Lecture 3-a
Least Squares – Introduction, &
Review of Linear Algebra

Brooks (4th edition): Chapter 3

Review – C.I. Application: Value-at-Risk (VaR)

• Q: What is the most an investor can lose with a particular investment over a given time framework? Or, what is the worst case scenario?

• **Value-at-Risk** (**VaR**) provides one answer, a (lower) bound with a probability attached to it.

• So far, we have measured risk of an asset/investment with its volatility, which is calculated including positive (right tail) and negative (left tail) returns. Investors, however, love the right tail of the returns distribution, but dislike the other tail. **VaR** focuses on the left tail.

• **VaR** gives a formal definition of left-tail risk, a "worst case scenario," for an asset over a time period.

Review – C.I. Application: Value-at-Risk

• VaR gives a formal definition of "worst case scenario" for an asset.

VaR: *Maximum expected amount (loss)* in a given time interval within a (*one-sided*) $(1 - \alpha)\%$ C.I.:

  VaR$(1 - \alpha)$ = Amount exposed * (1 + worst % change scenario in C.I.)

It is common to express the "*expected loss*" relative to today's expected value of asset/investment:
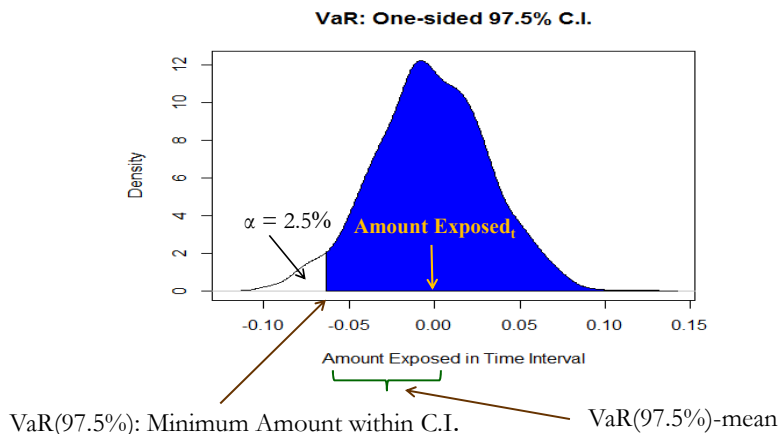
  VaR-mean$(1 - \alpha)$ = VaR – E[Amount exposed]

• There are different ways to compute the worst case scenario within a time interval. We go over two approaches:
  - Assuming a probability distribution (normal, in our case).
  - Using the empirical distribution (a bootstrap, using the past).

Review – C.I. Application: Value-at-Risk

Example: $\alpha = .05$
VaR = Amount exposed * (1 + worst change scenario in 97.5% C.I.).
VaR-mean(97.5%) = VaR – E[Amount exposed]

VaR: One-sided 97.5% C.I.

$\alpha = 2.5\%$ **Amount Exposed$_t$**

Density

Amount Exposed in Time Interval

VaR(97.5%): Minimum Amount within C.I.   VaR(97.5%)-mean

Review – C.I. Application: VaR in FX Markets

• When a company is involved with transactions denominated in foreign currency (FC), it is exposed to *currency risk*. Transaction exposure (TE) provides a simple measure of this exposure:

TE_t = Value of a fixed future transaction in FC * S_t

where S_t is the exchange rate expressed as units of domestic currency (USD for us) per unit of FC (say, EUR).

Example: A Swiss company, Swiss Cruises, sells packages in USD.
Amount = **USD 1 million**.
Payment: 30 days.
S_t = **0.92 CHF/USD**

$\Rightarrow$ **TE_t = USD 1M * 0.92 CHF/USD = CHF 0.92M**. ¶

Review – C.I. Application: VaR (Normal)

• Swiss Cruises wants a measure of the uncertainty related to the amount to receive in CHF in 30 days, since S_{t+30} is unknown.

We can use a range to quantify this uncertainty, we want to say
$TE_{t+30} \in [TE_{LB}, TE_{UB}]$ with high probability.

To determine this range for TE, we **assume** that (log) changes in S_t, $e_{f,t}$, are **normally distributed**: $e_{f,t} \sim N(\mu, \sigma^2)$.

Then, we build a (1 - α)% C.I.: $[\mu \pm z_{1-\alpha/2} * \sigma]$.

Usual α's in interval calculations: α =.05 $\Rightarrow |z_{.025}|$ = **1.96** (≈2)
                                                          α =.02 $\Rightarrow |z_{.01}|$ = **2.33**

As usual, we estimate (μ, σ) using ($\bar{X}$, s).

Review – C.I. Application: VaR (Normal)

Example: Range estimate based on a **Normal distribution**.
Assume Swiss Cruises believes that CHF/USD monthly changes follow a normal distribution. Swiss Cruises estimates the mean and variance using the last 15 years of data.

$\bar{X}$ = Monthly mean = **-0.00152** ≈ -0.15%
s^2 = Monthly variance = 0.001014 (⇒ s = **0.03184**, or 3.18%)
$e_{f,t}$ ~ N(**-0.00152**, **0.03184**^{2}) $e_{f,t}$ = CHF/USD log changes.

Swiss Cruises constructs a 95% CI for CHF/USD monthly changes.

Recall that a 95% C.I. for $e_{f,t+30}$ (which applies to any t) is given by:
$$e_{f,t} \in [\text{-0.00152} \pm \text{1.96} * \text{0.03184}] = [\text{-0.06393}; \text{0.06089}].$$

Based on this range for $e_{f,t}$, we can build a 95% C.I. for S_{t+30} and, then, for TE$_{t+30}$ (= **USD 1M** * S_{t+30}).

Review – C.I. Application: VaR (Normal)

Example (continuation):
$$e_{f,t+30} \in [\text{-0.00152} \pm \text{1.96} * \text{0.03184}] = [\text{-0.06393}; \text{0.06089}].$$

Now, based on the range $e_{f,t}$, we derive a range for S_{t+30}:

(A) Upper bound
$S_{t+30,UB} = S_t * (1 + e_{f,UB}) = $ **0.92 CHF/USD** $* (1 + $ **0.06089**$) =$
     = **0.97602**

(B) Lower bound
$S_{t+30,LB} = S_t * (1 + e_{f,LB}) = $ **0.92 CHF/USD** $* (1 - $ **0.06393**$)] =$
     = **0.86118**

⇒ $S_{t+30} \in [$**0.97602 CHF/USD**; **0.86118 CHF/USD**$]$.

