# Lecture 1:

# Basic Regression Analysis with Time Series Data



ISBN 9781408093757

# The nature of time series data

- Temporal ordering of observations; may not be arbitrarily reordered
- Typical features: serial correlation (tương quan chuỗi)/nonindependence of observations
- How should we think about the randomness in time series data?
  - The outcome of economic variables (e.g. GNP) is uncertain; they should therefore be modelled as random variables
  - Time series are sequences of r.v. (= stochastic processes)
  - Randomness does not come from sampling from a population
  - "Sample" = the one realized path of the time series out of the many possible paths the stochastic process could have taken


ISBN 9781408093757

# Remarks

- For time series data, there is a natural ordering; the observations are ordered according to time. More specifically, the time series data is a sequence of observations taken over time.
- Although time is continuous, the interval between observations is discrete. Usually the time interval of the observations is equally spaced. The interval can be yearly, quarterly, monthly, daily, hourly, etc.
- Example:
  - (a) Gross Domestic Product (GDP) : quarterly or yearly data
  - (b) Inflation rate or unemployment rate: monthly, quarterly, yearly data
  - (c) Stock Index, exchange rate: daily, weekly, monthly, quarterly, yearly data
  - (d) Corporate profits, dividends, etc: quarterly, yearly.
  - (e) Temperature:  hourly data.



ISBN 9781408093757

# Remarks

- (i) The frequency of the data needs to be consistent for all variables when we run regressions.
- (ii) To sum up 4 quarter data for any given years = yearly data
- (iii) Similarly, if you have daily data, we can convert the frequency of the data.
  - For instance, we have daily exchange rate data, and we can convert it to be monthly data.
- (iv) Suppose you have quarterly data. Can you change frequency of the data to monthly?
  - (cubic spline) interpolation (*but would violate autocorrelation*).


ISBN 9781408093757

# Example

- **Example: US inflation and unemployment rates 1948-2003**

**TABLE 10.1** Partial Listing of Data on U.S. Inflation and Unemployment Rates, 1948–2003

| Year | Inflation | Unemployment |
|---|---|---|
| 1948 | 8.1 | 3.8 |
| 1949 | −1.2 | 5.9 |
| 1950 | 1.3 | 5.3 |
| 1951 | 7.9 | 3.3 |
| . | . | . |
| . | . | . |
| . | . | . |
| 1998 | 1.6 | 4.5 |
| 1999 | 2.2 | 4.2 |
| 2000 | 3.4 | 4.0 |
| 2001 | 2.8 | 4.7 |
| 2002 | 1.6 | 5.8 |
| 2003 | 2.3 | 6.0 |



Here, there are only two time series. There may be many more variables whose paths over time are observed simultaneously.

Time series analysis focuses on modelling the dependency of a variable on its own past, and on the present and past values of other variables.



ISBN 9781408093757

# Examples of time series regression models

- **Static models (mô hình tĩnh)**
  - In static time series models, the current value of one variable is modelled as the result of the current values of explanatory variables
- **Examples for static models**

There is a contemporaneous relationship between unemployment and inflation (= Phillips-Curve).

$$inf_t = \beta_0 + \beta_1 unem_t + u_t$$

$$mrdrte_t = \beta_0 + \beta_1 convrte_t + \beta_2 unem_t + \beta_3 yngmle_t + u_t$$

The current murder rate (per 10,000) is determined by the current conviction rate, unemployment rate, and fraction of young males (18-25) in the population.


ISBN 9781408093757

# Examples of time series regression models (cont.)

- **Finite distributed lag models (mô hình phân phối trễ hữu hạn)**
  - In finite distributed lag models, the explanatory variables are allowed to influence the dependent variable with a time lag
- **Example for a finite distributed lag model**
  - The fertility rate may depend on the tax value of a child, but for biological and behavioural reasons, the effect may have a lag

$$gfr_t = \alpha_0 + \delta_0 \boxed{pe_t} + \delta_1 \boxed{pe_{t-1}} + \delta_2 \boxed{pe_{t-2}} + u_t$$

Children born per 1,000 women in year t

Tax exemption in year t

Tax exemption in year t-1

Tax exemption in year t-2



ISBN 9781408093757

# Examples of time series regression models (cont.)

- **Interpretation of the effects in finite distributed lag models**

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \cdots + \delta_q z_{t-q} + u_t$$

- **Effect of a past shock on the current value of the dep. variable**

$$\boxed{\frac{\partial y_t}{\partial z_{t-s}} = \delta_s} \qquad \boxed{\frac{\partial y_t}{\partial z_{t-q}} + \cdots + \frac{\partial y_t}{\partial z_t} = \delta_1 + \cdots + \delta_q}$$

Effect of a transitory shock (cú sốc tạm thời):
If there is a one time shock in a past period, the dep. variable will change temporarily by the amount indicated by the coefficient of the corresponding lag.

