beta_0 + \beta_1^{\star} X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \cdots + u_t$$

- Just K parameters remain to be estimated,
  but β_1 and β_4 cannot be estimated separately:
  - we can just estimate a linear combination of them:
    $$\beta_1^{\star} = \beta_1 + 2\beta_4,$$
  - *i.e.* combined effect of X_{1t} and X_{4t} on Y_t!!

- (Exercise: Try it yourself with $X_{2t} - 3X_{3t} = 10, \quad \forall t$.)

- multicollinearity $=$ *linear relationships*
  but... what if relationship *isn't linear*? *e.g.:*
  $$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{1t}^2 + u_t$$
  - X is of full column rank $\rightsquigarrow$ no problem.

Perfect Multicollinearity: consequences

- some parameters cannot be estimated separately.
- some estimates are just l.c. of parameters.
- R^2 is correct:
  correctly picks up proportion of (variance of) Y_t explained by the regression.
- Predictions of Y are still valid.

2.7c Imperfect Multicollinearity

Imperfect Multicollinearity

- Problem:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \cdots + u_t, t = 1, 2 \ldots, T,$$

$$X_{4t} = 2X_{1t} + v_t,$$

$v_t =$ gap between X_{4t} and $2X_{1t}$,

- approximate relationship:
- auxiliary regression X_{4t} on rest $\rightsquigarrow$ $R^2 \approx 1$.
- it's a matter of degree ($x'x$ not diagonal $\rightsquigarrow$ correlated variables)
- Note: whenever perfect/imperfect is not specified we mean imperfect mc.

Multicollinearity: Symptoms

- Typical symptom:
  - high R^2

    (relevant group of regressors)
  - but they appear to be not relevant individually

    (inability to separate effects of regressors).
- more formally:

$$\text{Var}(\widehat{\beta}^{\star}) = \sigma^2 (x'x)^{-1} = \frac{\sigma^2}{T}\text{Var}(X^{\star})^{-1}$$

$$\Rightarrow \quad \text{Var}(\widehat{\beta}_k) = \frac{\sigma^2}{T\,\text{Var}(X_k)(1-R_k^2)},$$

- so that, in the previous example $X_{4t} \approx 2X_{1t}$:
  - $\text{Corr}(X_4, X_1) \uparrow$
  - R_4^2 and $R_1^2 \upuparrows$
  - denominator $\downarrow$
  - variances $\upuparrows$

Multicollinearity: Consequences

- Some coefficients aren't significant, even if their variables have an important effect on dependent variable.
- Nevertheless, Gauss-Markov
  $\Rightarrow$ linear, unbiased and of minimum variance estimators,
  then *it isn't possible to find a Better LUE.*
- R^2 is correct:
  correctly picks up proportion of (variance of) Y_t
  explained by the regression.
- Predictions of Y are still valid.

Multicollinearity: How to detect

- Small changes in data
  ⇒ important changes in estimates
  (they can even affect their signs).
- Coefficient estimations
  not individually significant. . .
- . . . but they are jointly significant.
- High coefficient of determination R^2.
- Auxiliary regressions among regressors
  ⇒ high R_k^2.

Multicollinearity: Some solutions

Multicollinearity is not an easy problem to solve.
Nevertheless, from

$$\text{Var}(\widehat{\beta}_k) = \frac{\sigma^2}{T\,\text{Var}(X_k)(1 - \text{R}_k^2)},$$

it turns out that to lower the variance we may:

- $\mathbf{T}\uparrow$: Increase number of observations T .
  Also, differences among regressors may increase.
- $\text{Var}(\mathbf{X})\uparrow$: Increase data dispersion; *e.g.* study about consumption function:
  sample of families $\leftrightsquigarrow$ all possible incomes.
- $\text{Var}(\mathbf{X})\uparrow$:Include additional information.
  e.g. impose restrictions suggested by Ec. Th.
- $\sigma^{\mathbf{2}}\downarrow$: Add new relevant regressor not yet included.
  It would also avoid serious bias problems.
- $\mathbf{R_k^2}\downarrow$: Eliminate variables that may produce multicollinearity.
  (Take care of omitting some relevant regressor though).

2.8 The OLS Estimator under Restrictions.

GLRM under linear restrictions (1)

- previous chapter objectives:
  - Econometric model (GLRM), characteristics and basic assumptions. . .
  - but. . . no knowledge about model parameters.
  - Least Squares Method for parameter estimation (OLS).
  - Properties of resulting estimators.
- present chapter objectives:
  - *a priori* information about parameter values (or l.c.) . . .
  - given by
    - economic theory,
    - other empirical work,
    - own experience, etc.
  - Non-Restricted Model $\Rightarrow$ Ordinary LS.
  - Restricted Model $\Rightarrow$ Restricted LS.
  - Check, given the estimated model, if the information is compatible with available data.

GLRM under linear restrictions: examples

- production function with constant returns to scale: $\beta_K + \beta_L = 1$.
- product demands as function of price: $\beta = -1$ (say).
- in GLRM: let us assume that $\beta_2 = 0$ and $2\beta_3 = \beta_4 - 1$:
  - Full model:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \cdots + \beta_{Kt} X_{Kt} + u_t, \text{ with } \beta_2 = 0 \text{ and } 2\beta_3 + 1 = \beta_4;$$

  - Alternative transformed model:

$$Y_t = \beta_0 + \beta_1 X_{1t} + 0X_{2t} + \beta_3 X_{3t} + (2\beta_3 + 1) X_{4t} + \cdots + \beta_{Kt} X_{Kt} + u_t$$
$$Y_t - X_{4t} = \beta_0 + \beta_1 X_{1t} + \beta_3 (X_{3t} + 2X_{4t}) + \cdots + \beta_K X_{Kt} + u_t$$
$$Y_t^* = \beta_0 + \beta_1 X_{1t} + \beta_3 Z_t + \cdots + \beta_K X_{Kt} + u_t$$

    where $Y_t^* = Y_t - X_{4t}$ and $Z_t = X_{3t} + 2X_{4t}$.
  - This transformed model:
    - can be estimated by OLS:

$$\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_3, \widehat{\beta}_5, \ldots, \widehat{\beta}_K \text{ , together with } \widehat{\beta}_2 = 0 \text{ and } \widehat{\beta}_4 = 2\widehat{\beta}_3 + 1 \text{ .}$$

    - has new endogenous variable Y_t^* (not always so: e.g. if $\beta_2 = 0$ alone) and new explanatory variable Z_t.

GLRM under linear restrictions (2)

- The "transformation" method is good for simple cases only.
- In general, q (nonredundant) linear restrictions among parameters:

$$\begin{matrix} 1 \\ \vdots \\ q \end{matrix} \begin{pmatrix} \diamond & \diamond & \diamond & \dots & \diamond \\ \vdots & & & & \\ \diamond & \diamond & \diamond & \dots & \diamond \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \dots \\ \beta_K \end{pmatrix} = \begin{pmatrix} \diamond \\ \vdots \\ \diamond \end{pmatrix}$$

  - for given matrix R and vector r,

$$\underset{(q \times K+1)}{R} \quad \beta = \underset{(q \times 1)}{r}$$

  - example of non-valid case (why?):

$$\beta_3 = 0, \quad 2\beta_2 + 3\beta_4 = 1, \quad \beta_1 - 2\beta_4 = 3, \quad 6\beta_4 = 2 - 4\beta_2 + \beta_3$$

GLRM under linear restrictions (2cont)

- Write previous example $\beta_2 = 0$ and $2\beta_3 = \beta_4 - 1$ ($q = 2$ restrictions) as in general formula:

$$\underset{(2\times K+1)}{\underbrace{\begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 2 & -1 & 0 & \dots & 0 \end{pmatrix}}_{R}} \underset{(K+1\times 1)}{\underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \dots \\ \beta_K \end{pmatrix}}_{\beta}} = \underset{(2\times 1)}{\underbrace{\begin{pmatrix} 0 \\ -1 \end{pmatrix}}_{r}}.$$

- In general, we write GLRM subject to q linear restrictions as:

$$\underset{(T\times 1)}{Y} = \underset{(T\times K+1)}{X} \underset{(K+1\times 1)}{\beta} + \underset{(T\times 1)}{u},$$

$$\underset{(q\times K+1)}{R} \underset{(K+1\times 1)}{\beta} = \underset{(q\times 1)}{r}.$$

Estimation: restricted least squares (RLS).

- Typical optimization exercise:

$$\min_{\beta}(u'u) \quad \text{where} \quad u = Y - X\beta,$$

$$\text{subject to } R\beta = r.$$

- Lagrangian:

$$L(\beta,\lambda) = u'u - 2\lambda'(R\beta - r)$$

$$\min_{\beta,\lambda} L(\beta,\lambda).$$

- First derivatives:

$$\frac{\partial L(\beta,\lambda)}{\partial \beta} = -2X'u - 2R'\lambda,$$

$$\frac{\partial L(\beta,\lambda)}{\partial \lambda} = -2(R\beta - r),$$

Estimation: restricted least squares (RLS) (cont).

- 1st.o.c. $\rightsquigarrow$ normal equations:

$$X'\widehat{u}_R + R'\widehat{\lambda} = 0, \tag{4}$$

$$R\widehat{\beta}_R = r, \tag{5}$$

where $\widehat{\beta}_R$ and $\widehat{\lambda}$ are values of β, λ that satisfy 1st.o.c. and residuals

$$\widehat{u}_R = Y - X\widehat{\beta}_R. \tag{6}$$

- Solving for $\widehat{\beta}_R$ and $\widehat{\lambda}$:

$$\begin{aligned}
\widehat{\lambda} &= [R(X'X)^{-1}R']^{-1}(r - R\widehat{\beta}), \\
\widehat{\beta}_R &= \widehat{\beta} + (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(r - R\widehat{\beta}) \\
&= \widehat{\beta} + A(r - R\widehat{\beta}) = (I - AR)\widehat{\beta} + Ar
\end{aligned} \tag{7}$$

where $A = (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}$.

RLS estimation: characteristics

- Expression (7): $\widehat{\beta}_R = \widehat{\beta} + A(r - R\widehat{\beta}) \rightsquigarrow$
  - the restricted estimate $\widehat{\beta}_R$ can be obtained as a function of the (not restricted) ordinary estimate: $\widehat{\beta}$
  - $R\widehat{\beta} \simeq r \quad \Rightarrow \quad \widehat{\beta}_R$ (restricted) $\simeq \widehat{\beta}$ (not restricted) .
- Normal equations (4): $X'\widehat{u}_R + R'\widehat{\lambda} = 0 \rightsquigarrow$
  - satisfy the restrictions (obvious).
  - $X'\widehat{u}_R \neq 0$, *i.e.*:
    - sum of restricted residuals not zero,
    - restricted residuals not orthogonal to explanatory variables,
    - then, restricted residuals not orthogonal to fitted $\widehat{Y}_R$.
  - $\text{TSS} \neq \text{RSS}_R + \text{ESS}_R$
    (compare with ordinary case and with transformed equation: R^2 ??).

Properties of the RLS estimator (1)

Expression (7) : $\widehat{\beta}_R = (I - AR)\widehat{\beta} + Ar \quad \rightsquigarrow$

1. **Linear:** RLS estimator $\widehat{\beta}_R$ is l.c. of OLS estimator $\widehat{\beta}$, which is linear , then $\widehat{\beta}_R$ is linear also .

2. **Bias:** RLS estimator $\widehat{\beta}_R$ is $\begin{cases} \text{biased,} & \text{if } R\beta \neq r \text{ ,} \\ \text{unbiased,} & \text{if } R\beta = r \text{ true} \end{cases}$

   Demo:

$$\mathrm{E}\left(\widehat{\beta}_R\right) = (I - AR)\,\mathrm{E}\left(\widehat{\beta}\right) + Ar = (I - AR)\beta + Ar = \beta + A(r - R\beta).$$

3. **Covariance Matrix:** $\mathrm{Var}\left(\widehat{\beta}_R\right) = (I - AR)\mathrm{Var}\left(\widehat{\beta}\right) = \sigma^2 (I - AR)(X'X)^{-1}$

   Demo:

$$\begin{aligned} \mathrm{Var}\left(\widehat{\beta}_R\right) &= (I - AR)\,\mathrm{Var}\left(\widehat{\beta}\right)(I - AR)' = \sigma^2 (I - AR)(X'X)^{-1}(I - AR)' \\ &= \sigma^2 \left[(X'X)^{-1} + AR(X'X)^{-1}R'A' - AR(X'X)^{-1} - (X'X)^{-1}R'A'\right] \end{aligned}$$

where: $AR(X'X)^{-1}R'A' = (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R(X'X)^{-1}R'A'$

$$= (X'X)^{-1}R'A'.$$

Properties of the RLS estimator (2)

4. **Smaller variance** than OLS estimators,
   even if restrictions aren't true:

   Demo:

$$\text{Var}\left(\widehat{\beta}_R\right) = \text{Var}\left(\widehat{\beta}\right) - AR\,\text{Var}\left(\widehat{\beta}\right)$$
$$= \text{Var}\left(\widehat{\beta}\right) - \text{(psd matrix)}.$$

□

5. surprising result (apparently):
   - less "uncertainty" about parameters
     ⇝ greater precision in estimation. . .
   - but. . . towards an erroneous result (biased)
     if restriction isn't true.

Multicolinearity vs restrictions

Must clearly distinguish two different cases:

- linear relationships among regressors
  (*i.e.* multicollinearity):

  e.g. $X_{4t} = 2X_{1t}$

  $\Rightarrow$ missing information for individual estimates.