Review – C.I. Application: VaR (Normal)

Example (continuation): Finally, we derive the bounds for the TE:

(A) Upper bound ($S_{t+30,UB} = S_t * (1+ e_{f,UB}) = $ **0.97602 CHF/USD**)
TE$_{UB}$: **USD 1M** * **0.97602 CHF/USD = CHF 976,019**.

(B) Lower bound ($S_{t+30,LB} = S_t * (1+ e_{f,LB}) = $ **0.86118 CHF/USD**)
TE$_{LB}$: **USD 1M** * **0.86118 CHF/USD = CHF 861,184**.

$$\Rightarrow \text{TE}_{t+30} \in [\text{CHF 0.861 M}; \text{ CHF 0.976 M}]. \text{¶}$$

• The lower bound, for a receivable, represents the worst case scenario within the interval. This is the **VaR** interpretation:

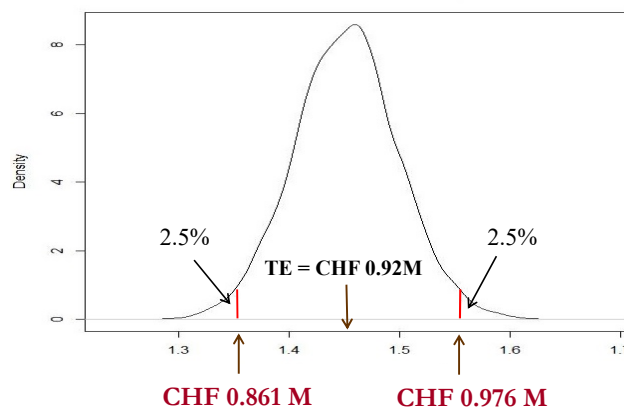VaR: *Maximum expected amount (loss)* in a given time interval within a *(one-sided)* confidence interval.
VaR = Amount exposed * (1+ worst % change scenario in C.I.)

Review – C.I. Application: VaR (Normal)

VaR = Amount exposed * (1+ worst % change scenario in C.I.).
Then, in our application: VaR = **TE$_t$** * (1 + $e_{f,LB}$)

VaR(97.5%): Minimum revenue within a 97.5% C.I.

Review – C.I. Application: VaR (Normal)

Example (continuation): The minimum revenue to be received by SC in the next 30 days, within a 97.5% CI.

$$\text{VaR}(97.5\%) = \textbf{CHF 0.92M} * [1+ (\textbf{-0.00152} - \textbf{1.96} * \textbf{0.03184})]$$
$$= \textbf{CHF 0.8612M}.$$

<u>Interpretation of VaR</u>: If SC expects to cover expenses with this USD inflow, the maximum amount in CHF to cover, within a 97.5% one-sided CI, should be **CHF 0.8612M**. ¶

It is common to express the "*expected loss*" relative to today's value of a transaction (or asset):

$$\text{VaR-mean} = \text{VaR} - \textbf{TE}_t = \textbf{TE}_t * (1 + e_{f,LB}) - \textbf{TE}_t$$
$$= \textbf{TE}_t * e_{f,LB}$$

Or just

$$\text{VaR-mean} = \text{Amount exposed} * \text{worst case scenario}$$

Review – C.I. Application: VaR (Normal)

Example (continuation): Relative to today's valuation (or *expected valuation*, according to RWM), the maximum expected loss in 30 days within a 97.5% one-sided C.I. is:
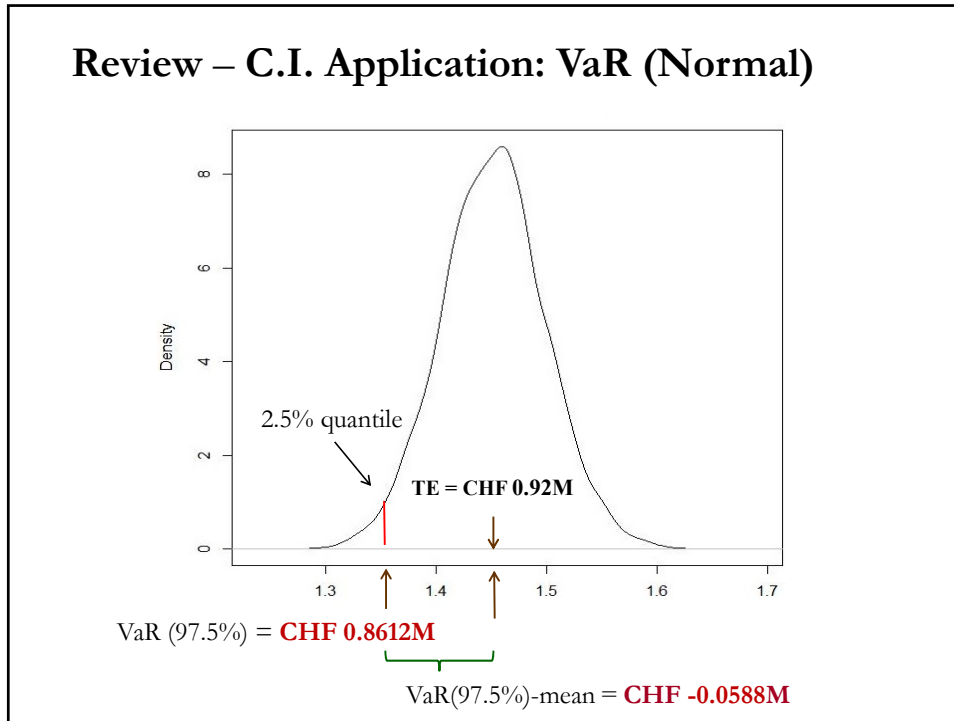
VaR-mean(.975) = **CHF 0.8612M** – **CHF 0.92M** = **CHF -0.0588M**.

Note that we can also compute the VaR-mean as:

VaR-mean(.975) = **CHF 0.92M** * (**-0.00152** – **1.96** * **0.03184**)
$$= \textbf{CHF -0.0588M}. ¶$$

• Technically speaking, the VaR is a *quantile*. A quantile is the fraction of observations that lie below a given value (in this case, the VaR).

In the previous example, the 0.025 quantile (or 2.5% quantile) for expected loses is **CHF -0.0588M**.