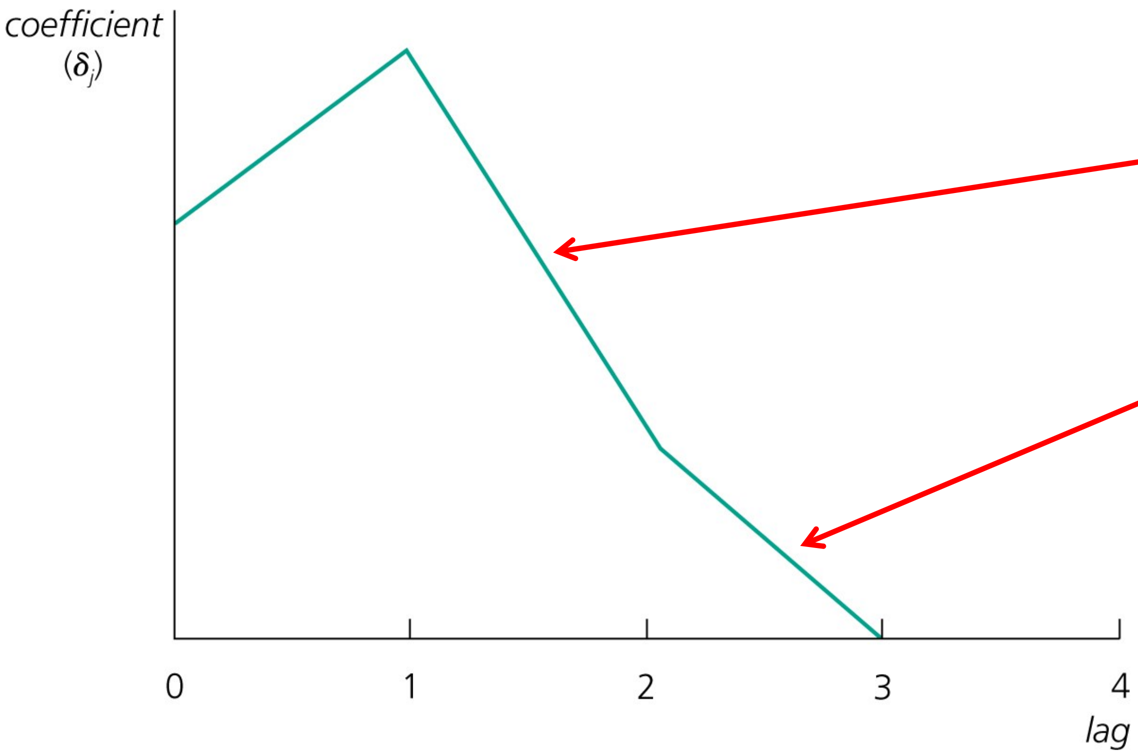
Effect of permanent shock (cú sốc vĩnh viễn):
If there is a permanent shock in a past period, i.e. the explanatory variable permanently increases by one unit, the effect on the dep. variable will be the cumulated effect of all relevant lags. This is a long-run effect on the dependent variable.



ISBN 9781408093757

# Examples of time series regression models (cont.)

- **Graphical illustration of lagged effects (ảnh hưởng có độ trễ)**




For example, the effect is biggest after a lag of one period. After that, the effect vanishes (if the initial shock was transitory).

The long run effect of a permanent shock is the cumulated effect of all relevant lagged effects. It does not vanish (if the initial shock is a permanent one).



ISBN 9781408093757

# Some differences between cross section and time series

| | **Cross-Sectional Data** | **Time Series Data** |
|---|---|---|
| Time | Time dimension is not involved. | Time dimension is important. |
| Subscript | i = 1,2,3…n<br>i indexes different persons, workers, firms, cities, etc. | t = 1,2,3…n<br>t indexes different time periods. |
| Correlations (tương quan) | Usually no correlation between observations. Therefore, we can randomly arrange the order of the observations. | Observations of a variable are usually auto-correlated. We cannot rearrange the order. Otherwise the structure is destroyed. |
| Outcome of the samples | Random sample | Random outcome |
| | A collected sample | A realization |

ISBN 9781408093757

# Quiz: Which type of data that corresponds to each of the following statements?

- Data on daily sales volume, revenue, number of customers for the past month at each Highlands Coffee location in Ho Chi Minh City.
- Data on daily sales revenue and expenses over past 12 months at Crescent Mall Highlands Coffee location.
- Data on daily sales volume, revenue, number of customers for the past month at all Highlands Coffee locations in Ho Chi Minh City.
- Data on 2019 Christmas day sales revenue and expenses in all Highlands Coffee locations in Vietnam.



ISBN 9781408093757

# Quiz: Which type of data that corresponds to each of the following statements?

- Data on daily sales volume, revenue, number of customers for the past month at each Highlands Coffee location in Ho Chi Minh City. → panel
- Data on daily sales revenue and expenses over past 12 months at Crescent Mall Highlands Coffee location. → time series
- Data on daily sales volume, revenue, number of customers for the past month at all Highlands Coffee locations in Ho Chi Minh City. → time series
- Data on 2019 Christmas day sales revenue and expenses in all Highlands Coffee locations in Vietnam. → cross section



ISBN 9781408093757

# Finite sample properties of OLS under classical assumptions

- **Assumption TS.1 (Linear in parameters) (tuyến tính theo tham số)**

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \ldots + \beta_k x_{tk} + u_t$$

The time series involved obey a linear relationship. The stochastic processes $y_t$, $x_{t1}, \ldots,$ $x_{tk}$ are observed, the error process $u_t$ is unobserved. The definition of the explanatory variables is general, e.g. they may be lags or functions of other explanatory variables.

- **Assumption TS.2 (No perfect collinearity) (không có đa cộng tuyến hoàn hảo)**

"In the sample (and therefore in the underlying time series process), no independent variable is constant nor a perfect linear combination of the others."


ISBN 9781408093757

# Finite sample properties of OLS under classical assumptions (cont.)

- **Notation**

$$\mathrm{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ \vdots & \vdots & & \vdots \\ x_{t1} & x_{t2} & \cdots & x_{tk} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}$$

This matrix collects all the information on the complete time paths of all explanatory variables

The values of all explanatory variables in period number t

- **Assumption TS.3 (Zero conditional mean) (trung bình có điều kiện bằng 0)**

$$E(u_t|\mathrm{X}) = 0$$

The mean value of the unobserved factors is unrelated to the values of the explanatory variables <u>in all periods</u>



ISBN 9781408093757

# Finite sample properties of OLS under classical assumptions (cont.)

- **Discussion of assumption TS.3**

Exogeneity (ngoại sinh cùng kỳ): $E(u_t|\mathbf{x}_t) = 0$ The mean of the error term is unrelated to the explanatory variables of the same period

Strict exogeneity (ngoại sinh nghiêm ngặt): $E(u_t|\mathbf{X}) = 0$ The mean of the error term is unrelated to the values of the explanatory variables of all periods

- **Strict exogeneity is stronger than contemporaneous exogeneity**
  - TS.3 rules out feedback from the dep. variable on future values of the explanatory variables; this is often questionable esp. if explanatory variables "adjust" to past changes in the dependent variable
  - If the error term is related to past values of the explanatory variables, one should include these values as contemporaneous regressors