- linear relationships among coefficients:

  e.g. $\beta_4 = 2\,\beta_1$

  $\Rightarrow$ extra information about parameters

  $\rightsquigarrow$ estimators with smaller variance.

- respective models to estimate:

$$Y_t = \beta_0 + (\underbrace{\beta_1 + 2\beta_4}_{\beta_1^*})X_{1t} + \beta_2 X_{2t} + + \cdots + u_t\,,$$

$\Rightarrow \widehat{\beta}_1^*$ but $\widehat{\beta}_1, \widehat{\beta}_4$?

$$Y_t = \beta_0 + \beta_1(\underbrace{X_{1t} + 2X_{4t}}_{X_{1t}^*}) + \beta_2 X_{2t} + + \cdots + u_t\,,$$

$\Rightarrow \widehat{\beta}_1$ and $\widehat{\beta}_4 = 2\,\widehat{\beta}_1$

3 The Linear Regression Model (II). Inference and Prediction.

3.1a Distribution of the Least-Squares Estimator under the Normality assumption.

OLS estimator under Normality

- If $Y = X\beta + u$, where $\boxed{u \sim \mathcal{N}(0, \sigma^2 I_T)}$, then (recall) OLS estimator:

$$\widehat{\beta}_{\text{OLS}} = (X'X)^{-1}X'Y = \beta + (X'X)^{-1}X'u$$
$$= \beta + \Gamma' u \quad \text{is linear in disturbances.}$$

- Therefore, same Multivariate Normal distribution, with (recall)

$$\begin{cases} \text{E}(\widehat{\beta}) & = \beta, \\ \text{Var}(\widehat{\beta}) & = \sigma^2 (X'X)^{-1}. \end{cases}$$

- That is:

$$\boxed{\widehat{\beta} \sim \mathcal{N}(\beta, \sigma^2 (X'X)^{-1})}$$

OLS estimator under Normality (cases)

Since $\boxed{\widehat{\beta} \sim \mathcal{N}(\beta, \sigma^2 (X'X)^{-1})}$:

- For the k-th coefficient:

$$\widehat{\beta}_k \sim \mathcal{N}(\beta_k, \sigma^2 a_{kk})$$

where a_{kk} is the $(k+1)$-th diagonal element of $(X'X)^{-1}$

- *for example:* $\widehat{\beta}_1 \sim \mathcal{N}(\beta_1, \sigma^2 a_{11})$,

$a_{11} =$ 2nd diagonal element.

- For a set of linear combinations:

$$R\widehat{\beta} \sim \mathcal{N}(R\beta, \sigma^2 R(X'X)^{-1}R').$$

- For a subvector of $\widehat{\beta}$: $R = [0_s \ldots 0_s | I_s]$; then

$$\widehat{\beta}^s \sim \mathcal{N}(\beta^s, \sigma^2 A_{ss})$$

where $\beta^s =$ subvector of β, $A_{ss} =$ submatrix of $(X'X)^{-1}$.

OLS estimator under Normality (cases)2

- In particular, if $R = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \Rightarrow$

  $R \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \beta^{\star}$ (without intercept):

- and

$$(X'X)^{-1} = \begin{pmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \\ a_{20} & a_{21} & a_{22} \end{pmatrix};$$

- then

$$\widehat{\beta}^{\star} \sim \mathcal{N}(\beta^{\star}, \sigma^2 \diamond)$$

OLS residuals under Normality

- Similarly, if $u \sim \mathcal{N}(0, \sigma^2 I_T)$,

  Then,

$$\widehat{u} \sim \mathcal{N}(0, \sigma^2 M)$$

- In particular, for the 4-th residual:

$$\widehat{u}_t \sim \mathcal{N}(0, \sigma^2 m_{44})$$

  where m_{44} is the 4-th diagonal element of matrix M.

3.1b Hypothesis Testing: a Review.

Hypothesis and Tests (rev1)

- Starting point:

$$\left.\begin{array}{ll} Y & = X\beta + u \\ u & \sim \mathcal{N}(0, \sigma^2 I_T) \end{array}\right\} \left\{ \begin{array}{l} \widehat{\beta} \sim \mathcal{N}(\beta, \sigma^2 (X'X)^{-1}) \\ \widehat{u} \sim \mathcal{N}(0, \sigma^2 M) \end{array}\right.$$

- Hypothesis: "conjecture about parameter(s) dn fn".
  For example:
  - in SLRM: $\widehat{\beta} \sim \mathcal{N}(\beta, v)$; assume $\beta = 2.5$.
  - in GLRM: $\widehat{\beta} \sim \mathcal{N}(\beta, \sigma^2 (X'X)^{-1})$; assume $\beta_1 + \cdots + \beta_K = 1$.
  - in general: Ec. Th. $\rightsquigarrow$ hypothesis
    e.g.: Cobb-Couglas Fn:

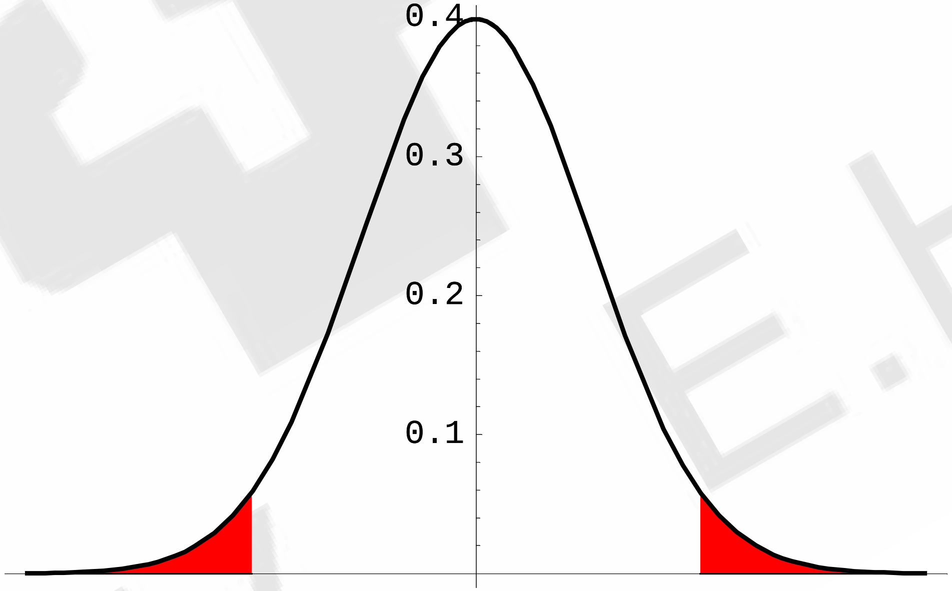
    $$Y_t = e^{\beta_0} L_t^{\beta_1} K_t^{\beta_2} e^{u_t}$$

    with Constant returns to scale: $\beta_1 + \beta_2 = 1$
- Test: "procedure to reject or accept the hypothesis"

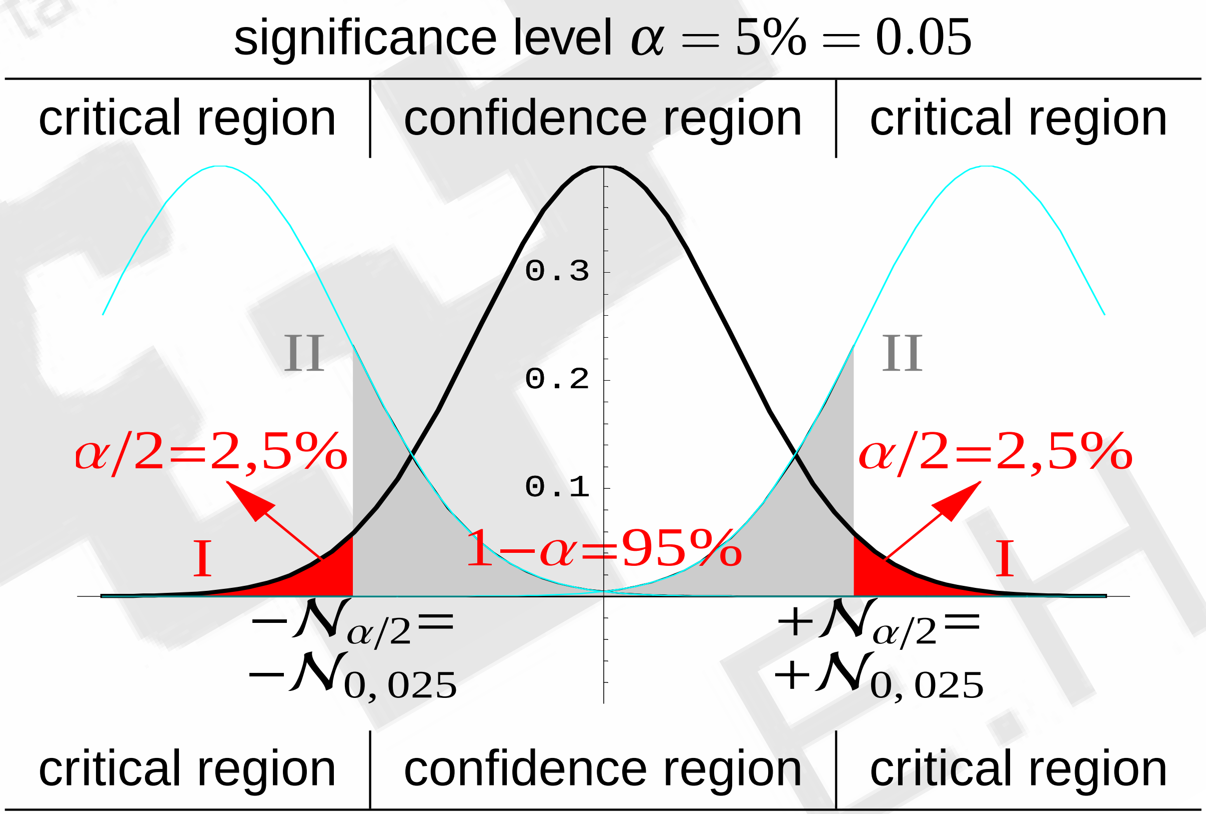
Hypothesis and Tests (rev2)

	elements	steps
a)	hypothesis to test (about estimator)	$H_0 : \ldots$ vs. $H_a : \ldots$ (disjoint)
b)	estimator dn	obtain test statistic with tabulated dn under H_0 :
c)	decision rule	calculated statistic: $\in$ critical region ("large") $\Downarrow$ Reject; $\notin$ critical region ("small") $\Downarrow$ not Reject

Example:

a)	$H_0 : \beta = 2.5$ vs. $H_a : \beta \neq 2.5$ (Var(β)=4)
b)	$\widehat{\beta} \sim \mathcal{N}(\beta, 4) \rightsquigarrow z = \frac{\widehat{\beta} - \beta}{2} \sim \mathcal{N}(0,1)$
c)	$z = \frac{\widehat{\beta} - 2.5}{2} \in$ 

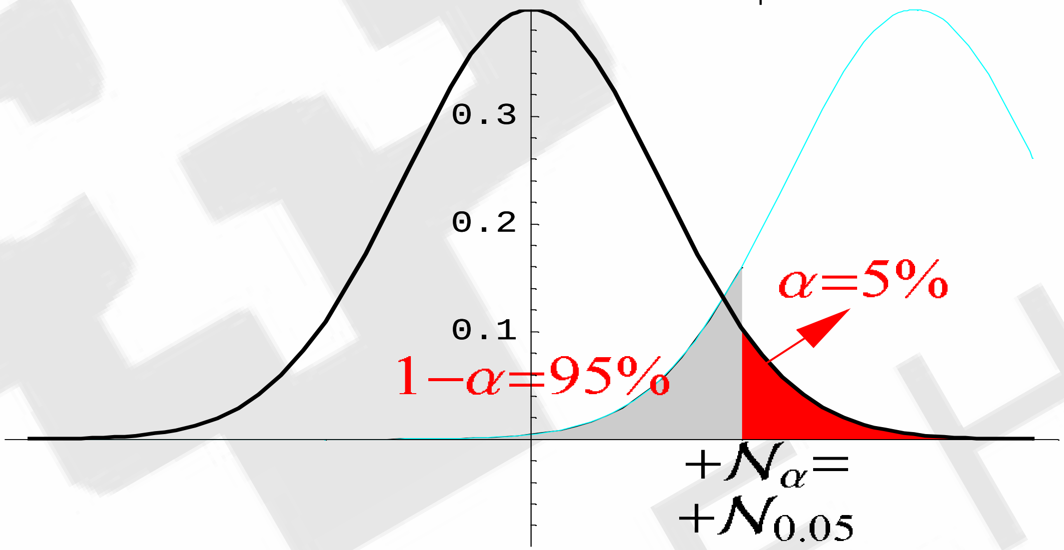
Hypothesis and Tests: Critical region



Hypothesis and Tests: Critical region (one sided)

significance level $\alpha = 5\% = 0.05$

confidence region	critical region



confidence region	critical region

Hypothesis and Tests: Distributions (rev)

1. Def of χ^2 (chi-square):

$$\left.\begin{array}{c} Z_i \sim \text{iid}\,\mathcal{N}(0,1) \\ Z \sim \mathcal{N}(0, I_m) \end{array}\right\} Z'Z = \sum_{i=1}^{m} Z_i^2 \sim \chi^2(m) \qquad \left\{\begin{array}{l} \mathrm{E}(\chi^2(m)) = m \\ \mathrm{Var}(\chi^2(m)) = 2m \end{array}\right.$$