Review – C.I. Application: VaR (Normal)

2.5% quantile

TE = CHF 0.92M

VaR (97.5%) = **CHF 0.8612M**

VaR(97.5%)-mean = **CHF -0.0588M**

Review – C.I. Application: VaR (Normal)

Example (continuation):

Note: We could have used a different quantile –i.e. a different significant level- to calculate the VaR, for example 1% ($\Rightarrow z_{.99} = $ **2.33**). Then,
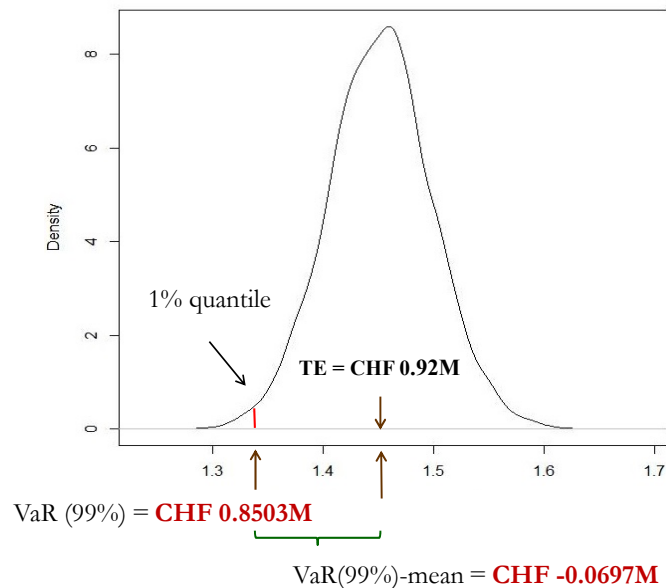
VaR(99%) = **CHF 0.92M** * [1+ (**-0.00152** – **2.33** * **0.03184**)]
= **CHF 0.8503M** (A more *conservative* bound.)

$\Rightarrow$ VaR-mean (.99) = **CHF 0.8503M** – **CHF 0.92M** = **CHF -0.0697M**

Interpretation of VaR-mean: Relative to today's valuation, the maximum *expected loss* with a 99% "chance" is **CHF -0.0697M**.

Note: As the C.I. gets wider, Swiss Cruises can spend less CHF on account of the USD 1M receivable. ¶

Review – C.I. Application: VaR (Normal)

Review – C.I. Application: VaR (Bootstrap)

• VaR is a statistic –a function of the data. We can do an empirical bootstrap to calculate the mean, SE (=SD), C.I., etc.

Example: We want to calculate the average VaR(97.5%) and its S.E., using all CHF/USD data from 1990:Jan - 2023:July. Then,

```
chfusd <- read.csv("http://www.bauer.uh.edu/rsusmel/4386/chfusd.csv",sep=",")  # Data
S <- chfusd$CHF_USD                         # Extract CHF_USD column of the data
T <- length(S)                              # Check total T (1971:1 to 2023:7)
Tstart <- 229                               # Start of sample period: 1990:1
SP <- S[Tstart: T]
T <- length(SP)
Val <- 1000000                              # Value of transaction in FC (in M)
S_0 <- S[T]                                 # Today's S_t
e_f <- log(SP[-1]/SP[-T])
T_s <- length(e_f)
alpha = .05                                 # Specify alpha level for VaR
T_s_low <- round(T_s*alpha/2)               # Obs corresponding to alpha/2*T_s
TE_o <- Val*S_0*(1+e_f)                      # calculate Original TE values
```

C.I. Application: VaR in FX Markets (Bootstrap)

Example (continuation):

```
STE_o <- sort(TE_o)                                    # sort Original TE
VaR_o <- STE_o[T_s_low]                                # Original VaR
> VaR_o
[1] 860293

# function to obtain VaR from the data
varisk <- function(data, i) {
d <-data[i]
TE <- Val*S_0*(1+d)                                    # calculate R TE values
STE <- sort(TE)                                        # sort TE
VaR <- STE[T_s_low]
return(VaR)
}
library(boot)
sim_size <- 1000
boot.samps <- boot(data=e_f, statistic=varisk, R=sim_size)   # use boot to build samples
```

C.I. Application: VaR in FX Markets (Bootstrap)

Example (continuation):

```
> boot.samps
Bootstrap Statistics :
    original   bias   std. error
t1*  860293 1929.305  4870.733

> boot.ci(boot.samps, type = "basic")
Intervals :
Level    Basic
95%   (849352, 867587 )

> mean(boot.samps$t)
[1] 862222.3
> sd(boot.samps$t)
[1] 4870.733

hist(boot.samps$t, xlab="VaR (in CHF)", breaks=30)
```
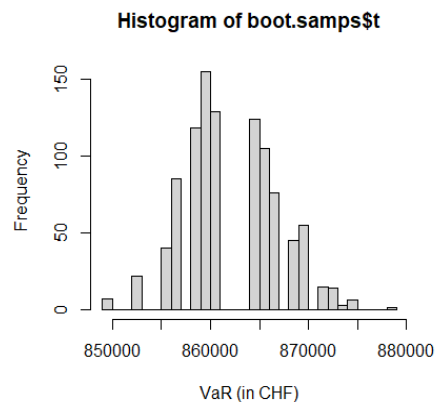

Histogram of boot.samps$t

Bootstrap estimated VaR(97.5%) = **CHF 0.8622M**

Review – C.I. Application: VaR (Bootstrap)

Example (continuation):

```
VaR_o <- STE_o[T_s_low]                          # Original VaR
> VaR_o
[1] 860293
# function to obtain VaR from the data
varisk <- function(data, i) {
d <-data[i]
TE <- Val*S_0*(1+d)                              # calculate R TE values
STE <- sort(TE)                                  # sort TE
VaR <- STE[T_s_low]
return(VaR)
}
boot.samps <- boot(data=e_f, statistic=varisk, R=sim_size)
> boot.samps
Bootstrap Statistics :
    original   bias    std. error
t1*   860293 1929.305    4870.733
```

Review – C.I. Application: VaR (Bootstrap)

Example (continuation):

```
> boot.ci(boot.samps, type = "basic")            # boot computes the CI.
Intervals :
Level     Basic
95%   (849352, 867587 )

> mean(boot.samps$t)
[1] 862222.3
> sd(boot.samps$t)
[1] 4870.733

hist(boot.samps$t, xlab="VaR (in CHF)",
breaks=30)
```
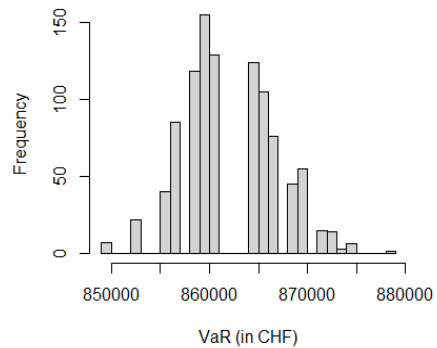

Histogram of boot.samps$t

Bootstrap estimated VaR(97.5%) = **CHF 0.8622M**

Review – Previous Classes: Summary

• Defined a random variable (say, stock returns) and how to describe its distribution with moments (mean, variance, skewness, etc).