ISBN 9781408093757

# Finite sample properties of OLS under classical assumptions (cont.)

- **<u>Theorem 10.1 (Unbiasedness of OLS) (tính không chệch của ước OLS)</u>**

$$TS.1 - TS.3 \quad \Rightarrow \quad E(\hat{\beta}_j) = \beta_j, \quad j = 0, 1, \ldots, k$$

- **Assumption TS.4 (Homoscedasticity) (phương sai không đổi)**

$$Var(u_t|\mathbf{X}) = Var(u_t) = \sigma^2$$

The volatility of the errors must not be related to the explanatory variables in any of the periods

  - A sufficient condition is that the volatility of the error is independent of the explanatory variables and that it is constant over time
  - In the time series context, homoscedasticity may also be easily violated, e.g. if the volatility of the dep. variable depends on regime changes



ISBN 9781408093757

# Finite sample properties of OLS under classical assumptions (cont.)

- **Assumption TS.5 (No serial correlation) (không có tương quan chuỗi)**

$$Corr(u_t, u_s|\mathbf{X}) = 0, \;\; t \neq s$$

Conditional on the explanatory variables, the un-observed factors must not be correlated over time

- **Discussion of assumption TS.5**
  - Why was such an assumption not made in the cross-sectional case?
  - The assumption may easily be violated if, conditional on knowing the values of the indep. variables, omitted factors are correlated over time
  - The assumption may also serve as substitute for the random sampling assumption if sampling a cross-section is not done completely randomly
  - In this case, given the values of the explanatory variables, errors have to be uncorrelated across cross-sectional units (e.g. areas)


ISBN 9781408093757

# Finite sample properties of OLS under classical assumptions (cont.)

- **Theorem 10.2 (OLS sampling variances) (phương sai mẫu của OLS)**

Under assumptions TS.1 – TS.5:

The same formula as in the cross-sectional case

$$Var(\hat{\beta}_j|\mathbf{X}) = \frac{\sigma^2}{SST_j(1-R_j^2)}, \quad j = 1, \ldots, k$$

SST: tổng bình phương toàn phần; R: hệ số xác định của hồi quy X

The conditioning on the values of the explanatory variables is not easy to understand. It effectively means that, in a finite sample, one ignores the sampling variability coming from the randomness of the regressors. This kind of sampling variability will normally not be large (because of the sums).

- **Theorem 10.3 (Unbiased estimation of the error variance) (ước lượng không thiên vị của phương sai sai số)**

$$TS.1 - TS.5 \quad \Rightarrow \quad E(\hat{\sigma}^2) = \sigma^2$$



ISBN 9781408093757

# Finite sample properties of OLS under classical assumptions (cont.)

- **<u>Theorem 10.4 (Gauss-Markov Theorem)</u>**
  - Under assumptions TS.1 – TS.5, the OLS estimators have the minimal variance of all linear unbiased estimators of the regression coefficients
  - This holds conditional as well as unconditional on the regressors
- **Assumption TS.6 (Normality) (phân phối chuẩn)**

This assumption implies TS.3 – TS.5

$$u_t \sim N(0, \sigma^2) \quad \text{independently of} \quad \mathbf{X}$$

- **<u>Theorem 10.5 (Normal sampling distributions)</u>**
  - Under assumptions TS.1 – TS.6, the OLS estimators have the usual normal distribution (conditional on $\mathbf{X}$). The usual F- and t-tests are valid.



ISBN 9781408093757

## Example 10.1

### [Static Phillips Curve]

To determine whether there is a tradeoff, on average, between unemployment and inflation, we can test $H_0$: $\beta_1 = 0$ against $H_1$: $\beta_1 < 0$ in equation (10.2). If the classical linear model assumptions hold, we can use the usual OLS $t$ statistic.

We use the file PHILLIPS.RAW to estimate equation (10.2), restricting ourselves to the data through 1996. (In later exercises, for example, Computer Exercises C10.12 and C11.10, you are asked to use all years through 2003. In Chapter 18, we use the years 1997 through 2003 in various forecasting exercises.) The simple regression estimates are

$$\widehat{inf}_t = \underset{(1.72)}{1.42} + \underset{(.289)}{.468}\, unem_t \qquad (10.14)$$
$$n = 49,\ R^2 = .053,\ \bar{R}^2 = .033.$$

This equation does not suggest a tradeoff between *unem* and *inf*: $\hat{\beta}_1 > 0$. The $t$ statistic for $\hat{\beta}_1$ is about 1.62, which gives a $p$-value against a two-sided alternative of about .11. Thus, if anything, there is a positive relationship between inflation and unemployment.

There are some problems with this analysis that we cannot address in detail now. In Chapter 12, we will see that the CLM assumptions do not hold. In addition, the static Phillips curve is probably not the best model for determining whether there is a short-run tradeoff between inflation and unemployment. Macroeconomists generally prefer the expectations augmented Phillips curve, a simple example of which is given in Chapter 11.

tsset year
reg inf unem



ISBN 9781408093757

# Finite sample properties of OLS under classical assumptions (cont.)

- **Example: Static Phillips curve**

$$\widehat{inf}_t = \underset{(1.72)}{1.42} + \underset{(.289)}{.468}\; unem_t$$

Contrary to theory, the estimated Phillips curve does not suggest a tradeoff between inflation and unemployment

$$n = 49, R^2 = .053, \bar{R}^2 = .033$$

- **Discussion of CLM assumptions**

The error term contains factors such as monetary shocks, income/demand shocks, oil price shocks, supply shocks, or exchange rate shocks

<u>TS.1:</u> $inf_t = \beta_0 + \beta_1 unem_t + u_t$

<u>TS.2:</u> A linear relationship might be restrictive, but it should be a good approximation. Perfect collinearity is not a problem as long as unemployment varies over time.