1b. $Z \sim \mathcal{N}(\mu, \Omega) \Rightarrow (Z-\mu)'\Omega^{-1}(Z-\mu) \sim \chi^2(m)$

2. Def of t (Student): $\left.\begin{array}{c} Z \sim \mathcal{N}(0,1), \quad W \sim \chi^2(m) \\ Z, W \text{ independent} \end{array}\right\} \dfrac{Z}{\sqrt{W/m}} \sim t(m)$

3. Def of $\mathscr{F}$ (Snedecor): $\left.\begin{array}{c} V \sim \chi^2(n) \quad W \sim \chi^2(m) \\ V, W \text{ independent} \end{array}\right\} \dfrac{V/n}{W/m} \sim \mathscr{F}_m^n$

4b. $n = 1 \Rightarrow \dfrac{Z^2}{W/m} \sim \mathscr{F}_m^1 \equiv t(m)^2$

Hypothesis and Tests: Useful result

From $\widehat{u} \sim \mathcal{N}(0, \sigma^2 M)$:

- $\dfrac{\text{RSS}}{\sigma^2} = \sum(\widehat{u}_t^2/\sigma^2) = \sum \mathcal{N}(0,1)^2\text{'s} \sim \chi^2(T{-}K{-}1)$
- Then: $\dfrac{\widehat{\sigma}^2}{\sigma^2} = \dfrac{\text{RSS}}{\sigma^2(T{-}K{-}1)} = \dfrac{\text{RSS}}{\sigma^2(T{-}K{-}1)} = \chi^2/\text{d.f.'s}$
  - $\dfrac{\text{expr}}{\sigma} \sim \mathcal{N}(0,1)$:
  - $\dfrac{\text{expr}}{\widehat{\sigma}} = \dfrac{\text{expr}/\sigma}{\widehat{\sigma}/\sigma} = \dfrac{\text{expr}/\sigma}{\sqrt{\widehat{\sigma}^2/\sigma^2}} = \dfrac{\mathcal{N}(0,1)}{\sqrt{\chi^2/\text{d.f.'s}}} = t$
  - $\dfrac{\text{expr}}{\sigma^2} \sim \chi^2(n)$:
  - $\dfrac{\text{expr}}{\widehat{\sigma}^2} = \dfrac{\text{expr}/\sigma^2}{\widehat{\sigma}^2/\sigma^2} \Rightarrow \dfrac{\frac{\text{expr}}{\sigma^2}/n}{\widehat{\sigma}^2/\sigma^2} = \dfrac{\chi^2(n)/n}{\chi^2/\text{d.f.'s}} \sim \mathcal{F}$
- In short: $\sigma^2 \to \widehat{\sigma}^2 \Rightarrow$ $\mathcal{N}(0,1) \to t$!!
  $\chi^2 \to \mathcal{F}$!!

3.2a Testing for the Significance of a single parameter. Confidence Intervals.

Single parameter Significance test: estimator dn

- Standardise $\widehat{\beta}_i \sim \mathcal{N}(\beta_i, \sigma^2 a_{ii})$

$$\frac{\widehat{\beta}_i - \beta_i}{\sqrt{\text{Var}(\widehat{\beta}_i)}} = \frac{\widehat{\beta}_i - \beta_i}{\sigma\sqrt{a_{ii}}} = \frac{\widehat{\beta}_i - \beta_i}{\sigma_{\widehat{\beta}_i}} \sim \mathcal{N}(0,1)$$

- change σ by $\widehat{\sigma}$:

$$\frac{\widehat{\beta}_i - \beta_i}{\widehat{\sigma}\sqrt{a_{ii}}} = \frac{\widehat{\beta}_i - \beta_i}{\sqrt{\widehat{\text{Var}(\widehat{\beta}_i)}}} = \boxed{\frac{\widehat{\beta}_i - \beta_i}{S_{\widehat{\beta}i}} \sim t\,(T-K-1)}$$

- Note how $\sigma_{\widehat{\beta}i} \rightarrow S_{\widehat{\beta}i} \quad \Rightarrow \quad \mathcal{N}(0,1) \rightarrow t$!!

Single parameter Significance test: rule

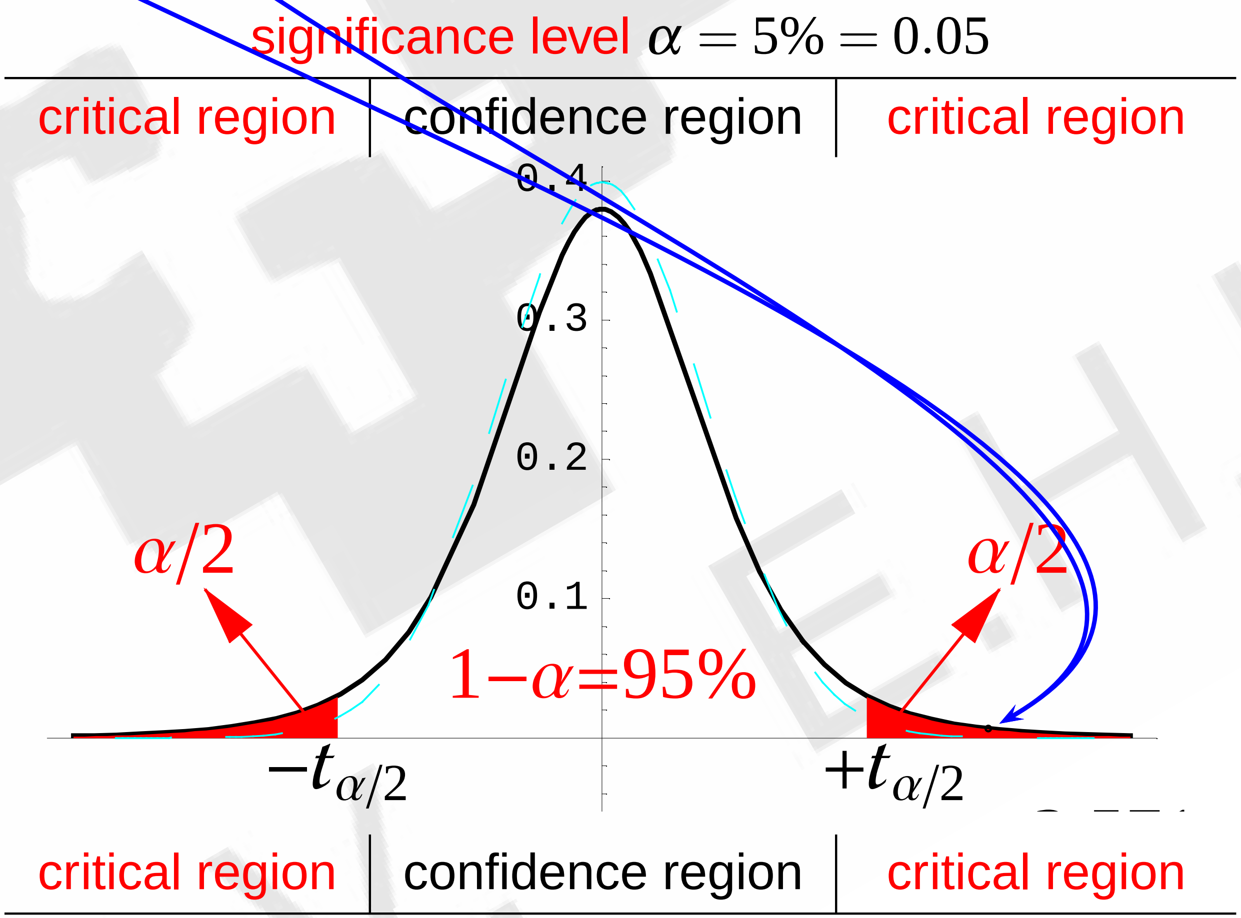
- $$\frac{\widehat{\beta}_i - \beta_i}{S_{\widehat{\beta}_i}} \sim t_{(T-K-1)}$$
- Which Test? $\begin{cases} H_0 : \beta_i = c & \text{(informative test)} \\ H_0 : \beta_i = 0 & \text{(test of significance)} \end{cases}$
- **Remember:** Hypothesis $\rightsquigarrow$ statistic $\rightsquigarrow$ rule...
- Test of Significance:
  - Hypothesis: $H_0 : \beta_i = 0$ vs. $H_a : \beta_i \neq 0$
  - Statistic: $t = \dfrac{\widehat{\beta}_i}{S_{\widehat{\beta}_i}} \sim t_{(T-K-1)}$ under H_0 :
  - Rule: $|t| = |\dfrac{\widehat{\beta}_i}{S_{\widehat{\beta}_i}}| > t_{\alpha/2(T-K-1)} \quad \Rightarrow$ reject H_0 :

    $\Rightarrow$ β_i is (statistically or significantly) different from zero

    $\Rightarrow$ X_i is a (statistically) relevant or significant variable.
- similarly for informative test $H_0 : \beta_i = c$ (Exercise: Try it!!)

Single parameter Significance test: rule (cont)

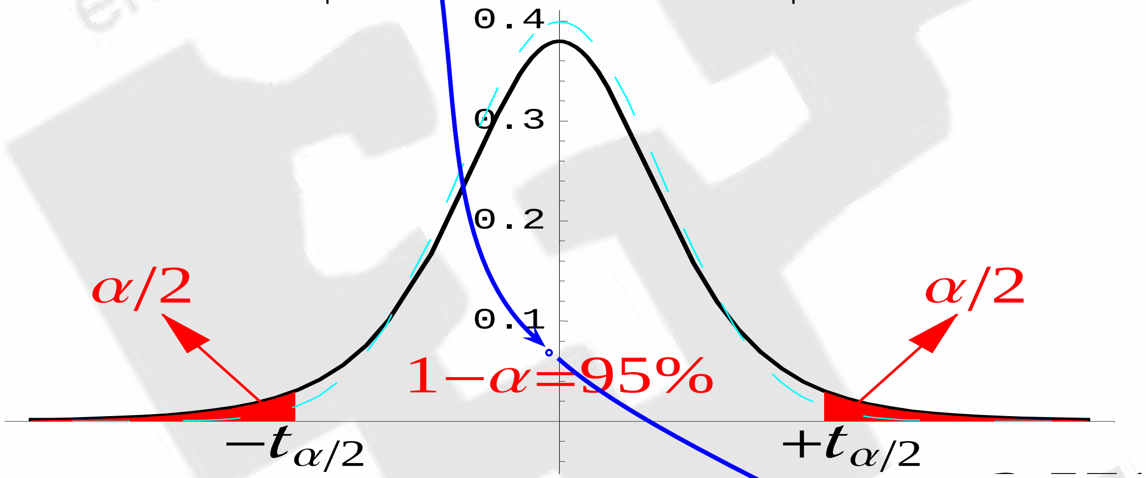
- Rule: $|t| = |\frac{\widehat{\beta}_i}{S_{\widehat{\beta}_i}}| > t_{\alpha/2}(T{-}K{-}1) \quad \Rightarrow \quad$ reject H_0 :
-



Confidence interval for β_i

- Recall that $\dfrac{\widehat{\beta}_i - \beta_i}{S_{\widehat{\beta}_i}} \sim t_{(T-K-1)}$

critical region | confidence region | critical region



critical region | confidence region | critical region

i.e.: $\Pr[-t_{\alpha/2} \leq \dfrac{\widehat{\beta}_i - \beta_i}{S_{\widehat{\beta}_i}} \leq +t_{\alpha/2}] = 1-\alpha$

$$\Pr[\underbrace{\widehat{\beta}_i - t_{\alpha/2}\, S_{\widehat{\beta}_i} \leq \beta_i \leq \widehat{\beta}_i + t_{\alpha/2}\, S_{\widehat{\beta}_i}}_{\text{CI}_{1-\alpha}(\beta_i)}] = 1-\alpha$$

Confidence interval for β_i (cont)

- That is:

$$\text{CI}_{1-\alpha}(\beta_i) = [\widehat{\beta}_i \pm t_{\alpha/2}\, S_{\widehat{\beta}_i}]$$

- *e.g.* for $\alpha = 5\%$, $T–K–1 = 25$, $\widehat{\beta}_i = 2.12$ and $S_{\widehat{\beta}_i} = 0.08$:

$$\text{CI}_{95\%}(\beta_i) = [\widehat{\beta}_i \pm t_{2.5\%}(25)\, S_{\widehat{\beta}_i}]$$

$$= [\widehat{\beta}_i \pm 2.06\, S_{\widehat{\beta}_i}] = [2.12 \pm 2.06 \cdot 0.08] = [1.9552\,;\, 2.2848]$$

- testing by means of confidence interval:

Hypothesis:	$H_0 : \beta_i = c$ vs. $H_a : \beta_i \neq c$
Interval:	$\text{CI}_{95\%}(\beta_i)$
Rule:	Reject H_0 : if $c \notin \text{CI}_{95\%}(\beta_i)$, with 5% significance.

- *e.g.* $H_0 : \beta_i = 0$? $\Rightarrow$ Reject $\Rightarrow$ β_i is significant (at 5% level).

Testing a Single Linear Combination

- Let's have a restricted GLRM with 1 restriction ($q = 1$):
  $R\beta = r$ but now simpler...
  $R = d'$ (any row of $K+1$ values $d_0, d_1, \ldots, +d_K$) and
  $r = c$ (any single value):
- Let $H_0 : v = d'\beta = d_0\beta_0 + d_1\beta_1 + \cdots + d_K\beta_K = c$
  that is,
  an informative test about the value c that takes a single linear combination v of the parameters.