• Testing hypothesis (H_0) about the behavior of the RV. For example, the mean return of the S&P 500 is 0 (H_0: $\mu=0$).

• Building Confidence Intervals: Assuming a distribution or a bootstrap.

• Basic R and simple applications.

Modeling a Dependent Variable as a Function

• So far, we focused on one RV only, say stock returns and learning about its distribution, for example, using descriptive statistics. In econometrics, we are more interested in describing or measuring the expected effect of size on stock returns or the expected effect of education on an employee's salary or a CEO's compensation.

• That is, we usually care about a functional relation between y, the *dependent variable,* and x, a set (a vector!) of *explanatory variables.*

• In this lecture, we will **linearly** relate y to x & an *error term,* ε:
$$y_i = \alpha + \beta\, x_i + \varepsilon_i, \qquad\qquad i = 1, 2, ..., T$$
where α & β are parameters to be estimated and ε_i is the error term or *disturbance* that has zero mean and constant variance, σ^2. That is, ε_i is a RV with $E[\varepsilon_i] = 0$ & $Var[\varepsilon_i] = \sigma^2$.

Modeling a Dependent Variable: Linear Model

• In our linear model:
$$y_i \ = \ \alpha + \beta \, x_i + \varepsilon_i, \qquad\qquad i = 1, 2, ..., N.$$
We think of ε_i as the effect of individual variation that is not "controlled for" with x_i. The disturbance, ε_i, is part of the model. We call the above equation the **Data Generating Process** (DGP).

Then, under the assumption $E[\varepsilon_i] = 0$, we have:
$$E[y_i] = \ \alpha + \beta \, E[x_i].$$

Example: The Sharpe-Litner CAPM posits a relation between the excess return of asset i, $r_{i,t} - r_f$, and the excess return of the market, $r_{m,t} - r_f$. In equilibrium, the CAPM states:
$$E[(r_{i,t} - r_f)] = \beta_i \, E[(r_{m,t} - r_f)],$$
where β_i is the sensitivity of asset i to market risk.

Modeling a Dependent Variable: Linear Model

Example (continuation): Let asset i = IBM. The previous mathematical structure allows us to estimate $\beta_{i=IBM}$, compute expected excess returns for IBM, and test the CAPM for IBM. Define

y = excess returns for IBM

x = excess returns for the market (the "Market").

We express the underlying relation behind the CAPM as:
$$y_i \ = \ \alpha + \beta \, x_i + \varepsilon_i, \qquad\qquad i = 1, 2, ..., N.$$
where α & β are parameters to be estimated and ε is the *error term* with $E[\varepsilon] = 0$ & $Var[\varepsilon] = \sigma^2$. Taking expectations:
$$E[y_i] = \ \alpha + \beta \, E[x_i].$$

Then, once we estimate α & β, we test the CAPM for IBM, since according to the CAPM $\alpha = 0$. That is, we test: H_0: $\alpha = 0$ vs H_1: $\alpha \neq 0$.

Univariate Regression Equation

• To gain intuition and easy interpretation of the model, it is useful to think of x as a given or *predetermined* (realized before y) variable. Then, we can express the relation between y & x, in terms of the conditional expectation of y, conditioning on the predetermined value of x:

$$E[y_i \mid x_i] = \alpha + \beta\, x_i. \qquad (\text{"}Regression\ equation\text{"})$$

• The conditional expectation of y is what we model; in general, based on finance theory or the experience of the practitioner. To be technically precise, for the regression equation we require $E[\varepsilon_i \mid x_i] = 0$.

Note: We started with: $y_i = \alpha + \beta\, x_i + \varepsilon_i$

which can be converted to: $y_i = E[y_i \mid x_i] + \varepsilon_i$

That is, y_i is what we model plus something unexpected, a surprise.

Multivariate Regression Equation

• In the CAPM example, we have that IBM's excess returns are only related to ("explained by") the market. This is a **one variable model**.

But, we could have used more variables, for example the 3 factors in the standard Fama-French model: Market, SMB (size factor), and HML (book-to-market or value factor). This represents a **multivariate model** for IBM returns:

$$y_i = \alpha + \beta_1\, x_{1,i} + \beta_2\, x_{2,i} + \beta_3\, x_{3,i} + \varepsilon_i$$

• Though not necessary correct, we usually think of y as the *endogenous* variable and x as the *exogenous*, determined "outside" the model.

This lecture: Estimation of population parameters α & β and the properties of estimators.

Multivariate Regression Equation: Normality

<u>Technical Note</u>: We can study the joint distribution of y & x, $f(x, y)$ and describe the joint behavior in terms of expectations, conditional expectations, correlations, etc. For example, assuming joint normality for y & x, we can derive the conditional expectation of y, given x:

$$E[y_i| x_i] = \alpha + \beta\, x_i$$

which gives a functional (& linear) relation between y & x.

Thus, in a joint normality context, we can study the effect of a change in x on y. Moreover, after a lot of manipulations and applying statistical definitions, we get a formula to estimate α & β in terms of moments of y & x.

Q: Why do we need other methods to estimate α & β?

Multivariate Regression Equation: Normality

• Q: More specific to this lecture, why do we need a least squares estimation (regression analysis)?

A: Two things to consider:

1) Joint normality is not a realistic assumption in economics and finance.

2) In many situations, we think of the explanatory variable, x, as control, not necessarily as RV.

<u>Remark</u>: Without making any reference to a joint distribution, we will derive the formulas to estimate parameters in a linear relation.

Least Squares Estimation

Carl F. Gauss (1777 – 1855, Germany)

• Old method: Gauss (1795, 1801) used it in astronomy.

Idea:

• We relate a dependent variable y to a set of k explanatory variables $\boldsymbol{x}$. This function depends on unknown parameters, $\boldsymbol{\theta}$, which we want to estimate. The relation between y and $\boldsymbol{x}$ is not exact; there is an error, ε. We have T observations of y and $\boldsymbol{x}$.

$$y_i = f(x_{1,i}, x_{2,i}, \ldots, x_{k,i}; \boldsymbol{\theta}) + \varepsilon_i, \quad i = 1, 2, \ldots, T.$$

• If the functional form is known, we estimate the parameters $\boldsymbol{\theta}$ by **minimizing** a sum of squared errors:

$$\min_\theta \{ S(\boldsymbol{x}; \boldsymbol{\theta}) = \sum_i^T \varepsilon_i^2 = \sum_i^T (y_i - f(x_{1,i}, x_{2,i}, \ldots, x_{k,i}; \boldsymbol{\theta}))^2 \}$$

Ordinary Least Squares Estimation

• The estimator obtained is called the *Least Squares* (LS) estimator.