ISBN 9781408093757

# Finite sample properties of OLS under classical assumptions (cont.)

- **Discussion of CLM assumptions (cont.)**

<u>TS.3:</u> $E(u_t|unem_1,\ldots,unem_n)=0$ Easily violated

$unem_{t-1}\uparrow\ \rightarrow\ u_t\downarrow$ For example, past unemployment shocks may lead to future demand shocks which may dampen inflation

$u_{t-1}\uparrow\ \rightarrow\ unem_t\uparrow$ For example, an oil price shock means more inflation and may lead to future increases in unemployment

<u>TS.4:</u> $Var(u_t|unem_1,\ldots,unem_n)=\sigma^2$ Assumption is violated if monetary policy is more "nervous" in times of high unemployment

<u>TS.5:</u> $Corr(u_t,u_s|unem_1,\ldots,unem_n)=0$ Assumption is violated if exchange rate influences persist over time (they cannot be explained by unemployment)

<u>TS.6:</u> $u_t\sim N(0,\sigma^2)$ Questionable



ISBN 9781408093757

# Finite sample properties of OLS under classical assumptions (cont.)

- **Example: Effects of inflation and deficits on interest rates**

Interest rate on 3-months T-bill

Government deficit as percentage of GDP

$$\widehat{i3}_t = \underset{(0.43)}{1.73} + \underset{(.082)}{.606}\ inf_t + \underset{(.118)}{.513}\ def_t$$

$$n = 56, R^2 = .602, \bar{R}^2 = .587$$

The error term represents other factors that determine interest rates in general, e.g. business cycle effects

- **Discussion of CLM assumptions**

<u>TS.1:</u> $i3_t = \beta_0 + \beta_1 inf_t + \beta_2 def_t + u_t$

<u>TS.2:</u> A linear relationship might be restrictive, but it should be a good approximation. Perfect collinearity will seldomly be a problem in practice.



ISBN 9781408093757

# Finite sample properties of OLS under classical assumptions (cont.)

- **Discussion of CLM assumptions (cont.)**

TS.3: $E(u_t|inf_1,\ldots,inf_n,def_1,\ldots,def_n)=0$ ← Easily violated

$def_{t-1}\uparrow\;\rightarrow\;u_t\uparrow$ ← For example, past deficit spending may boost economic activity, which in turn may lead to general interest rate rises

$u_{t-1}\uparrow\;\rightarrow\;inf_t\uparrow$ ← For example, unobserved demand shocks may increase interest rates and lead to higher inflation in future periods

TS.4: $Var(u_t|inf_1,\ldots,def_n)=\sigma^2$ ← Assumption is violated if higher deficits lead to more uncertainty about finances and possibly more abrupt rate changes

TS.5: $Corr(u_t,u_s|inf_1,\ldots,def_n)=0$ ← Assumption is violated if business cylce effects persist across years (and they cannot be completely accounted for by inflation and the evolution of deficits)

TS.6: $u_t\sim N(0,\sigma^2)$ ← Questionable



ISBN 9781408093757

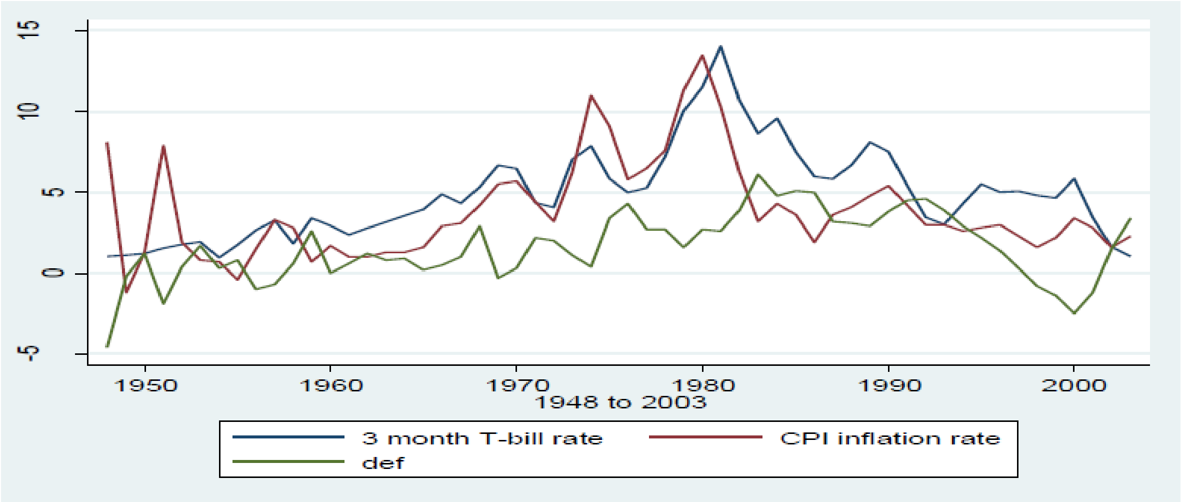
# Illustration by Stata

Empirical example:

(Effects of inflation and deficit on interest rates) The data in INTDEF.dta come from the 2004 Economic Report of the President and span the years 1948 through 2003. The variable $i3$ is the three-month T-bill rate, $inf$ is the annual inflation rate based on the consumer price index (CPI), and $def$ is the federal budget deficit as a percentage of GDP. The regression model is

$$i3_t = \beta_0 + \beta_1 inf_t + \beta_2 def_t + u_t, \text{ where } t = 1948, 1949, \ldots, 2003$$

>**tsline** i3 inf def






ISBN 9781408093757

# Illustration by Stata

```
. reg i3 inf def

      Source |       SS       df       MS              Number of obs =      56
-------------+------------------------------           F(  2,    53) =   40.09
       Model |  272.420338     2  136.210169           Prob > F      =  0.0000
    Residual |  180.054275    53  3.39725047           R-squared     =  0.6021
-------------+------------------------------           Adj R-squared =  0.5871
       Total |  452.474612    55  8.22681113           Root MSE      =  1.8432

------------------------------------------------------------------------------
          i3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         inf |   .6058659   .0821348     7.38   0.000     .4411243    .7706074
         def |   .5130579   .1183841     4.33   0.000     .2756095    .7505062
       _cons |   1.733266    .431967     4.01   0.000     .8668497    2.599682
------------------------------------------------------------------------------
```

These estimates show that increases in inflation and the relative size of the deficit work together to increase short-term interest rates, both of which are expected from basic economics. For example, a ceteris paribus one percentage point increase in the inflation rate increases *i3* by .605 points. Both *inf* and *def* are very statistically significant, assuming, of course, that the CLM assumptions hold.