Testing a Single Linear Combination: Example

- Let's have the linearised Cobb-Douglas fn

$$\log Y_t = \alpha + \beta_L \log L_t + \beta_K \log K_t + u_t$$

$d' = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix}$ and $c = 1$:

-

$$H_0 : v = d'\beta = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix} \begin{pmatrix} \alpha \\ \beta_L \\ \beta_K \end{pmatrix} = \beta_L + \beta_K$$

$$= c \qquad = 1$$

that is, $H_0 : \beta_L + \beta_K = 1$;

the test of the constant returns to scale hypothesis.

Testing a Single Linear Combination: dn

- Since $\widehat{\beta} \sim \mathcal{N}(\beta, \sigma^2 (X'X)^{-1})$, we have that

$$d'\widehat{\beta} \sim \mathcal{N}(d'\beta, \sigma^2 d'(X'X)^{-1} d)$$

$$\widehat{\nu} \sim \mathcal{N}(\nu, \text{Var}(\widehat{\nu}))$$

where $\text{Var}(\widehat{\nu}) = \sigma^2 \sum_{i,j=0}^{K} d_i d_j a_{ij}$

- As before, standardise $\widehat{\nu}$

$$\frac{\widehat{\nu} - \nu}{\sqrt{\text{Var}(\widehat{\nu})}} \sim \mathcal{N}(0,1)$$

- Therefore (recall $\sigma \rightarrow \widehat{\sigma}$):

$$\Rightarrow \boxed{\frac{\widehat{\nu} - \nu}{S_{\widehat{\nu}}} \sim t_{(T-K-1)}}$$

where $S_{\widehat{\nu}} = \widehat{\sigma} \sqrt{\sum_{i,j=0}^{K} d_i d_j a_{ij}}$.

Testing a Single Linear Combination: rule

- $$\boxed{\frac{\widehat{v} - v}{S_{\widehat{v}}} \sim t_{(T-K-1)}}$$
- Which Test? $\left\{ H_0 : v (= d'\beta) = c \right.$ (informative test)
- **Remember:** Hypothesis $\rightsquigarrow$ statistic $\rightsquigarrow$ rule. . .
- Test for a linear combination:
  - Hypothesis: $H_0 : v = c$ vs. $H_a : v \neq c$
  - Statistic:

$$t = \frac{\widehat{v} - c}{S_{\widehat{v}}} \sim t_{(T-K-1)} \text{ under } H_0 :$$

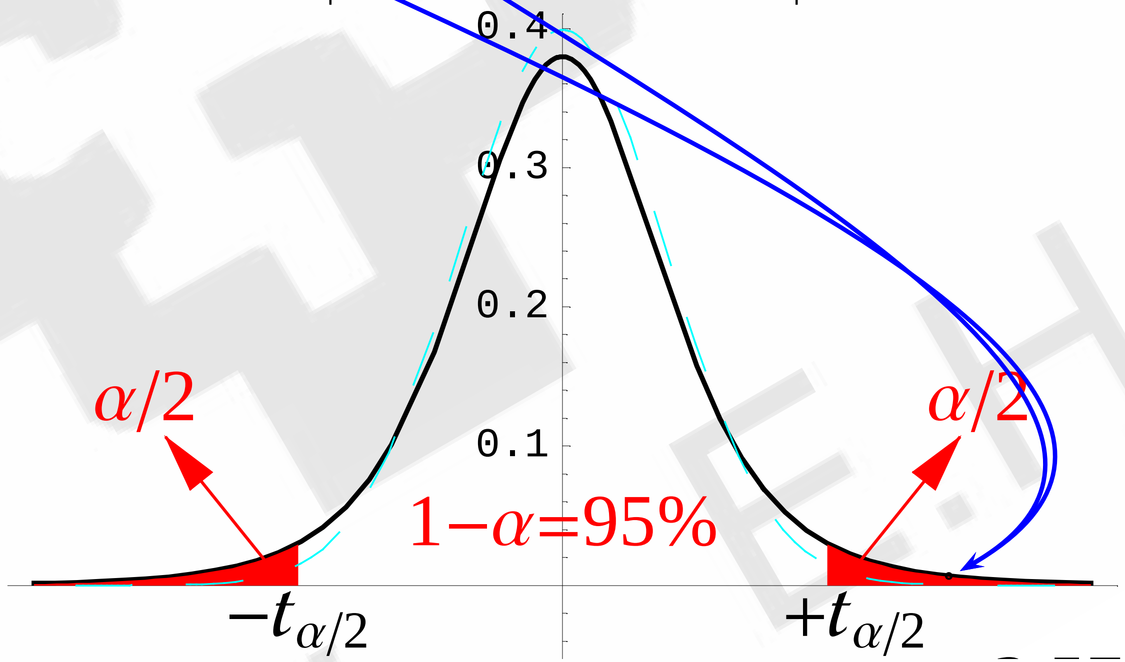
  - Rule: $|t| > t_{\alpha/2}(T-K-1) \quad \Rightarrow \quad$ reject H_0 :
    $\Rightarrow$ value of linear combination isn't right.
  - *cf* test of single parameter β_k, any similarities?.

Testing a Single Linear Combination: rule (cont)

- Rule: $|t| = \left| \frac{\widehat{\nu} - c}{S_{\widehat{\nu}}} \right| > t_{\alpha/2}(T-K-1) \quad \Rightarrow \quad$ reject H_0 :
-

significance level $\alpha = 5\% = 0.05$

critical region	confidence region	critical region



critical region	confidence region	critical region

Testing a Single Linear Combination: Example

- In the linearised Cobb-Douglas fn:
  $\widehat{\log Y}_t = \widehat{\alpha} + \widehat{\beta}_L \log L_t + \widehat{\beta}_K \log K_t, \quad T = 53;$
- $\widehat{\log Y}_t = 2.10 + 0.67 \log L_t + 0.32 \log K_t, \quad \widehat{\sigma}^2 = 4$;

$$(X'X)^{-1} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & -1 \\ 0 & -1 & 7 \end{pmatrix}$$

- Test the H_0 : constant returns to scale

  at the $\alpha = 5\%$ significance level:

Testing a Single Linear Combination: Example (cont)

- Hypothesis: $H_0 : \beta_L + \beta_K = 1$ vs. $H_a : \beta_L + \beta_K \neq 1$
- Statistic:
  - $$\widehat{\nu} = \widehat{\beta}_L + \widehat{\beta}_K = 0.67 + 0.27 = 0.89$$
  - $$S_{\widehat{\nu}} = \sqrt{\widehat{\text{Var}(\widehat{\beta}_L)} + \widehat{\text{Var}(\widehat{\beta}_K)} + 2\widehat{\text{Cov}(\widehat{\beta}_L, \widehat{\beta}_K)}} = \widehat{\sigma}\sqrt{a_{11} + a_{22} + 2a_{12}} = 2\sqrt{4 + 7 + 2(-1)} = 2\sqrt{9} = 6$$
  - $$t = \frac{\widehat{\nu} - 1}{S_{\widehat{\nu}}} = \frac{0.89 - 1}{6} = \frac{-0.11}{6} = -0.018.$$
- Rule: $|t| = 0.018 < t_{0.025}(50) = 2.01 \Rightarrow$ don't reject H_0 :

$\Rightarrow$ “constant returns to scale" is supported by data.

3.2b Testing for Overall Significance.

Overall Significance Test: estimator dn

- $H_0 : \beta_1 = \beta_2 = \cdots = \beta_K = 0 \quad \rightsquigarrow$
- $H_0 : \beta^\star = \mathbf{0} \quad \rightsquigarrow$
- $$\widehat{\beta}^\star \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \begin{bmatrix} a_{00} & a_{01} & \dots & a_{0K} \\ a_{10} & a_{11} & \dots & a_{1K} \\ \vdots & \vdots & \dots & \vdots \\ a_{K0} & a_{K1} & \dots & a_{KK} \end{bmatrix}\right) \sim \mathcal{N}(\mathbf{0}, \sigma^2 (x'x)^{-1})$$
- Standardise and write Sum of Squares:
  - $$\frac{\widehat{\beta}^{\star\prime} x' x \widehat{\beta}^\star}{\sigma^2} \sim \chi^2(K) \quad \text{under } H_0 :$$
- Therefore (recall changing $\sigma^2 \rightarrow \widehat{\sigma}^2$):

$$\boxed{F = \frac{\widehat{\beta}^{\star\prime} x' x \widehat{\beta}^\star / K}{\widehat{\sigma}^2} \sim \mathcal{F}^K_{T-K-1}}$$

Overall Significance Test: rule

- $$F = \frac{\widehat{\beta}^{\star\prime} x' x \widehat{\beta}^{\star} / K}{\widehat{\sigma}^2} \sim \mathscr{F}^{K}_{T-K-1}$$ under H_0 :

- Overall significance test: $\left\{ H_0 : \beta^\star = 0 \right.$

- **Remember:** Hypothesis $\rightsquigarrow$ statistic $\rightsquigarrow$ rule. . .
  - Hypothesis: $H_0 : \beta^\star = 0$ vs. $H_a : \beta^\star \neq 0$ (*i.e.* $\exists \beta_i \neq 0$)
  - Statistic:

$$F = \frac{\widehat{\beta}^{\star\prime} x' x \widehat{\beta}^{\star} / K}{\widehat{\sigma}^2} = \frac{\widehat{y}'\widehat{y}/K}{\widehat{u}'\widehat{u}/(T-K-1)} = \frac{\text{ESS}/K}{\text{RSS}/(T-K-1)}$$

$$= \frac{(\text{ESS}/\text{TSS})/K}{(\text{RSS}/\text{TSS})/(T-K-1)} = \frac{R^2/K}{(1-R^2)/(T-K-1)} \sim \mathscr{F}^{K}_{T-K-1} \text{ under } H_0 :$$

  - Rule: $F > \mathscr{F}_{\alpha}(K, T-K-1) \quad \Rightarrow \quad$ reject H_0 :

    $\Rightarrow$ all coefs are jointly significant (different from zero)

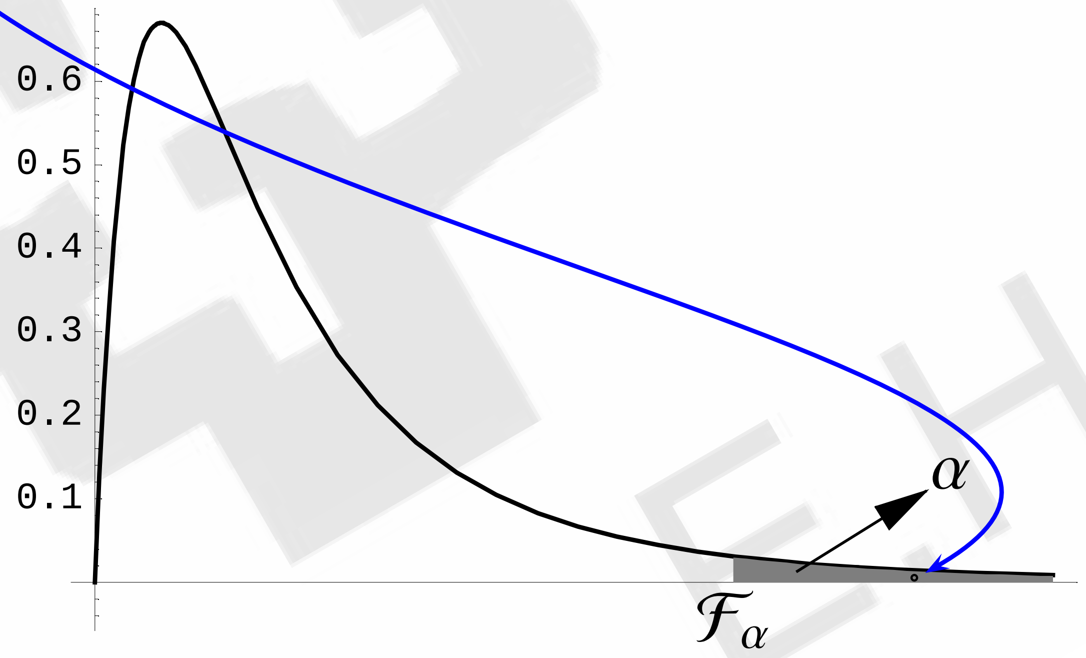
    $\Rightarrow$ whole regression is (statistically) relevant.

Overall Significance Test: rule (cont)

- Rule: $F > \mathscr{F}_{\alpha}(K, T{-}K{-}1) \quad \Rightarrow \quad$ reject H_0 :
-

significance level $\alpha = 5\% = 0.05$

confidence region	critical region



confidence region	critical region

Overall Significance Test: Example

- In the previous example (linearised Cobb-Douglas fn:)

$$\widehat{\log Y}_t = \widehat{\alpha} + \widehat{\beta}_L \log L_t + \widehat{\beta}_K \log K_t, \quad T = 53;$$

$$\widehat{\log Y}_t = 2.10 + 0.67 \log L_t + 0.32 \log K_t, \quad \widehat{\sigma}^2 = 4; R^2 = 0.88$$

- Test the overall significance

at the $\alpha = 5\%$ significance level:

-

$$F = \frac{R^2 / K}{(1 - R^2) / (T - K - 1)}$$

$$= \frac{0.88 / 2}{(1 - 0.88) / (50)} = \frac{0.44}{0.024} = 183.33 > \mathscr{F}_{0.05}(2, 50) = 3.19$$

$\Rightarrow$ β_K & β_L are jointly significant
$\Rightarrow$ regression is (statistically) relevant.