• LS is a general estimation method. It can be applied to almost any function.

• The functional form, $f(\boldsymbol{x}_i, \boldsymbol{\theta})$, is dictated by theory or experience. In this class, we mainly work with the **linear** case:

$$f(\boldsymbol{x}_i, \boldsymbol{\theta}) = \beta_1\ x_{1,i} + \beta_2\ x_{2,i} + \beta_3\ x_{3,i} + \ldots + \beta_k\ x_{k,i}.$$

• Now, we estimate the vector $\boldsymbol{\theta} = \{\beta_1, \beta_2, \ldots, \beta_k\}$ by minimizing

$$S(\boldsymbol{x}; \boldsymbol{\theta}) = \sum_i^T \varepsilon_i^2 = \sum_i^T (y_i - \beta_1\ x_{1,i} - \beta_2\ x_{2,i} - \cdots - \beta_k\ x_{k,i})^2$$

In this case, we call this estimator the *Ordinary Least Squares* (OLS) estimator. (Ordinary = Linear functional form.)

Least Squares Estimation – Notation

<u>Notation</u>: In lecture 2, we used ^ over the estimator of the parameter of interest. For example, $\hat{\theta}$ is the estimator of the parameter θ. Sometimes, to emphasize the method of estimation, we add to the estimated parameter the initials of the method used, say $\hat{\theta}_{LS}$.

• For historical reasons, in the linear model, **b** is popularly used to denote the OLS estimator of β.

LS Estimation – Example 1: One Variable Model

Example 1: We want to study the effect of the tech boom (x) on the San Francisco housing market (y). We rely on a simple linear model, with only one explanatory variable, the tech boom variable. That is,

$$y_i = \alpha + \beta \ x_i + \varepsilon_i.$$

In this model, we are interested in estimating β, our parameter of interest. β measures the *marginal effect* of x on y. We can use the estimate of β to check if the tech boom has a positive effect on SF housing prices. In this case, we test:
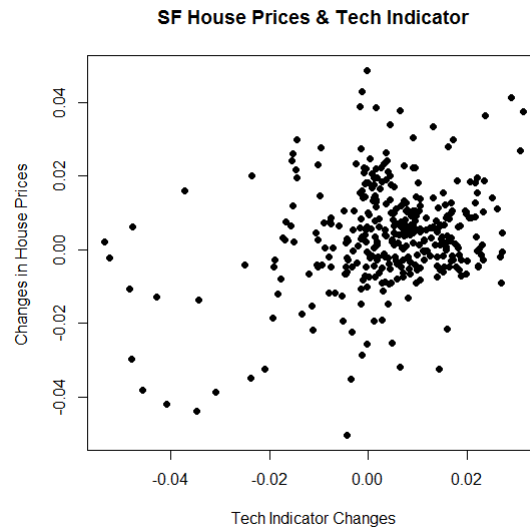
H$_0$ (No or Negative effect): $\beta \leq 0$.
H$_1$ (Positive effect): $\beta > 0$.

We have monthly data on SF Housing Prices and a Tech Indicator, developed by the Federal Reserve. We transform the data in percentage changes.
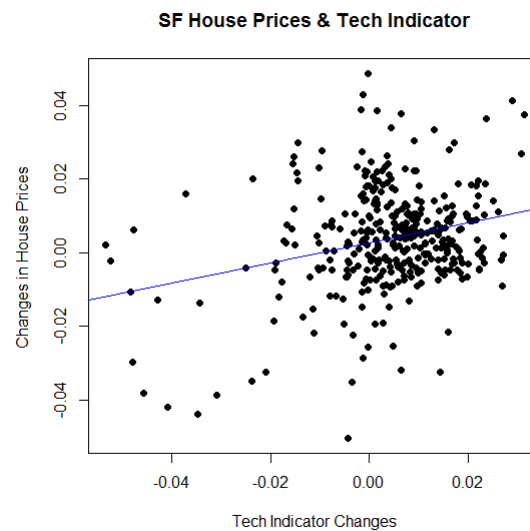
LS Estimation – Example 1: One Variable Model

Plot data: SF House Prices vs Tech Indicator (both in % changes).

LS Estimation – Example 1: One Variable Model

It looks like $\beta > 0$.

LS Estimation – Example 2: Multivariate Model

Example 2: We want to study the effect of a CEO's education (x) on a firm's CEO's compensation (*y)*. We build a CEO's compensation model including a CEO's education (x) and other "*control variables*" (**W**: experience, gender, etc.), controlling for other features that make one CEO's compensation different from another. That is,

$$y_i = f(x_i, W_i, \theta) + \varepsilon_i, \qquad i = 1, 2, ..., T.$$

The term ε_i represents the effects of individual variation that have not been controlled for with W_i or x_i and θ is a vector of parameters.

Usually, $f(x, \theta)$ is linear. Then, the compensation model becomes:

$$y_i = \alpha + \beta\ x_i + \gamma_1\ W_{1,i} + \gamma_2\ W_{2,i} + ... + \varepsilon_i$$

Again, in this model, we are interested in estimating β, our parameter of interest, which measures the effect of a CEO's education on a CEO's compensation.

OLS Estimation – One Variable: Derivation

• LS estimation can be applied to any functional form. In this class, we use a **linear function**. In this lecture, we derive the OLS formulas for the simplest case: **one explanatory variable**. Then,

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i \qquad (\text{two parameters} \Rightarrow k = 2)$$

Objective function:

$$S(x; \beta_1, \beta_2) = \sum_i^T \varepsilon_i^2 = \sum_i^T (y_i - \beta_1 - \beta_2 x_i)^2 =$$
$$= \{y_1 - \beta_1 - \beta_2 x_1)^2 + \{y_2 - \beta_1 - \beta_2 x_2)^2 + ... + \{y_T - \beta_1 - \beta_2 x_T)^2$$

Taking first derivatives with respect to β_1 & β_2 :
(β_1): $2 \sum_i^T (y_i - \beta_1 - \beta_2 x_i) (-1) \qquad \Rightarrow -2 \sum_i^T (y_i - \beta_1 - \beta_2 x_i)$
(β_2): $2 \sum_i^T (y_i - \beta_1 - \beta_2 x_i) (-x_i) \qquad \Rightarrow -2 \sum_i^T (y_i x_i - \beta_1 x_i - \beta_2 x_i^2)$

OLS Estimation – One Variable: Derivation

• Now, we set f.o.c.'s

(β_1): $\sum_i^T (y_i - b_1 - b_2\, x_i) = 0$ 　　　　　　(1)

(β_2): $\sum_i^T (y_i\, x_i - b_1\, x_i - b_2\, x_i^2) = 0$ 　　　(2)

Since we have $k = 2$, the f.o.c.'s form a 2x2 system of equations, the *normal equations.*

• Next, we solve for b_1 & b_2, the OLS estimators.