Remark:

In practice, we should always test the ________ of the time series data before running a regression model. If the time series data is not _________, then we cannot run the regression model by using the raw data directly. Here for the sake of simplicity and illustration, we assume all variables are all _________ at this moment.



ISBN 9781408093757

## [Effects of Personal Exemption on Fertility Rates]

The general fertility rate (*gfr*) is the number of children born to every 1,000 women of childbearing age. For the years 1913 through 1984, the equation,

$$gfr_t = \beta_0 + \beta_1 pe_t + \beta_2 ww2_t + \beta_3 pill_t + u_t,$$

explains *gfr* in terms of the average real dollar value of the personal tax exemption (*pe*) and two binary variables. The variable *ww2* takes on the value unity during the years 1941 through 1945, when the United States was involved in World War II. The variable *pill* is unity from 1963 on, when the birth control pill was made available for contraception.

Using the data in FERTIL3.RAW, which were taken from the article by Whittington, Alm, and Peters (1990), gives

$$\begin{aligned} \widehat{gfr}_t &= \underset{(3.21)}{98.68} + \underset{(.030)}{.083}\, pe_t - \underset{(7.46)}{24.24}\, ww2_t - \underset{(4.08)}{31.59}\, pill_t \\ n &= 72,\ R^2 = .473,\ \bar{R}^2 = .450. \end{aligned} \qquad \boxed{10.18}$$

Each variable is statistically significant at the 1% level against a two-sided alternative. We see that the fertility rate was lower during World War II: given *pe*, there were about 24 fewer births for every 1,000 women of childbearing age, which is a large reduction. (From 1913 through 1984, *gfr* ranged from about 65 to 127.) Similarly, the fertility rate has been substantially lower since the introduction of the birth control pill.

```
tsset year
reg gfr pe ww2 pill
```



ISBN 9781408093757

# Using dummy explanatory variables in time series

Children born per 1,000 women in year t

Tax exemption in year t

Dummy for World War II years (1941–45)

Dummy for availabity of contraceptive pill (1963–present)

$$\widehat{gfr}_t = \underset{(3.68)}{98.68} + \underset{(.030)}{.083}\ pe_t - \underset{(7.46)}{24.24}\ \boxed{ww2_t} - \underset{(4.08)}{31.59}\ \boxed{pill_t}$$

$$n = 72, R^2 = .473, \bar{R}^2 = .450$$

- **Interpretation**
  - During World War II, the fertility rate was temporarily lower
  - It has been permanently lower since the introduction of the pill in 1963



ISBN 9781408093757

# Example: Election Outcomes and Economic Performance

- Fair (1996) summarizes his work on explaining presidential election outcomes in terms of economic performance. He explains the proportion of the two-party vote going to the Democratic candidate using data for the years 1916 through 1992 (every four years) for a total of 20 observations.
- We estimate a simplified version of Fair's model (using variable names that are more descriptive than his):

$$
\begin{aligned}
demvote = \beta_0 + \beta_1 partyWH + \beta_2 incum + \beta_3 partyWH \cdot gnews \\
+ \beta_4 partyWH \cdot inf + u,
\end{aligned}
$$

- where *demvote* is the proportion of the two-party vote going to the Democratic candidate.



ISBN 9781408093757

- *partyWH* is similar to a dummy variable, but it takes on the value 1 if a Democrat is in the White House and -1 if a Republican is in the White House.
- *incum* is defined to be 1 if a Democratic incumbent is running, -1 if a Republican incumbent is running, and zero otherwise.
- *gnews* is the number of quarters, during the current administration's first 15 quarters, where the quarterly growth in real per capita output was above 2.9% (at an annual rate), and
- *inf* is the average annual inflation rate over the first 15 quarters of the administration.