3.3 A General Test for Linear Restrictions.

Testing for Linear Restrictions: Example 1

- Recall GLRM subject to q linear restrictions:

$$\underset{(T\times 1)}{Y} = \underset{(T\times K+1)}{X}\;\underset{(K+1\times 1)}{\beta} + \underset{(T\times 1)}{u},$$
$$H_0:\ \underset{(q\times K+1)}{R}\;\underset{(K+1\times 1)}{\beta} = \underset{(q\times 1)}{r}.$$

- Previous tests $\equiv$ special cases of LRs:
  1. Let's have the GLRM with
     $q=1$, $R = \begin{bmatrix} 0 & 0 & 1 & \ldots & 0 \end{bmatrix}$ and $r = 0$:

$$H_0: R\beta = \begin{bmatrix} 0 & 0 & 1 & \ldots & 0 \end{bmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \ldots \\ \beta_K \end{pmatrix} = \beta_2$$
$$= r \qquad = 0$$

es decir, $H_0: \beta_2 = 0$;

the test of individual significance of X_2 .

Testing for Linear Restrictions: Example 2

- $H_0: \underset{(q \times K+1)}{R} \; \underset{(K+1 \times 1)}{\beta} = \underset{(q \times 1)}{r}$.
  3. Let's assume $q = 2$ restrictions such that

$$R = \begin{bmatrix} 0 & 2 & 3 & 0 & \dots & 0 \\ 1 & 0 & 0 & -2 & \dots & 0 \end{bmatrix} \text{ and } r = \begin{bmatrix} 5 \\ 3 \end{bmatrix}:$$

-

$$H_0 : R\beta = \begin{bmatrix} 0 & 2 & 3 & 0 & \dots & 0 \\ 1 & 0 & 0 & -2 & \dots & 0 \end{bmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \dots \\ \beta_K \end{pmatrix} = \begin{bmatrix} 2\beta_1 + 3\beta_2 \\ \beta_0 - 2\beta_3 \end{bmatrix}$$

$$= r = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$$

that is, the GLRM under $H_0 : \begin{cases} 2\beta_1 + 3\beta_2 = 5 \\ \beta_0 - 2\beta_3 = 3 \end{cases}$

Testing for Linear Restrictions: Example 3

- $H_0: \underset{(q\times K+1)}{R} \; \underset{(K+1\times 1)}{\beta} = \underset{(q\times 1)}{r}$.

2. Let's assume $q = K$ restrictions such that

$$R = \left[\mathbf{0} \;\middle|\; \mathbf{I}_K \right] = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix} \text{ and } r = \mathbf{0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

-

$$H_0 : R\beta = \left[\mathbf{0}_K \;\middle|\; \mathbf{I}_K \right] \begin{pmatrix} \beta_0 \\ \vdots \\ \beta^{\star} \end{pmatrix} = \beta^{\star}$$

$$= r \qquad = \mathbf{0}$$

that is, $H_0 : \beta^{\star} = \mathbf{0}$;

the test of overall significance of the regression.

Testing for Linear Restrictions: dn

- …so, can have a general test statistic to cover for all hypothesis of the form

$$H_0: \underset{(q \times K+1)}{R} \underset{(K+1 \times 1)}{\beta} = \underset{(q \times 1)}{r} \quad ?$$

- Given that $\widehat{\beta} \sim \mathcal{N}(\beta, \sigma^2 (X'X)^{-1})$, we have that

$$R\widehat{\beta} \sim \mathcal{N}(R\beta, \sigma^2 R(X'X)^{-1}R')$$

- As before, standardise $R\widehat{\beta}$ and construct SS,

$$\frac{(R\widehat{\beta} - R\beta)'[R(X'X)^{-1}R']^{-1}(R\widehat{\beta} - R\beta)}{\sigma^2} \sim \chi^2(q)$$

- Therefore (recall changing $\sigma^2 \to \widehat{\sigma}^2$):

$$\boxed{\frac{(R\widehat{\beta} - R\beta)'[R(X'X)^{-1}R']^{-1}(R\widehat{\beta} - R\beta)/q}{\widehat{\sigma}^2} \sim \mathcal{F}^{q}_{T-K-1}}$$

General Test for Linear Restrictions: rule

- Which Test? $\left\{ H_0 : R\beta = r \right.$
- **Remember:** Hypothesis $\rightsquigarrow$ statistic $\rightsquigarrow$ rule...
- Test for linear restrictions:
  - Hypothesis: $H_0 : R\beta = r$ vs. $H_a : R\beta \neq r$
  - Statistic:

$$F = \frac{(R\widehat{\beta} - r)'[R(X'X)^{-1}R']^{-1}(R\widehat{\beta} - r)/q}{\widehat{\sigma}^2} \sim \mathscr{F}^{q}_{T\text{–}K\text{–}1} \text{ under } H_0 :$$

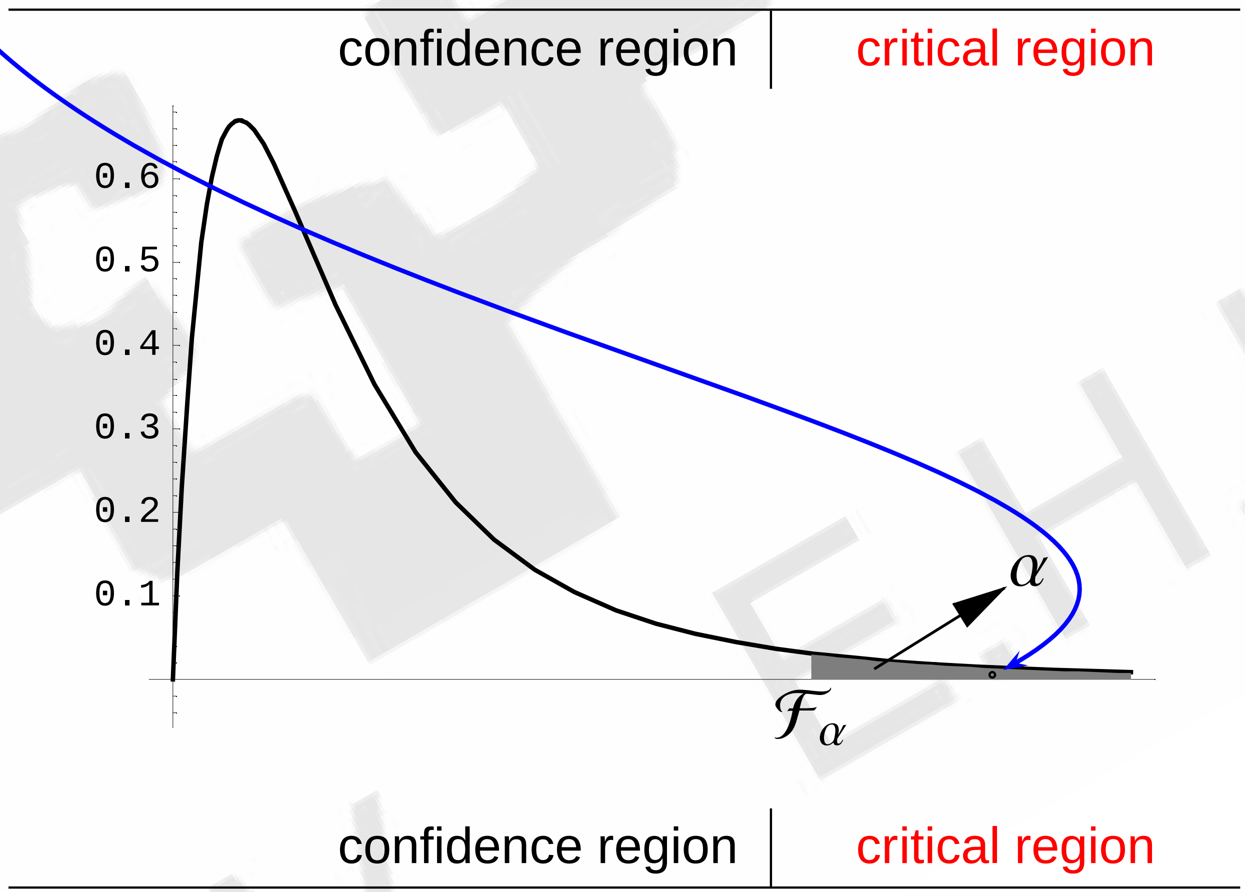
  - Rule: $F > \mathscr{F}_{\alpha}(q, T\text{–}K\text{–}1) \quad \Rightarrow \quad$ reject H_0 :

    $\Rightarrow$ linear restrictions aren't (jointly) true.

General Test for Linear Restrictions: rule (cont)

- Rule: $F > \mathscr{F}_{\alpha}(q, T{-}K{-}1) \quad \Rightarrow \quad$ reject H_0 :
-

significance level $\alpha = 5\% = 0.05$



3.4 Tests based on the Residual Sum of Squares.

General Test for Linear Restrictions: rule 2

- Hypothesis: $H_0 : R\beta = r$ vs. $H_a : R\beta \neq r$
- Statistic:

$$F = \frac{(R\widehat{\beta} - r)'[R(X'X)^{-1}R']^{-1}(R\widehat{\beta} - r)/q}{\widehat{\sigma}^2}$$

- using result on $\widehat{\beta}_R = (I - AR)\widehat{\beta} + Ar$, numerator is difference between SS's:

$$F = \frac{(\text{RSS}_R - \text{RSS})/q}{\text{RSS}/(T–K–1)} \sim \mathscr{F}^{q}_{T–K–1} \text{ under } H_0 :$$

- Rule: $F > \mathscr{F}_{\alpha}(q, T–K–1)$ $\Rightarrow$ reject H_0 :
  $\Rightarrow$ linear restrictions aren't (jointly) true.

General Test for Linear Restrictions: Summary

- Hypothesis: $H_0 : R\beta = r$ vs. $H_a : R\beta \neq r$
- Statistic:

$$F = \frac{(R\widehat{\beta} - r)'[R(X'X)^{-1}R']^{-1}(R\widehat{\beta} - r)/q}{\widehat{\sigma}^2}$$

$$= \frac{(\text{RSS}_R - \text{RSS})/q}{\text{RSS}/(T\text{–}K\text{–}1)} \sim \mathscr{F}^{q}_{T\text{–}K\text{–}1} \text{ under } H_0:$$

- Rule: $F > \mathscr{F}_{\alpha}(q, T\text{–}K\text{–}1)$ $\Rightarrow$ reject H_0 :
  $\Rightarrow$ linear restrictions aren't (jointly) true.
- Note that, SS form needs estimating twice: unrestricted and restricted regressions.
- and, of course, they can also be used to test for individual significance, overall significance, informative restrictions, etc.

Test based on SS: Example Cobb-Douglas

- Hypothesis: $H_0 : \beta_L + \beta_K = 1$ vs. $H_a : \beta_L + \beta_K \neq 1$
- Statistic:
  - $$\widehat{v} = \widehat{\beta}_L + \widehat{\beta}_K = 0.67 + 0.27 = 0.89$$
  - $$S_{\widehat{v}} = \sqrt{\widehat{\text{Var}(\widehat{\beta}_L)} + \widehat{\text{Var}(\widehat{\beta}_K)} + 2\widehat{\text{Cov}(\widehat{\beta}_L, \widehat{\beta}_K)}} = \widehat{\sigma}\sqrt{a_{11} + a_{22} + 2a_{12}} = 2\sqrt{4 + 7 + 2(-1)} = 2\sqrt{9} = 6$$
  - $$t = \frac{\widehat{v} - 1}{S_{\widehat{v}}} = \frac{0.89 - 1}{6} = \frac{-0.11}{6} = -0.018.$$
- Rule: $|t| = 0.018 < t_{0.025}(50) = 2.01$ $\Rightarrow$ don't reject H_0 :

$\Rightarrow$ the “constant returns to scale" hypothesis is supported by data.

Test based on SS: Example Cobb-Douglas (2)

- Alternatively, use SS form to calculate this t ratio:
  unrestricted:
  $$\log Y = \alpha + \beta_L \log L + \beta_K \log K + u, \quad \rightsquigarrow \quad \text{RSS} = 200$$
- restricted:
  $$\log Y = \alpha + \beta_L \log L + (1 - \beta_L) \log K + u$$
  $$\log(Y/K) = \alpha + \beta_L \log(L/K) + u, \quad \rightsquigarrow \quad \text{RSS}_R = 200.001296$$
-
  $$F = \frac{(\text{RSS}_R - \text{RSS})/q}{\text{RSS}/(T-K-1)}$$
  $$= \frac{(200.001296 - 200)/1}{200/50} = \frac{.001296}{4} = 0.000324$$
  $$< \mathscr{F}_{0.05}(1,50) = 4.04$$
- or (recall $t(m) = \sqrt{\mathscr{F}(1,m)}$)
  $$t = \sqrt{F} = \sqrt{0.000324} = 0.018$$
  $$< t_{0.05}(50) = 2.01$$

General Test: Example 2

- GLRM with $q = 2$, $R = \begin{bmatrix} 0 & 2 & 3 & 0 & \dots & 0 \\ 1 & 0 & 0 & -2 & \dots & 0 \end{bmatrix}$ and $r = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$:

$$R\widehat{\beta} = \begin{bmatrix} d_1'\widehat{\beta} \\ d_2'\widehat{\beta} \end{bmatrix} = \begin{bmatrix} 2\widehat{\beta}_1 + 3\widehat{\beta}_2 \\ \widehat{\beta}_0 - 2\widehat{\beta}_3 \end{bmatrix}$$

-

$$R(X'X)^{-1}R' = \begin{bmatrix} d_1'(X'X)^{-1}d_1 & d_1'(X'X)^{-1}d_2 \\ d_2'(X'X)^{-1}d_1 & d_2'(X'X)^{-1}d_2 \end{bmatrix}$$

$$= \begin{bmatrix} 4a_{11} + 9a_{22} + 12a_{12} & 2a_{10} - 4a_{13} + 3a_{02} - 6a_{23} \\ & a_{00} + 4a_{33} - 4a_{03} \end{bmatrix}$$

- Therefore $F =$

$$\frac{\begin{bmatrix} 2\beta_1 + 3\beta_2 - 5 & \beta_0 - 2\beta_3 - 3 \end{bmatrix} \begin{bmatrix} 4a_{11} + 9a_{22} + 12a_{12} & 2a_{10} - 4a_{13} + 3a_{02} - 6a_{23} \\ & a_{00} + 4a_{33} - 4a_{03} \end{bmatrix}^{-1} \begin{bmatrix} 2\beta_1 + 3\beta_2 - 5 \\ \beta_0 - 2\beta_3 - 3 \end{bmatrix} /2}{\widehat{\sigma}^2}$$

$\sim \mathscr{F}^2_{T-K-1}$ under H_0 :

- *es decir*, an "F" statistic for testing two linear restrictions jointly.