From (1):　$\sum_i^T y_i - \sum_i^T b_1 - b_2 \sum_i^T x_i = 0$

　　　　　$\Rightarrow b_1 = \bar{y} - b_2\, \bar{x}$

From (2):　$\sum_i^T y_i x_i - (\bar{y} - b_2\, \bar{x}) \sum_i^T x_i - b_2 \sum_i^T x_i^2 = 0 = 0$

　　　　　$\Rightarrow \sum_i^T y_i x_i - \bar{y} \sum_i^T x_i - b_2\, (\sum_i^T x_i^2 - \bar{x} \sum_i^T x_i) = 0$

　　　　　$\Rightarrow b_2 = \dfrac{\sum_i (y_i - \bar{y})\, x_i}{\sum_i (x_i - \bar{x})\, x_i}$

OLS Estimation – One Variable: Derivation

• The estimator of β_2:

$$b_2 = \frac{\sum_i^T (y_i - \bar{y})\, x_i}{\sum_i^T (x_i - \bar{x})\, x_i}$$

after some algebra can be written as:

$$b_2 = \frac{\sum_i^T (y_i - \bar{y})(x_i - \bar{x})}{\sum_i^T (x_i - \bar{x})^2} = \frac{\sum_i^T (y_i - \bar{y})(x_i - \bar{x})/(T-1)}{\sum_i^T (x_i - \bar{x})^2/(T-1)}$$

$$= \frac{cov(y_i, x_i)}{var(x_i)}$$

<u>Note:</u> We need $var(x_i) \neq 0$ to get b_2.

• The estimators b_1 & b_2 are a function of the data. Moreover, if we think of x as predetermined, b_1 & b_2 are a linear function of y. This is a property of the OLS estimators.

OLS Estimation: Interpretation & Prediction

• Interpretation of coefficients

- b_1 estimates the *constant* of the regression, the value of y_i, when x_i equals to 0.

- b_2 estimates the *slope* of the regression, the marginal effect –i.e., the first derivative of y_i with respect to x_i:

$$\frac{\delta y_i}{\delta x_i} = \beta_2 \qquad \Rightarrow \text{estimated by } b_2.$$

That is, if x increases by one unit (say, 1%), then, we estimate that y increases by b_2 units (say, b_2%).

• Conditional Prediction

Suppose analysts estimate that x_i will be **z%**, then, you estimate (or predict, given the **z%** value of x_i):

$$\text{Predicted } [y_i \mid x_i = \textbf{z\%}] = b_1 + b_2 * \textbf{z}.$$

LS Estimation – General Case: f.o.c.

• General functional form:

$$f(x_i, \boldsymbol{\theta}) \qquad \text{-}\boldsymbol{\theta} \text{ is a vector of } k \text{ parameters.}$$

• The model:

$$y_i = f(x_i, \boldsymbol{\theta}) + \varepsilon_i$$

• Objective function:

$$S(\boldsymbol{x}; \theta) = \sum_i^T \varepsilon_i^2 = \sum_i^T \{y_i - f(x_i, \boldsymbol{\theta})\}^2$$
$$= \{y_1 - f(x_1, \boldsymbol{\theta})\}^2 + \{y_2 - f(x_2, \boldsymbol{\theta})\}^2 + \dots + \{y_T - f(x_T, \boldsymbol{\theta})\}^2$$

• We minimize $S(\boldsymbol{x}, \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$:

$$\frac{\partial S(\boldsymbol{x}, \theta)}{\partial \theta} = 2 \{y_1 - f(x_1, \boldsymbol{\theta})\}(-f'(x_1, \boldsymbol{\theta})) + \dots + 2 \{y_T - f(x_T, \boldsymbol{\theta})\}(-f'(x_T, \boldsymbol{\theta}))$$
$$= -2 \sum_i^T \{y_i - f(x_i, \boldsymbol{\theta})\} f'(x_i, \boldsymbol{\theta})\}$$

LS Estimation – General Case: OLS & NLLS

• We minimize S($\boldsymbol{x}$, $\boldsymbol{\theta}$) with respect to $\boldsymbol{\theta}$.

$$\frac{\partial S(\boldsymbol{x},\, \theta)}{\partial \theta} = -2 \sum_i^T \{y_i - f(x_i, \theta)\}\, f'(x_i, \theta)\}$$

• We set the f.o.c.'s:

$$-2 \sum_i^T \{y_i - f(x_i, \hat{\theta}_{LS})\}\, f'(x_i, \hat{\theta}_{LS})\} = 0$$
$$\sum_i^T \{y_i - f(x_i, \hat{\theta}_{LS})\}\, f'(x_i, \hat{\theta}_{LS})\} = 0 \qquad \textit{(normal equations)}$$

• The *normal equations* (a $k \times k$ system) do not always have an analytic solution. When $f(x_i, \boldsymbol{\theta})$ is linear, we get an explicit solution, $\hat{\theta}_{OLS} = \mathbf{b}$.

• When $f(x_i, \boldsymbol{\theta})$ is *non-linear*, we **do not** have an explicit solution for $\hat{\theta}_{LS}$. The system can be solved numerically. In this case, the estimator is usually referred as *Non-linear Least Squares* estimator, $\hat{\theta}_{NLLS}$.

OLS Estimation – One Variable: CAPM

• A typical finance application of a one variable linear model is the CAPM. Recall that the (Sharpe-Litner) CAPM, in equilibrium, implies:

$$E[r_{i,t} - r_f] = \beta_i\, E[(r_{m,t} - r_f)],$$

where

$r_{i,t}$ = return on asset i at time t.

r_f = return of riskless asset at time t.

$r_{m,t}$ = return on the market portfolio at time t.

β_i = asset i's sensitivity to market (systematic) risk.

That is, the expected excess return on asset i is proportional to the market risk premium. β_i is the proportionality factor, it measures the sensitivity to market (systematic) risk.

OLS Estimation – One Variable: CAPM

• The CAPM is a particular case of what in financial theory we call *"factor models."* Factors represent the systematic component that drives the cross-section of returns over time; they can be observed or unobserved. For example, a *k-factor model* for returns is given by:

$$r_{i,t} - r_f = \alpha + \beta_1 f_{1,t} + \beta_2 f_{2,t} + \dots + \beta_k f_{k,t} + \varepsilon_{i,t}$$

where $f_{j,t}$ is the j (common) factor at time t, and constant over i, and $\varepsilon_{i,t}$ represents the idiosyncratic component of asset i.