ISBN 9781408093757

# Estimation results:

The estimated equation using the data in FAIR.RAW is

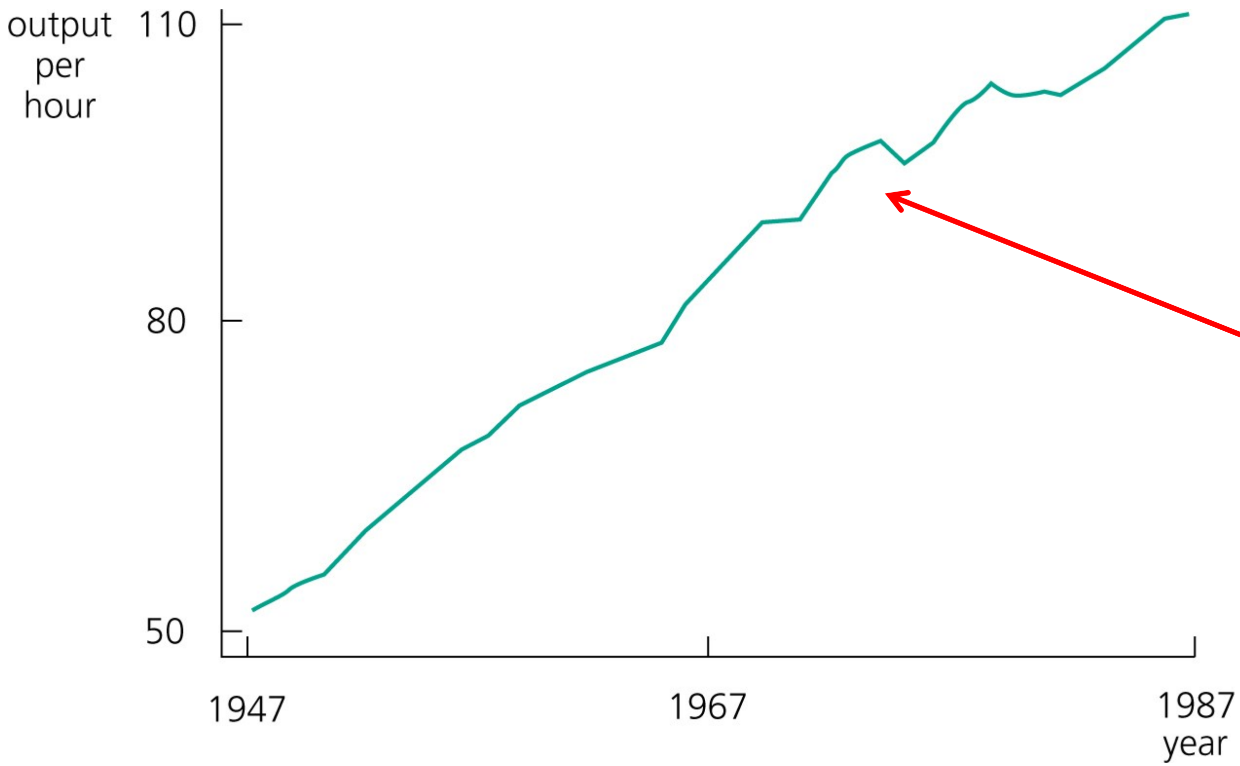
$$\widehat{demvote} = \underset{(.012)}{.481} - \underset{(.0405)}{.0435}\, partyWH + \underset{(.0234)}{.0544}\, incum$$
$$+ \underset{(.0041)}{.0108}\, partyWH \cdot gnews - \underset{(.0033)}{.0077}\, partyWH \cdot inf$$
$$n = 20,\ R^2 = .663,\ \bar{R}^2 = .573.$$

```
tsset year
drop if year==1996
reg demvote partyWH incum pWHgnews pWHinf
```



ISBN 9781408093757

# Time series with trends




Example for a time series with a linear upwards trend



ISBN 9781408093757

# Time series with trends (cont.)

- **Modelling a linear time trend**

$$y_t = \alpha_0 + \alpha_1 t + e_t \quad \Leftrightarrow \quad E(\Delta y_t) = E(y_t - y_{t-1}) = \alpha_1$$

$$\partial y_t / \partial t = \alpha_1$$

Abstracting from random deviations, the dependent variable increases by a constant amount per time unit

$$E(y_t) = \alpha_0 + \alpha_1 t$$

Alternatively, the expected value of the dependent variable is a linear function of time

- **Modelling an exponential time trend**

$$\log(y_t) = \alpha_0 + \alpha_1 t + e_t \quad \Leftrightarrow \quad E(\Delta \log(y_t)) = \alpha_1$$

$$(\partial y_t / y_t) / \partial t = \alpha_1$$

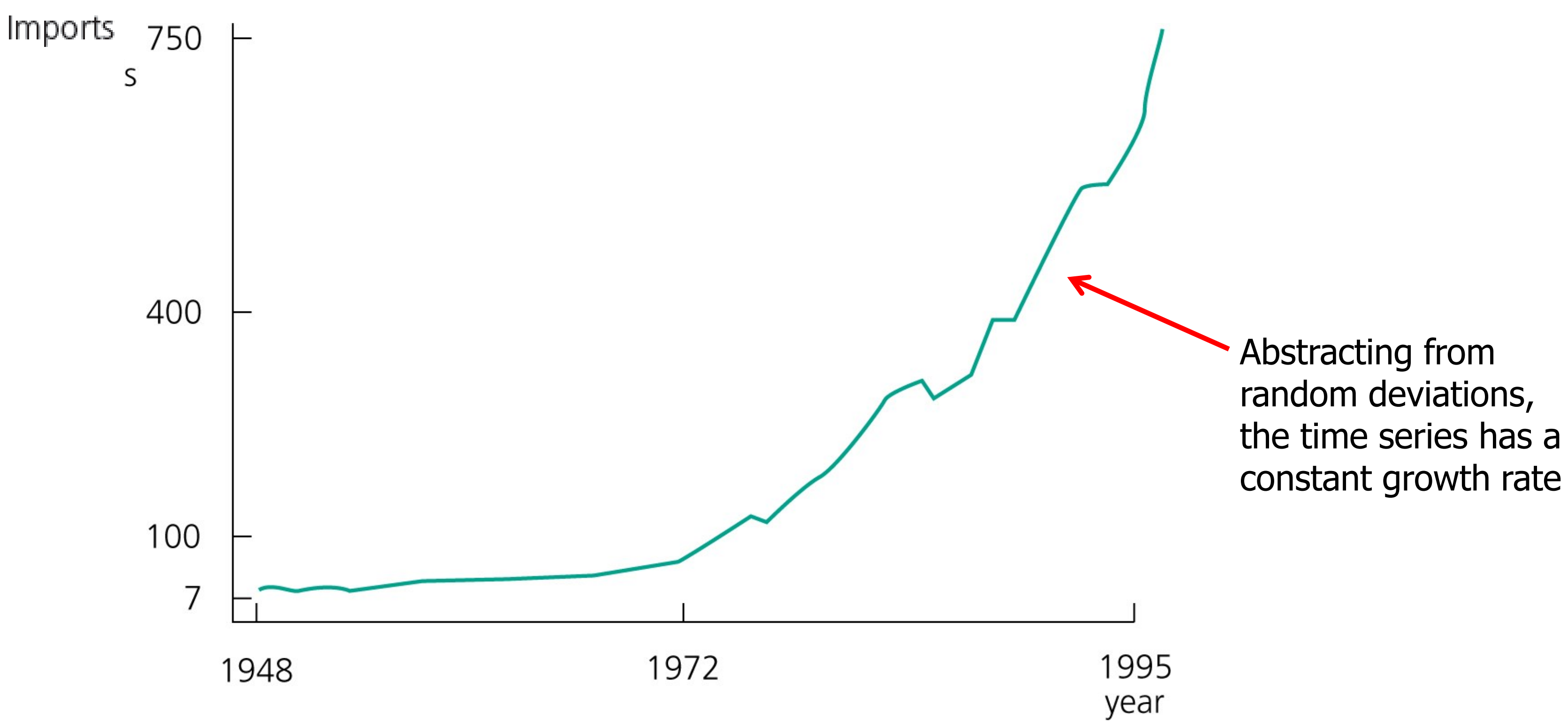
Abstracting from random deviations, the dependent variable increases by <u>a constant percentage</u> per time unit