General Test: Example 2

- Alternatively (easier), use SS form to calculate this F statistic:

$$H_0 : \begin{cases} 2\beta_1 + 3\beta_2 = 5 \\ \beta_0 - 2\beta_3 = 3 \end{cases}$$

$$\beta_1 = \frac{5 - 3\beta_2}{2}, \qquad \beta_0 = 3 + 2\beta_3$$

- unrestricted:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 \cdots + u \ \rightsquigarrow \text{RSS}$$

- restricted:

$$Y = (3 + 2\beta_3) + (2.5 - 1.5\beta_2) X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 \cdots + u$$

$$\underbrace{Y - 3 - 2.5X_1}_{Y^*} = \beta_2(\underbrace{X_2 - 1.5X_1}_{X_2^*}) + \beta_3(\underbrace{X_3 + 2}_{X_3^*}) + \beta_4 X_4 \cdots + u$$

$$Y^* = \beta_2 X_2^* + \beta_3 X_3^* + \beta_4 X_4 \cdots + u \ \rightsquigarrow \text{RSS}_R$$

- and $F = \frac{(\text{RSS}_R - \text{RSS})/q}{\text{RSS}/(T-K-1)}$, etc.

General Test: Example 3

- GLRM with $q = K$, $R = \begin{bmatrix} \mathbf{0}_K & | & \mathbf{I}_K \end{bmatrix}$ and $r = \mathbf{0}_K$:

$$R\widehat{\beta} \quad \rightsquigarrow \text{selects} \quad \beta^\star$$

$$R(X'X)^{-1}R' \quad \rightsquigarrow \text{selects} \quad \begin{bmatrix} a_{00} & a_{01} & \dots & a_{0K} \\ a_{10} & a_{11} & \dots & a_{1K} \\ \vdots & \vdots & \dots & \vdots \\ a_{K0} & a_{K1} & \dots & a_{KK} \end{bmatrix} = (x'x)^{-1}$$

- Therefore:

$$F = \frac{(\widehat{\beta}^\star - 0)'[(x'x)^{-1}]^{-1}(\widehat{\beta}^\star - 0)/K}{\widehat{\sigma}^2}$$

$$= \frac{\widehat{\beta}^{\star\prime} x'x\widehat{\beta}^\star/K}{\widehat{\sigma}^2}$$

- *es decir*, the usual "F" statistic for testing the overall significance of the regression.

General Test: Example 3

- Alternatively, use SS form to calculate this F:

$$\text{unrestricted: } Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_K X_K + u \quad \rightsquigarrow \quad \text{RSS}$$

$$\text{restricted: } Y = \beta_0 + u \quad \rightsquigarrow \quad \text{RSS}_R = \text{TSS}$$

- Statistic:

$$F = \frac{(\text{RSS}_R - \text{RSS})/q}{\text{RSS}/(T\text{–}K\text{–}1)} = \frac{(\text{TSS} - \text{RSS})/K}{\text{RSS}/(T\text{–}K\text{–}1)} = \frac{\text{ESS}/K}{\text{RSS}/(T\text{–}K\text{–}1)}$$

$$= \frac{R^2/K}{(1 - R^2)/(T\text{–}K\text{–}1)}$$

obtaining same formula as before.

3.5 Point Prediction and Prediction Interval.

Prediction

- Previous chapters: **Specification, Estimation and Validation**.
- This chapter: Final stage: **Use = Prediction**.
- Starting point: appropriate model to describe behaviour of variable Y:

Model Specification

⇓

Parameter Estimation

⇓

Validation and Hypothesis Testing

-

⇓

not satisfactory result	satisfactory result
⇓	⇓
respecify model	go to prediction stage

Concept

- Time series: prediction (of future values)
  $\Rightarrow$ Forecasting
- Cross-section: prediction (of unobserved values)
  $\Rightarrow$ Simulation
- In general: prediction $\Rightarrow$ answer to
  "what if. . . ?" questions,
  es decir what value would take Y *if* $X = X_p$ *?*

Basic Elements

- Model or PRF:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \cdots + \beta_K X_{Kt} + u_t$$

$$Y_t = X_t'\beta + u_t, \quad t = 1, \ldots, T.$$

- Estimated model or SRF:

$$\widehat{Y}_t = X_t'\widehat{\beta}, \quad t = 1, \ldots, T. \tag{8}$$

- Prediction observation: with subindex $p =$ (usually $p \notin [1, T]$):

$$Y_p = X_p'\beta + u_p.$$

- Random disturbance u_p:

$$\mathsf{E}(u_p) = 0, \quad \mathsf{E}(u_p^2) = \sigma^2, \quad \mathsf{E}(u_p u_s) = 0 \quad \forall s \neq p.$$

- Known value of vector X_p' .

Point Prediction

- Substituting in SRF (8):

$$\boxed{\widehat{Y}_p = X_p' \widehat{\beta}.}$$

es decir, numeric value as approximation to unknown value.

Prediction Error

- The error made (when taking $\widehat{Y}_p$ instead of the true Y_p) is

$$e_p = Y_p - \widehat{Y}_p,$$

- which can be expressed as:
- a function of the two error sources introduced in the prediction.
- Under normality:

$$(\widehat{\beta} - \beta) \sim \mathcal{N}(0, \sigma^2 (X'X)^{-1}), \quad \text{and} \quad u_p \sim \mathcal{N}(0, \sigma^2),$$

- so that

$$\boxed{e_p \sim \mathcal{N}(0, \sigma_e^2)},$$

- where the prediction error variance is:

$$\sigma_e^2 = X_p' \underbrace{\text{Var}\left(\widehat{\beta}\right)}_{\sigma^2 (X'X)^{-1}} X_p + \underbrace{\text{Var}\left(u_p\right)}_{\sigma^2} + \underbrace{\text{Cov}\left(\widehat{\beta}, u_p\right)}_{0}$$

$$= \sigma^2 (1 + X_p' (X'X)^{-1} X_p).$$

Interval Prediction

- Standardised prediction error:

$$\frac{e_p - 0}{\sigma_e} = \frac{e_p}{\sigma\sqrt{1 + X_p'(X'X)^{-1}X_p}} \sim \mathcal{N}(0,1),$$

- Recall how changing $\sigma \to \hat{\sigma} \quad \Rightarrow \quad \mathcal{N}(0,1) \to \boldsymbol{t}$!!, then

$$\frac{e_p}{\hat{\sigma}_e} = \frac{e_p}{\hat{\sigma}\sqrt{1 + X_p'(X'X)^{-1}X_p}} \sim \boldsymbol{t}_{(T-K-1)}.$$

- Therefore:

$$Pr(-\boldsymbol{t}_{\alpha/2} \leq \frac{e_p}{\hat{\sigma}_e} \leq \boldsymbol{t}_{\alpha/2}) = 1 - \alpha,$$

- and solving for Y_p :

$$Pr(\hat{Y}_p - \hat{\sigma}_e \boldsymbol{t}_{\alpha/2} \leq Y_p \leq \hat{Y}_p + \hat{\sigma}_e \boldsymbol{t}_{\alpha/2}) = 1 - \alpha.$$

- Then, the $(1-\alpha)$ confidence interval for the unknown Y_p is:

$$CI(Y_p)_{(1-\alpha)} = \left[\hat{Y}_p \pm \hat{\sigma}_e \boldsymbol{t}_{\alpha/2}\right],$$

which measures the precision of the point prediction.

Prediction: Example

- In the previous example (linearised Cobb-Douglas fn:)

$$\widehat{\log Y_t} = \widehat{\alpha} + \widehat{\beta}_L \log L_t + \widehat{\beta}_K \log K_t, \quad T = 53;$$

$$\widehat{\log Y_t} = 2.10 + 0.67 \log L_t + 0.32 \log K_t, \quad \widehat{\sigma}^2 = 4$$

- "What value would Y_p take if $\log L_p = 2.5; \log K_p = 2.0$?":
- $X_p' = \begin{bmatrix} 1 & 2.5 & 2.0 \end{bmatrix}$
-

$$\widehat{\log Y_p} = X_p' \widehat{\beta} = \begin{bmatrix} 1 & 2.5 & 2.0 \end{bmatrix} \begin{bmatrix} 2.10 \\ 0.67 \\ 0.32 \end{bmatrix}$$

$$= 2.10 + 0.67 \cdot 2.5 + 0.32 \cdot 2.0 = 4.42$$

Prediction: Example

- Construct a 95% CI for the true Y_p:

$$\widehat{\sigma}_e^2 = \sigma^2(1 + X_p'(X'X)^{-1}X_p)$$

$$= 4\left(1 + \begin{bmatrix} 1 & 2.5 & 2.0 \end{bmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & -1 \\ 0 & -1 & 7 \end{pmatrix} \begin{bmatrix} 1 \\ 2.5 \\ 2.0 \end{bmatrix}\right)$$

$$= 4\left(1 + \begin{bmatrix} 2 & 8 & 11.5 \end{bmatrix} \begin{bmatrix} 1 \\ 2.5 \\ 2.0 \end{bmatrix}\right)$$

$$= 4(1 + 45) = 4 \cdot 46 = 184$$

-

$$CI(\log Y_p)_{0.95} = \left[\widehat{\log Y_p} \pm \widehat{\sigma}_e t_{0.025}(50)\right]$$

$$= \left[4.42 \pm \sqrt{184} \cdot 2.01\right]$$

$$= [4.42 \pm 27.25]$$

$$= [-22.84 \quad ; \quad 31.68]$$

4.1 Dummy Variables. Definition and use in the GLRM.

Dummy Variables: Definition

- Qualitative explanatory var $\rightsquigarrow$ subsamples $T_1, T_2, \ldots$

  according to category or characteristics
- examples:
  - pure qualitative vars:
    - individual diffs: sex, race, civil state, etc.
    - time diffs: season, war/peace, etc.
    - spatial diffs: countries, A.C.'s, urban/rural, etc.
  - quantitative vars by sections: income, age, etc.
- Recall: we cannot use qualitative vars...

  then substitute by **dummy vars**...
- Def. of Dummy Variable:

$$D_{jt} = \begin{cases} 1, & \text{if } t \in \text{category } j \text{ ;} \\ 0, & \text{otherwise.} \end{cases}$$

$$\Rightarrow D_{jt} = \mathscr{I}(t \in T_j)$$

1 QV with 2 categories

$$\text{Consumption} = f([\text{ctnt}],\ \text{income},\ \text{sex})$$

$$\downarrow \qquad \downarrow \qquad \downarrow \qquad \swarrow\ \searrow$$

$$Y_t \qquad [1] \qquad R_t \qquad M \quad F$$

$$S_{1t} = \mathscr{I}(t \in M) \qquad S_{2t} = \mathscr{I}(t \in F)$$

Sample:

t	Y	ctnt	R	S
1	Y_1	1	R_1	M
2	Y_2	1	R_2	F
3	Y_3	1	R_3	F
⋮	⋮	⋮	⋮	⋮
T	Y_T	1	R_T	M
		X ?		

$\Rightarrow$

t	Y	ctnt	R	S_1	S_2
1	Y_1	1	R_1	1	0
2	Y_2	1	R_2	0	1
3	Y_3	1	R_3	0	1
⋮	⋮	⋮	⋮	⋮	
T	Y_T	1	R_T	1	0
		X			

In principle: substitute QV by

as many DVs as categories we have.

Dummy Var Trap: 1 qualitative var

- Model: $Y_t = \beta_0 + \beta_1 R_t + \gamma_1 S_{1t} + \gamma_2 S_{2t} + u_t$
- Problem (Dummy Variable trap):
  X is a $(T \times 4)$ matrix, but

$$S_1 + S_2 = [1] \text{ (exact l.c.)} \Rightarrow \text{rk}(X) = 3 < 4 \quad (\textit{i.e.} \text{ perfect MC})$$

- $$\Rightarrow \det(X'X) = 0$$
  $$\Rightarrow (X'X)^{-1} \text{ doesn't exist!! and}$$
  $$\widehat{\beta} \text{ cannot be calculated!!}$$

- General Solution: eliminate ONE of the col's causing the problem: [1] or S_1 or S_2 .
- (POSSIBLE Solution: eliminate intercept. . . but. . .