• We think of returns as driven by common factors (undiversifiable) & idiosyncratic factors (diversifiable in large portfolios.) Thus, investors get compensated only for the systematic risk they take. The higher the exposure –i.e., β_i– the higher the expected compensation.

• The CAPM has only one factor: market excess returns (*"the market"*).

OLS Estimation – One Variable: CAPM

• A linear data generating process (DGP) consistent with the CAPM:

$$r_{i,t} - r_f = \alpha_i + \beta_i(r_{m,t} - r_f) + \varepsilon_{i,t}, \qquad i = 1, ..., N \ \& \ t = 1, …, T$$

α_i & β_i are the coefficients to be estimated by LS, and

$$\text{Cov}(r_{m,t}, \varepsilon_{i,t}) = 0 \qquad \text{-i.e., market returns are } exogenous.$$

If $\beta_i = 0$, asset i is not exposed to market risk. Thus, the investor is not compensated with a higher return than r_f.

If $\beta_i > 0$, asset i is exposed to market risk & $r_{i,t} \geq r_f$, provided that $E[r_{m,t} - r_f] > 0$.

If $\beta_i > 1$ ($\beta_i < 1$), asset i is *"riskier"* (*"safer"*) than the market.

If $\alpha_i > 0$, then asset i has higher expected returns than what is expected in equilibrium –i.e., what the CAPM implies.

OLS Estimation – One Variable: CAPM

• Then, in our linear model let y_i represent IBM excess returns $(r_{i=IBM,t} - r_f)$ at time t and let x_i represent Market excess returns $(r_{m,t} - r_f)$ at time t. Then, b_2 estimates IBM's beta in the CAPM.

Then, b_2 estimates the stock's beta in the CAPM:

$$b_2 = \hat{\beta}_{IBM} = \frac{cov(r_{i=IBM,t} - r_f, r_{m,t} - r_f)}{var(r_{m,t} - r_f)}$$

That is, the CAPM β is the ratio of a covariance over a variance.

Recall that the CAPM β measures a stock's risk in relation to the risk (volatility) of the market. We think of β as a measure of the relative risk exposure of holding a particular stock (IBM, in this case) in relation to the market.

OLS Estimation – One Variable: CAPM

• Interpretation of coefficients

- b_1 estimates the *constant* of the regression: IBM excess returns in excess of Market excess returns. In the CAPM, it should be 0 (= α_i).

- b_2 estimates the *slope* of the regression. In the CAPM: β_i

$$\frac{\delta y_i}{\delta x_i} = \beta_i$$

That is, if Market excess returns increase by one 1%, then IBM excess returns are expected to increase by b_2 (= β_i) units (say, b_2%). The β_{IBM} also tells us if IBM is riskier ($\beta_{IBM} > 1$) or safer ($\beta_{IBM} < 1$) than the market.

OLS Estimation – One Variable: CAPM

• Conditional Prediction

Suppose analysts estimate that Market excess returns are **10%,** then, we estimate (or predict, given the **10%** value for Market excess returns):

$$\text{Predicted } [\text{IBM excess returns} | (r_{m,t} - r_f) = .10] = b_1 + b_2 * .10.$$

We will call the Predicted $y_i = \hat{y}_i$ = fitted value.

OLS Estimation: CAPM – R Estimation

Example: Estimate the CAPM for **IBM returns** using *lm* function.

• Import data with **read** function

SFX_da <- read.csv("http://www.bauer.uh.edu/rsusmel/4397/Stocks_FX_1973.csv", head=TRUE, sep=",")

• Extract variables from imported data

x_ibm <- SFX_da$IBM	# extract IBM price data
x_Mkt_RF <- SFX_da$Mkt_RF	# extract Market excess returns (in %)
x_RF <- SFX_da$RF	# extract Risk-free rate (in %)

• Define log returns & adjust size of variables accordingly

T <- length(x_ibm)	# sample size
lr_ibm <- log(x_ibm[-1]/x_ibm[-T])	# create IBM log returns (in decimal returns)
Mkt_RF <- x_Mkt_RF[-1]/100	# Adjust sample size to (T-1) by removing 1st obs
RF <- x_RF[-1]/100	# Adjust sample size and use decimal returns.

OLS Estimation: CAPM – R Estimation

Example (continuation):

• Define excess returns and estimate CAPM with **lm** function:

ibm_x <- lr_ibm – RF	# **IBM excess returns**
fit_ibm_capm <- lm(ibm_x ~ Mkt_RF)	# lm (=*linear model*) package in R
summary(**fit_ibm_capm**)	# print lm results

> summary(**fit_ibm_capm**)

Call:

lm(formula = ibm_x ~ Mkt_RF)

Residuals:

Min	1Q	Median	3Q	Max
-0.314401	-0.031692	-0.000537	0.031447	0.248201

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	**-0.005791**	0.002487	-2.329	0.0202	*
xMkt_RF	**0.895774**	0.053867	16.629	<2e-16	***

OLS Estimation: CAPM – R Estimation

Example (continuation):

> summary(**fit_ibm_capm**)

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)		
						b_1
(Intercept)	**-0.005791**	0.002487	-2.329	0.0202	*	
xMkt_RF	**0.895774**	0.053867	16.629	<2e-16	***	
---						$b_2 = \widehat{\beta}_{CAPM}$

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05887 on 567 degrees of freedom
Multiple R-squared: 0.3278,   Adjusted R-squared: 0.3266
F-statistic: 276.5 on 1 and 567 DF, p-value: < 2.2e-16

<u>Interpretation of b_1:</u>

b_1 = constant. The additional IBM return, after excess market returns are incorporated, is **-0.58%**. Under the CAPM, b_1 should be close to 0.

OLS Estimation: CAPM – R Estimation

Example (continuation):

Interpretation of b_2:

b_2 = slope. If market excess returns increase by 1%, IBM excess returns increase by **0.90%**. The estimate of β_{IBM} ($\beta < 1$) implies that IBM is less volatile ("safer") than the market.

Conditional prediction of IBM excess returns:

Suppose market excess returns increase are 10%, then we predict IBM excess returns = **-0.005791** + **0.895774** * .10 = 0.08378 (**8.38%**).

Note: According to the CAPM, IBM underperformed:

- IBM excess returns (CAPM) = **0.895774** * mean(Mkt_RF)

$\qquad\qquad\qquad\qquad\qquad\qquad$ = **0.895774** * 0.0056489 = **0.0050601**

- IBM excess returns (sample) = mean(ibm_x) = **-0.00073141**

OLS Estimation – Multivariate Case

• LS is a general estimation method. It allows any functional form for the relation between y_i and x_i. But, in this lecture, we cover the case where $f(x_i, \theta)$ is **linear**. When the relation is linear, we do OLS estimation.