ISBN 9781408093757

# Time series with trends (cont.)

- **Example for a time series with an exponential trend**






ISBN 9781408093757

# Time series with trends (cont.)

- **Using trending variables in regression analysis**
  - If trending variables are regressed on each other, a spurious re-lationship may arise if the variables are driven by a common trend
  - In this case, it is important to include a trend in the regression
- **Example: Housing investment and prices**

Per capita housing investment

Housing price index

$$\widehat{\log}(invpc) = \underset{(.043)}{-\ .550} + \underset{(.382)}{1.241}\ \log(price)$$

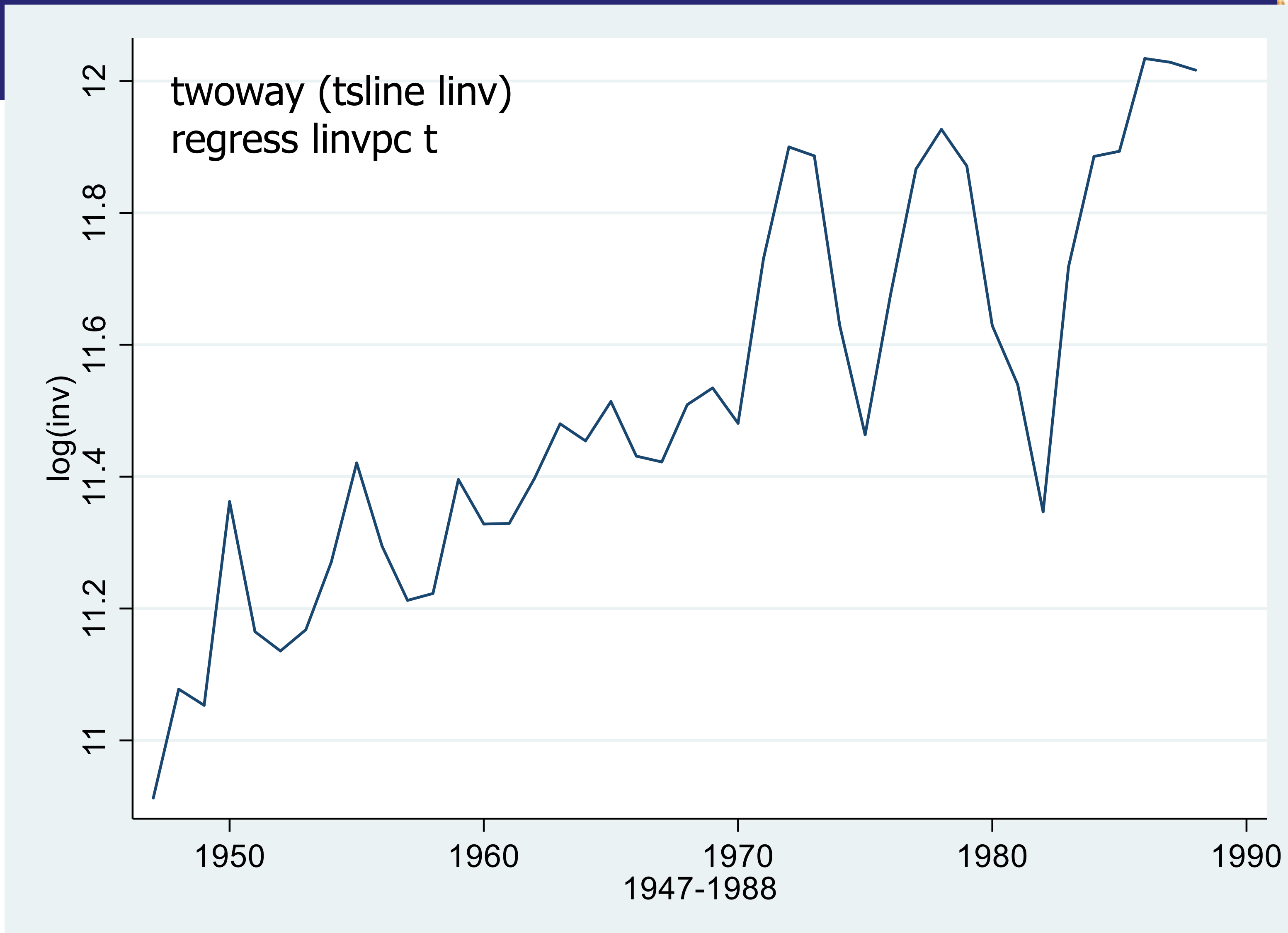
It looks as if investment and prices are positively related