Solution: DV without a category

MOST USUAL SOLUTION: eliminate category: *e.g.* F (S_2):

- Model to estimate:

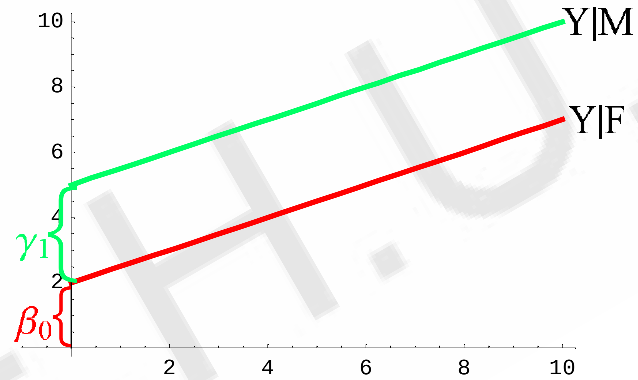
$$Y_t = \beta_0 + \beta_1 R_t + \gamma_1 S_{1t} + \not{\gamma_2} S_{2t} + u_t$$
$$= \beta_0 + \beta_1 R_t + \gamma_1 S_{1t} + u_t$$

without category F

- Subsample Models:

$$\mathrm{E}(Y_t | S = F) = \beta_0 + \beta_1 R_t$$
$$\mathrm{E}(Y_t | S = M) = \beta_0 + \beta_1 R_t + \gamma_1$$

$\Rightarrow$

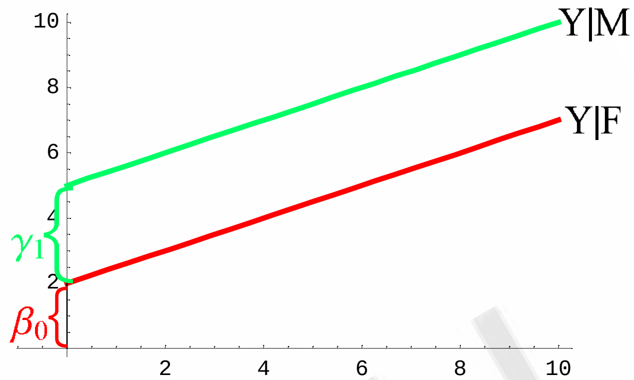


- Coefficient interpretation:

$$\mathrm{E}(Y_t | R_t = 0, S = F) = \beta_0$$
$$\mathrm{E}(Y_t | S = M) - \mathrm{E}(Y_t | S = F) = \gamma_1$$

Coefficient Interpretation

without category F



$$\mathrm{E}\left(Y_t|S=M\right)-\mathrm{E}\left(Y_t|S=F\right)=\gamma_1$$

$$\mathrm{E}\left(Y_t|R_t=0,S=F\right)=\beta_0$$

- that is,

$$\beta_0 = \text{expected consumption Women (base) if } R_t = 0.$$

$$\gamma_1 = \text{diff expected consumption of Men}$$

$$\text{(vs. base} = \text{Women).}$$

$$\beta_1 = \Delta \text{ consumption if } \Delta R_t = 1 \text{ (c.p.).}$$

Recall: This case just means different intercepts for each category.
Note: Eliminating a category $\rightsquigarrow$ transforms it into reference base.

Usual Tests with 1 QV

Hypothesis: qualitative variable (Sex) not significant (it doesn't affect Consumption)

i.e. M and F same Consumption:

- Unrestricted Model

$$Y_t = \beta_0 + \beta_1 R_t + \gamma S_{1t} + u_t$$

- Hypothesis: $H_0 : \gamma = 0$ vs. $H_a : \gamma \neq 0$
- Restricted Model:

$$Y_t = \beta_0 + \beta_1 R_t + u_t$$

- Use usual t Statistic (or F Statistic based on RSS)

1 QV with 2 cats + 1 QV with 3 cats

$$\text{Consumption} = f([\text{ctnt}],\ \text{income},\ \text{sex},\ \text{territory CAV})$$

- Consumption → Y_t
- ctnt → $[1]$
- income → R_t
- sex → M, F
- territory CAV → A, B, G

$$S_{1t} = \mathscr{I}(t \in M) \qquad S_{2t} = \mathscr{I}(t \in F)$$

$$T_{1t} = \mathscr{I}(t \in A) \qquad T_{2t} = \mathscr{I}(t \in B) \qquad T_{3t} = \mathscr{I}(t \in G)$$

Sample:

t	Y	ctnt	R	S	T
1	Y_1	1	R_1	M	B
2	Y_2	1	R_2	F	G
3	Y_3	1	R_3	F	B
⋮	⋮	⋮	⋮	⋮	⋮
T	Y_T	1	R_T	M	A

X ?

⇒

t	Y	ctnt	R	S_1	S_2	T_1	T_2	T_3
1	Y_1	1	R_1	1	0	0	1	0
2	Y_2	1	R_2	0	1	0	0	1
3	Y_3	1	R_3	0	1	0	1	0
⋮	⋮	⋮	⋮	⋮		⋮		
T	Y_T	1	R_T	1	0	1	0	0

X

Recall: In principle, substitute qualitative var by as many Dummy vars as categories we have.

Dummy Var Trap: 2 qualitative vars

- Model: $Y_t = \beta_0 + \beta_1 R_t + \gamma_1 S_{1t} + \gamma_2 S_{2t} + \delta_1 T_{1t} + \delta_2 T_{2t} + \delta_3 T_{3t} + u_t$
- Problem (DV trap):
  X is a $(T \times 7)$ matrix, but

$$S_1 + S_2 = T_1 + T_2 + T_3 = [1]$$

(2 exact l.c.) $\Rightarrow \mathrm{rk}(X) = 5 < 7$ (*i.e.* perfect MC)

- $\Rightarrow \det(X'X) = 0$
  $\Rightarrow (X'X)^{-1}$ doesn't exist!! and
  $\widehat{\beta}$ cannot be calculated!!

- General Solution: eliminate ONE of the col's causing the problem: [1] or (S_1 or S_2) or (T_1 or T_2 or T_3).

Solution: DV without combination of categories

MOST USUAL SOLUTION:

eliminate last category of each DV: S_2 and T_3:

- Model to estimate:

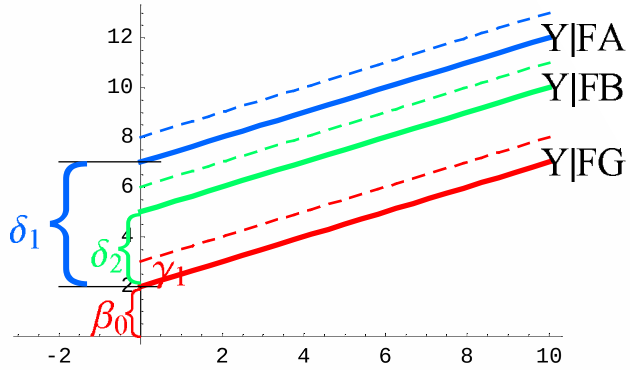
$$Y_t = \beta_0 + \beta_1 R_t + \gamma_1 S_{1t} + \cancel{\gamma_2} S_{2t} + \delta_1 T_{1t} + \delta_2 T_{2t} + \cancel{\delta_3} T_{3t} + u_t$$
$$= \beta_0 + \beta_1 R_t + \gamma_1 S_{1t} + \delta_1 T_{1t} + \delta_2 T_{2t} + u_t$$

- Subsample Models:

	$S = M$	$S = F$	$M - F$
$T = A$	$\beta_0 + \beta_1 R_t + \gamma_1 + \delta_1$	$\beta_0 + \beta_1 R_t + \delta_1$	γ_1
$T = B$	$\beta_0 + \beta_1 R_t + \gamma_1 + \delta_2$	$\beta_0 + \beta_1 R_t + \delta_2$	γ_1
$T = G$	$\beta_0 + \beta_1 R_t + \gamma_1$	$\beta_0 + \beta_1 R_t$	γ_1
$A - G$	δ_1	δ_1	
$B - G$	δ_2	δ_2	
$A - B$	$\delta_1 - \delta_2$	$\delta_1 - \delta_2$	

Coefficient Interpretation

without categories F nor G



$$\mathsf{E}(Y_t|S=M) - \mathsf{E}(Y_t|S=F) = \gamma_1$$

$$\mathsf{E}(Y_t|T=A) - \mathsf{E}(Y_t|T=G) = \delta_1$$

$$\mathsf{E}(Y_t|T=B) - \mathsf{E}(Y_t|T=G) = \delta_2$$

$$\mathsf{E}(Y_t|R_t=0, S=F, T=G) = \beta_0$$

- that is,

$\beta_0 =$ expected consumption Women G (base) if $R_t = 0$.

$\gamma_1 =$ diff. expected consumption Men vs. Women .

$\delta_1 =$ diff. expected consumption A vs. G.

$\delta_2 =$ diff. expected consumption B vs. G.

$\beta_1 = \Delta$ consumption if $\Delta R_t = 1$ (*c.p.*).

Recall: This case just means different intercepts for each category. Recall: Eliminating a (combination of) category(ies)

$\rightsquigarrow$ transforms it into reference base.

Usual Tests with 2 QVs

Hypothesis: Variable Sex doesn't affect Consumption
(but place of residence might do)

- Unrestricted Model:

$$Y_t = \beta_0 + \beta_1 R_t + \gamma_1 S_{1t} + \delta_1 T_{1t} + \delta_2 T_{2t} + u_t$$

($\gamma_1 =$ diff. exp. C of M vs. F)

- Hypothesis: $H_0 : \gamma_1 = 0$ vs. $H_a : \gamma_1 \neq 0$
- Restricted Model:

$$Y_t = \beta_0 + \beta_1 R_t + \delta_1 T_{1t} + \delta_2 T_{2t} + u_t$$

- Use usual t Statistic (or F Statistic based on RSS)

Other usual Tests with 2 QVs

- Unrestricted Model (without S_2 nor T_3):

$$Y_t = \beta_0 + \beta_1 R_t + \gamma_1 S_{1t} + \delta_1 T_{1t} + \delta_2 T_{2t} + u_t$$

  - Recall: γ_1 is diff. expected C of M vs. F (base)
    δ_1 and δ_2 are diff. exp. C of A and B vs. G (base)

- Hypothesis: Same Consumption overall (independently of Sex and Residence):
  - $H_0 : \gamma_1 = \delta_1 = \delta_2 = 0$
  - Restricted Model:

$$Y_t = \beta_0 + \beta_1 R_t + u_t$$

- Hypothesis: Place of Residence doesn't affect Consumption (but M vs. F might do):
  - $H_0 : \delta_1 = \delta_2 = 0$
  - Restricted Model:

$$Y_t = \beta_0 + \beta_1 R_t + \gamma_1 S_{1t} + u_t$$

Other usual Tests with 2 QVs

- Unrestricted Model (without S_2 nor T_3):

$$Y_t = \beta_0 + \beta_1 R_t + \gamma_1 S_{1t} + \delta_1 T_{1t} + \delta_2 T_{2t} + u_t$$

  - Recall: δ_1 and δ_2 are diff. expected C of A and B vs. G (base)

- Hypothesis: Residents of same sex in A and B have same consumption level (but G might be different):
  - $H_0 : \delta_1 = \delta_2$ vs. $H_a : \delta_1 \neq \delta_2$
  - Restricted Model:

$$Y_t = \beta_0 + \beta_1 R_t + \gamma_1 S_{1t} + \delta \underbrace{(T_{1t} + T_{2t})}_{1-T_{3t}} + u_t$$

- Hypothesis: Residents of same sex in B and G have same consumption level (but A might be different):
  - $H_0 : \delta_2 = 0$ vs. $H_a : \delta_2 \neq 0$
  - Restricted Model:

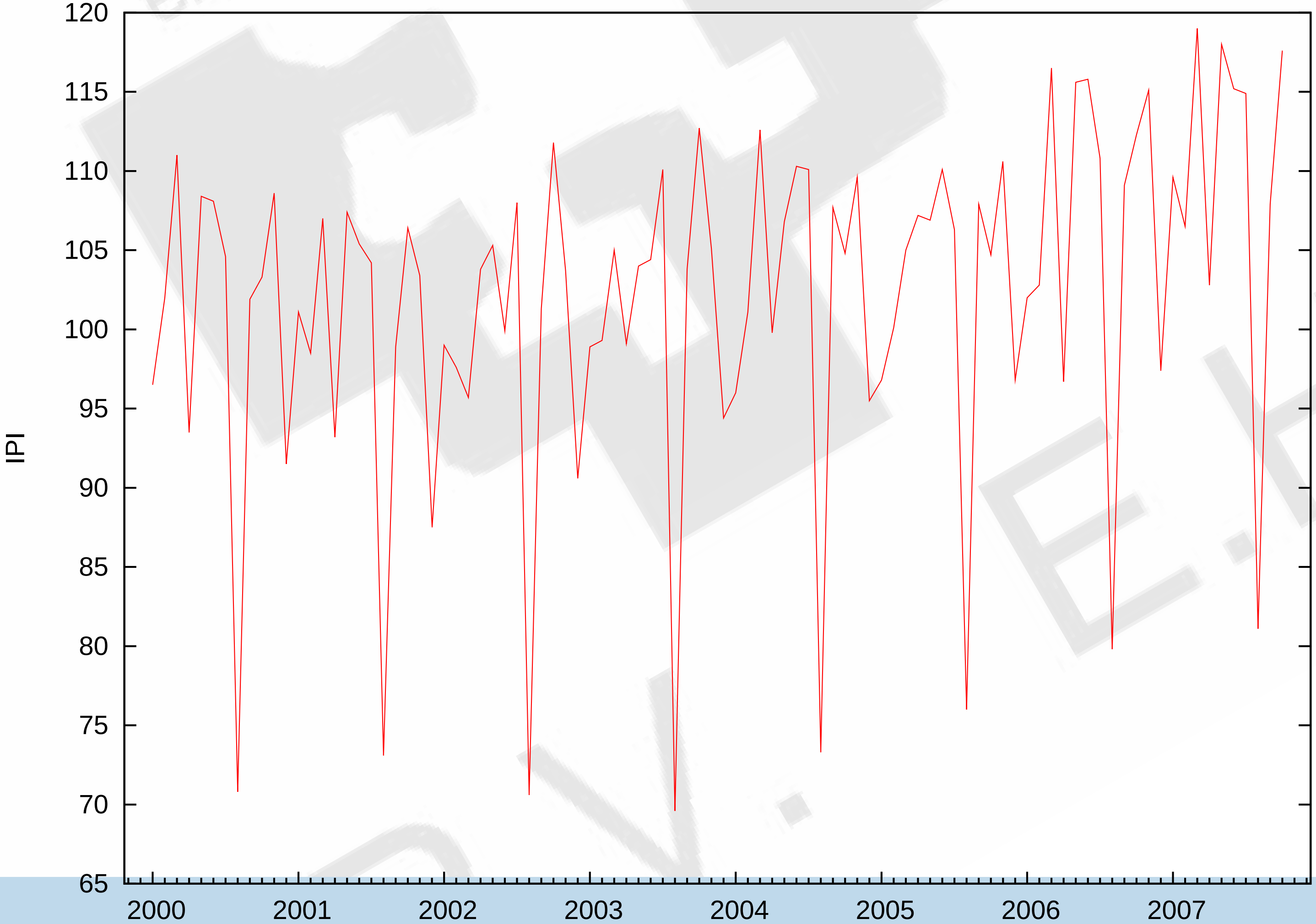
$$Y_t = \beta_0 + \beta_1 R_t + \gamma_1 S_{1t} + \delta_1 T_{1t} + u_t$$