We assume a linear system with k independent variables and T observations. That is,

$$y_i = \beta_1\, x_{1,i} + \beta_2\, x_{2,i} + \dots + \beta_k\, x_{k,i} + \varepsilon_i, \qquad i = 1, 2, \dots, T$$

The whole system (for all i) is:

$$y_1 = \beta_1\, x_{11} + \beta_2\, x_{12} + \dots + \beta_k\, x_{k1} + \varepsilon_1$$
$$y_2 = \beta_1\, x_{12} + \beta_2\, x_{22} + \dots + \beta_k\, x_{k2} + \varepsilon_2$$
$$\dots \qquad \dots \qquad \dots \qquad \dots$$
$$y_T = \beta_1\, x_{1T} + \beta_2\, x_{2T} + \dots + \beta_k\, x_{kT} + \varepsilon_T$$

OLS Estimation – Multivariate Case

• It is cumbersome to write the whole system. Using linear algebra, we can rewrite the system in a more compact and simplify derivations.

Example: Using vector & matrix notation, we write the system as:
$$\mathbf{y} = f(\mathbf{X}, \theta) + \boldsymbol{\varepsilon} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Notation: $\mathbf{y}$, $\boldsymbol{\beta}$ & $\boldsymbol{\varepsilon}$ are vectors:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_T \end{bmatrix}, \quad \& \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$$

$\mathbf{X}$ is a matrix:
$$X = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{k1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1T} & x_{2T} & \cdots & x_{kT} \end{bmatrix}$$

Linear Algebra: Brief Review – Matrix

• Life (& notation) becomes easier with linear Algebra. Concepts:

• A Matrix.
A matrix is a set of elements, organized into rows and columns

columns

rows $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$

• a and d are the diagonal elements.
• b and c are the off-diagonal elements.

• Matrices are like plain numbers in many ways: they can be added, subtracted, and, in some cases, multiplied and inverted (divided).

27

Linear Algebra: Matrices and Vectors

Examples:

$$A = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix}; \quad b = \begin{bmatrix} b_1 & b_2 & b_3 \end{bmatrix}$$

• Dimensions of a matrix: numbers of rows by numbers of columns. The Matrix **A** is a 2x2 matrix, **b** is a 1x3 matrix.

• A matrix with only 1 column or only 1 row is called a *vector*.

• If a matrix has an equal numbers of rows and columns, it is called a *square* matrix. Matrix **A**, above, is a square matrix.

• <u>Usual Notation</u>: Upper case letters $\Rightarrow$ matrices
Lower case $\Rightarrow$ vectors

55

Linear Algebra: Matrices – Information

• Information is described by data. A tool to organize the data is a list, which we call a vector. Lists of lists are called matrices. That is, we organize the data using matrices, say, **X**.

• We think of the elements of **X** as data points ("data entries", "observations"), in economics, we usually have numerical data.

• We store the data in rows. In a Txk matrix, **X**, over time we build a database:

$$X = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{k1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1T} & x_{2T} & \cdots & x_{kT} \end{bmatrix} \quad \begin{array}{l} \text{row 1} = k \text{ entries at time 1} \\ \\ \text{row } T = k \text{ entries at time T} \end{array}$$

• Once the data is organized in matrices it can be easily manipulated: multiplied, added, etc. (This is what Excel does very well).

56

Linear Algebra: Matrices in Econometrics

• We want to estimate a model: $y = f(x_1, x_2, ..., x_k)$. We collect data, T (or N) observations, on a dependent variable, $\mathbf{y}$, and on k explanatory variables, $\mathbf{X}$.

• Usual notation: vectors are column vectors: $\mathbf{y}$ & $\mathbf{x}_j$ are Tx1 vectors:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} \qquad \& \qquad \mathbf{x}_j = \begin{bmatrix} x_{j1} \\ x_{j2} \\ \vdots \\ x_{jT} \end{bmatrix} \qquad j = 1, ..., k$$

$\mathbf{X}$ is a Txk matrix:
$$X = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{k1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1T} & x_{2T} & \cdots & x_{kT} \end{bmatrix}$$

Its columns are the k Tx1 vectors $\mathbf{x}_j$. It is common to treat $\mathbf{x}_1$ as vector of ones, $\mathbf{i}$.

57

Linear Algebra: Matrices in Econometrics

• In general, we import matrices (information) to our programs.

Example: In R, we use the **read** function, usually followed by the type of data we are importing. Below, we import a comma separated values (csv) file with monthly CPIs and exchange rates for 20 different countries, then we use the **read.csv** function:

PPP_da <-
read.csv("http://www.bauer.uh.edu/rsusmel/4397/ppp_2020_m.csv",head=TRUE
,sep=",")

The **names()** function describes the headers of the file imported (41 headers):
> names(PPP_da)
 [1] "Date"   "BG_CPI"   "IT_CPI"   "GER_CPI" "UK_CPI"
 [6] "SWED_CPI" "DEN_CPI" "NOR_CPI" "IND_CPI" "JAP_CPI"
[11] "KOR_CPI" "THAI_CPI" "SING_CPI" "MAL_CPI" "KUW_CPI"
[16] "SUAD_CPI" "CAN_CPI" "MEX_CPI" "US_CPI"   "EGY_CPI"
[...]

58

Linear Algebra: Matrices in Econometrics

Example (continuation):

The **summary()** function provides some stats of variables imported:

```
>summary(PPP_da)
      Date       BG_CPI         IT_CPI        GER_CPI
 1/15/1971: 1   Min.  : 19.77  Min.  :  5.90  Min.  : 31.20
 1/15/1972: 1   1st Qu.: 49.32  1st Qu.: 32.25  1st Qu.: 57.17
 1/15/1973: 1   Median : 69.91  Median : 67.30  Median : 75.30
 1/15/1974: 1   Mean  : 67.92  Mean  : 60.14  Mean  : 72.29
 1/15/1975: 1   3rd Qu.: 89.40  3rd Qu.: 89.65  3rd Qu.: 91.17
 1/15/1976: 1   Max.  :109.71  Max.  :103.50  Max.  :106.60
 (Other)  :588
```

We extract a variable from the matrix by the name of file followed by $ and the header of variable:

```
x_chf <- PPP_da$CHF_USD                 # extract CHF/USD exchange rate data
```

We can transform the vector x_chf. For example, for % changes:

```
T <- length(x_chf)                      # length of CHF/USD exchange rate data
lr_chf <- log(x_chf[-1]/x_chf[-T])      # create log returns (changes) for the CHF/USD
```

59