4.2 Seasonal effects

Seasonal effect

- Seasonal effect:
-
- Seasonal var $\rightsquigarrow$ subsamples $T_1, T_2, \ldots$

according to seasons/months



Seasonal Dummy Variables: Definition

- Def. of Seasonal Dummy Variable:

$$D_{jt} = \begin{cases} 1, & \text{if } t \in \text{season } j = 1, 2, 3, 4, \ldots; \\ 0, & \text{otherwise.} \end{cases}$$

- *e.g.* for quarterly data:

date (t)	IPI_t	X_t	D_{1t}	D_{2t}	D_{3t}	D_{4t}
1975.1	·	·	1	0	0	0
1975.2	·	·	0	1	0	0
1975.3	·	·	0	0	1	0
1975.4	·	·	0	0	0	1
1976.1	·	·	1	0	0	0
1976.2	·	·	0	1	0	0
1976.3	·	·	0	0	1	0
1976.4	·	·	0	0	0	1
1977.1	·	·	1	0	0	0
⋮	⋮	⋮	...			
2000.1	·	·	1	0	0	0
2000.2	·	·	0	1	0	0
2000.3	·	·	0	0	1	0
2000.4	·	·	0	0	0	1
2001.1	·	·	1	0	0	0
...	...	...	...			

Seasonal Dummy Variables: Definition (2)

- Model to estimate:

$$IPI_t = \beta_0 + \beta_1 X_t + \gamma_1 D_{1t} + \gamma_2 D_{2t} + \gamma_3 D_{3t} + \cancel{\gamma_4} D_{4t} + u_t$$
$$= \beta_0 + \beta_1 X_t + \gamma_1 D_{1t} + \gamma_2 D_{2t} + \gamma_3 D_{3t} + u_t$$

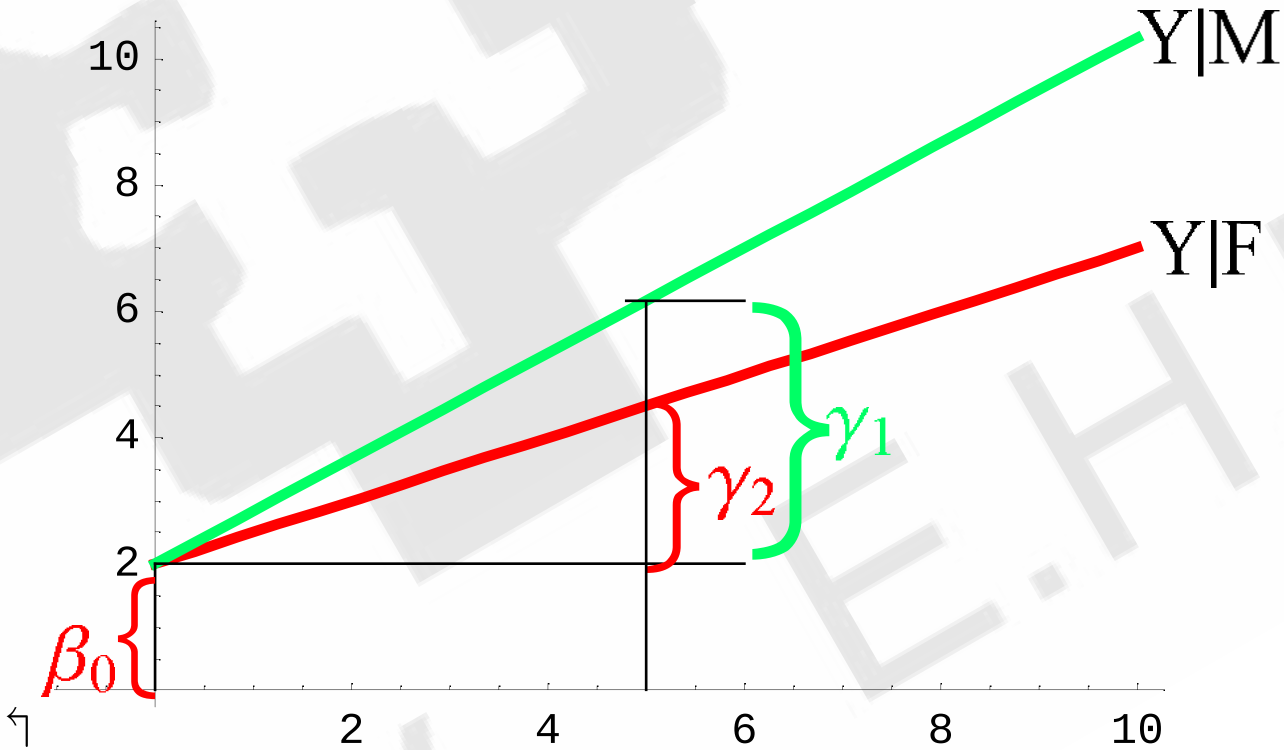
- interpretation of γ parameters?
- What if data are monthly observations (as in the IPI example actually)?

date (t)	IPI_t	X_t	D_{1t}	D_{2t}	D_{3t}	D_{4t}		D_{1t} ...											$\ldots D_{12t}$
1975.jan	·	·	1	0	0	0		1	0	0	0	0	0	0	0	0	0	0	0
1975.feb	·	·	1	0	0	0		0	1	0	0	0	0	0	0	0	0	0	0
1975.mar	·	·	1	0	0	0		0	0	1	0	0	0	0	0	0	0	0	0
1975.apr	·	·	0	1	0	0		0	0	0	1	0	0	0	0	0	0	0	0
1975.may	·	·	0	1	0	0		0	0	0	0	1	0	0	0	0	0	0	0
1975.jun	·	·	0	1	0	0		0	0	0	0	0	1	0	0	0	0	0	0
1975.jul	·	·	0	0	1	0		0	0	0	0	0	0	1	0	0	0	0	0
1975.ago	·	·	0	0	1	0	or	0	0	0	0	0	0	0	1	0	0	0	0
1975.sep	·	·	0	0	1	0		0	0	0	0	0	0	0	0	1	0	0	0
1975.oct	·	·	0	0	0	1		0	0	0	0	0	0	0	0	0	1	0	0
1975.nov	·	·	0	0	0	1		0	0	0	0	0	0	0	0	0	0	1	0
1975.dec	·	·	0	0	0	1		0	0	0	0	0	0	0	0	0	0	0	1
1976.jan	·	·	1	0	0	0		1	0	0	0	0	0	0	0	0	0	0	0
1976.feb	·	·	1	0	0	0		0	1	0	0	0	0	0	0	0	0	0	0
1976.mar	·	·	1	0	0	0		0	0	1	0	0	0	0	0	0	0	0	0
...	...	...	...					...											

4.3 Interaction between DVs and quantitative Vars

Interaction between DVs and quantitative Vars

Instead of different *intercepts*, we require different slopes for each category:



that is, different response "Y" for same "X"

Dummy Var Trap: interaction

- Matrix X:

ctnt	R	$R \times S_1$	$R \times S_2$
1	R_1	$R_1 \times 1$	$R_1 \times 0$
1	R_2	$R_2 \times 0$	$R_2 \times 1$
1	R_3	$R_3 \times 0$	$R_3 \times 1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
1	R_T	$R_T \times 1$	$R_T \times 0$

$\Rightarrow$

ctnt	R	RS_1	RS_2
1	R_1	R_1	0
1	R_2	0	R_2
1	R_3	0	R_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$
1	R_T	R_T	0

- Model:

$$Y_t = \beta_0 + \beta_1 R_t + \gamma_1 R_t S_{1t} + \gamma_2 R_t S_{2t} + u_t$$

- Problem (DV trap): X is $T \times 4$, but

$$RS_1 + RS_2 = R \Rightarrow \mathrm{rk}(X) = 3 < 4 \quad \text{(exact MultiCol!)}$$

- General Solution: eliminate ONE of the col's causing the problem: R or RS_1 or RS_2 .

Solution: Interaction without a category

- MOST USUAL SOLUTION:

  eliminate last category of the DV: $F\ (RS_2\ )$:

- Model to estimate:

$$Y_t = \beta_0 + \beta_1 R_t + \gamma_1 R_t S_{1t} + u_t$$

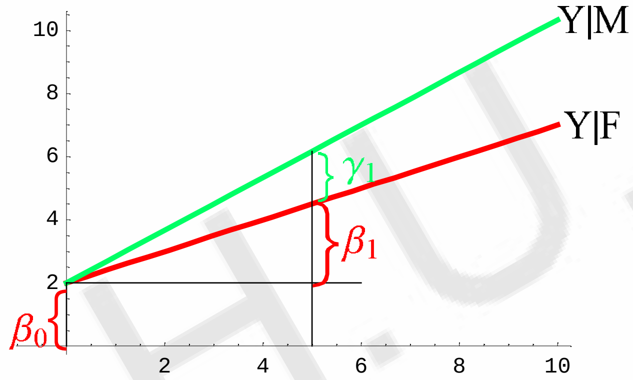
no RS_2

- Subsample Models:

$$E\left(Y_t | S = F\right) = \beta_0 + \beta_1 R_t$$

$$E\left(Y_t | S = M\right) = \beta_0 + \underbrace{(\beta_1 + \gamma_1)}_{\beta_1^*} R_t$$

$\Rightarrow$



- Coefficient interpretation:

$$E\left(Y_t | R_t = 0\right) = \beta_0$$

$$\frac{\partial E\left(Y_t | S = F\right)}{\partial R_t} = \beta_1$$

$$\frac{\partial E\left(Y_t | S = M\right)}{\partial R_t} = \beta_1 + \gamma_1$$

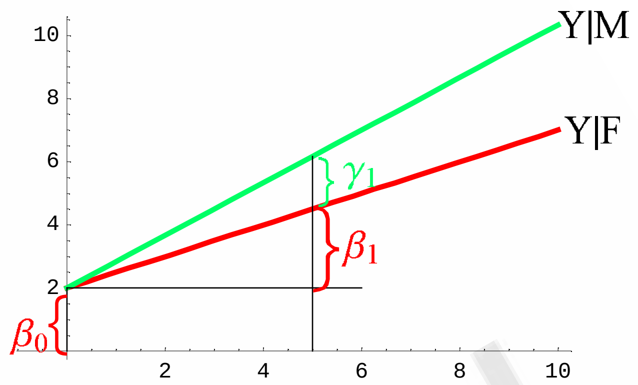
Coefficient Interpretation

$$\mathsf{E}\left(Y_t | R_t = 0\right) = \beta_0$$

$$\frac{\partial \mathsf{E}\left(Y_t | S = F\right)}{\partial R_t} = \beta_1$$

$$\frac{\partial \mathsf{E}\left(Y_t | S = M\right)}{\partial R_t} = \beta_1 + \gamma_1$$

no RS_2



- that is,

$\beta_0 =$ expected consumption if $R_t = 0$.

$\beta_1 = \Delta$ consumption Women if $\Delta R_t = 1$ (*c.p.*).

$\gamma_1 =$ diff Δ consumption for Men (vs. base $=$ Female).

Recall: This case means different slopes for each category. Recall: again eliminating a category $\rightsquigarrow$

transforms it into reference base.

Usual Tests with Interaction

Hypothesis: M and F equal Consumption
or variable Sex doesn't affect Consumption:

- Unrestricted Model:

$$Y_t = \beta_0 + \beta_1 R_t + \gamma_1 R_t S_{1t} + u_t$$

- Hypothesis: $H_0 : \gamma_1 = 0$ vs. $H_a : \gamma_1 \neq 0$
- Restricted Model:

$$Y_t = \beta_0 + \beta_1 R_t + u_t$$

- Use usual t Statistic (or F Statistic based on RSS)

THE END