$$n = 42, R^2 = .208, \bar{R}^2 = .189$$

tsset year
reg linvpc lprice

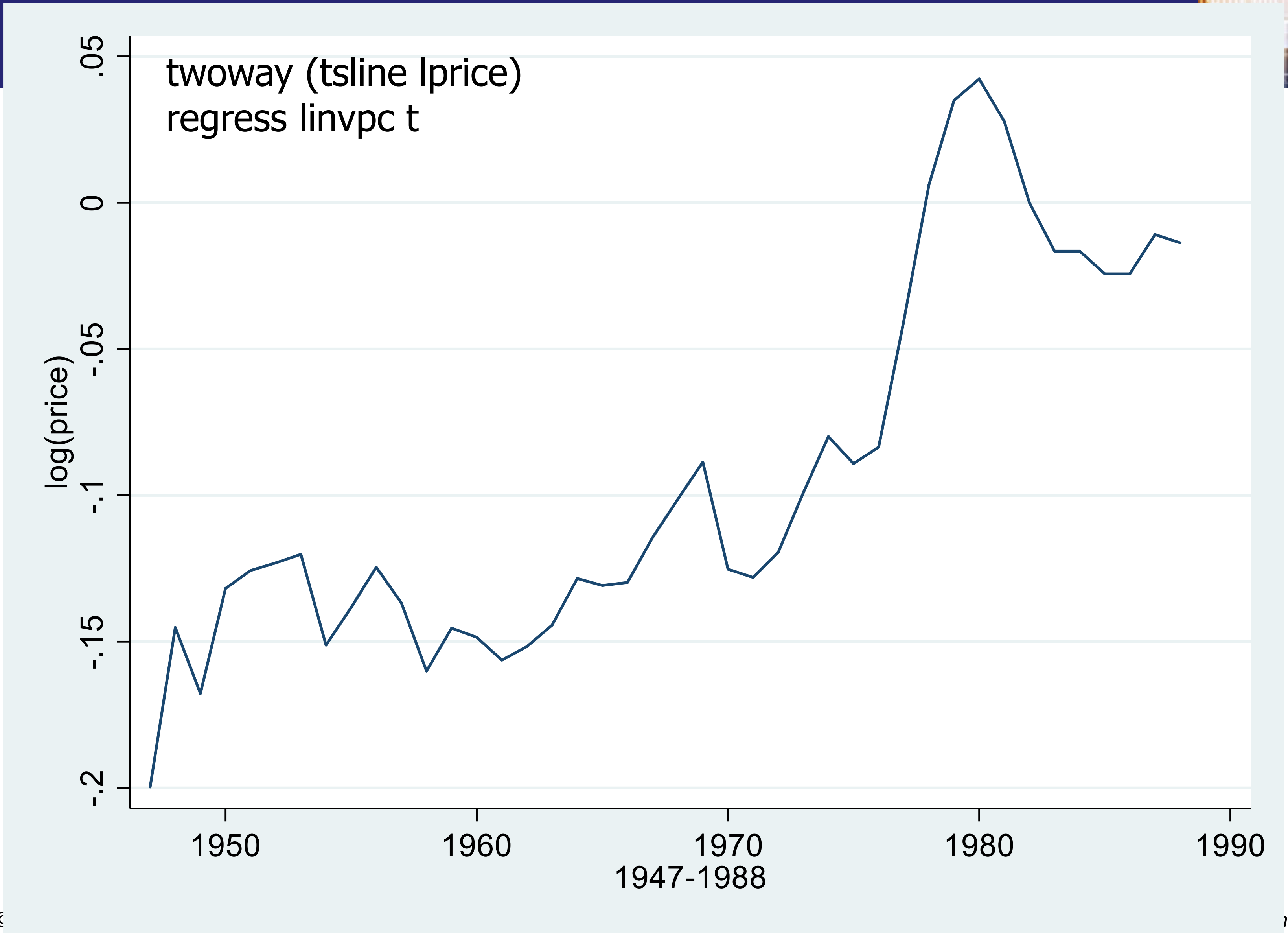


ISBN 9781408093757






ISBN 9781408093757




ISBN 9781408093757

# Time series with trends (cont.)

- **Example: Housing investment and prices (cont.)**

$$\widehat{\log}(invpc) = -\underset{(.136)}{.913} - \underset{(.679)}{.381}\ \log(price) + \boxed{\underset{(.0035)}{.0098\ t}}$$

$$n = 42, R^2 = .341, \bar{R}^2 = .307$$

There is no significant relationship between price and investment anymore

- **When should a trend be included?**
  - If the dependent variable displays an obvious trending behaviour
  - If both the dependent and some independent variables have trends
  - <u>If only some of the independent variables have trends</u>; their effect on the dep. var. may only be visible after a trend has been substracted

reg linvpc lprice t



ISBN 9781408093757

# Time series with trends (cont.)

- **A Detrending interpretation of regressions with a time trend**
  - It turns out that the OLS coefficients in a regression including a trend are the same as the coefficients in a regression without a trend but where all the variables have been detrended before the regression
  - This follows from the general interpretation of multiple regressions
- **Computing R-squared when the dependent variable is trending**
  - Due to the trend, the variance of the dep. var. will be overstated
  - It is better to first detrend the dep. var. and then run the regression on all the indep. variables (plus a trend if they are trending as well)
  - The R-squared of this regression is a more adequate measure of fit


ISBN 9781408093757

# Example: Housing investment and prices (cont.)

```
tsset year
reg linvpc t
predict resid1, resid
reg lprice t
predict resid2, resid
reg resid1 resid2
```

| Source | SS | df | MS |
|---|---|---|---|
| Model | .006498133 | 1 | .006498133 |
| Residual | .804674939 | 40 | .020116873 |
| Total | .811173072 | 41 | .019784709 |

| | | |
|---|---|---|
| Number of obs | = | 42 |
| F(1, 40) | = | 0.32 |
| Prob > F | = | 0.5730 |
| R-squared | = | 0.0080 |
| Adj R-squared | = | -0.0168 |
| Root MSE | = | .14183 |

| resid1 | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| resid2 | -.3809612 | .670296 | -0.57 | 0.573 | -1.73568 | .9737577 |
| _cons | 3.63e-10 | .0218855 | 0.00 | 1.000 | -.0442322 | .0442322 |



ISBN 9781408093757

# Example: Housing investment and prices (cont.)

reg linvpc lprice t

| Source | SS | df | MS |
|---|---|---|---|
| Model | .415945108 | 2 | .207972554 |
| Residual | .804674927 | 39 | .02063269 |
| Total | 1.22062003 | 41 | .02977122 |

| | | |
|---|---|---|
| Number of obs | = | 42 |
| F(2, 39) | = | 10.08 |
| Prob > F | = | 0.0003 |
| R-squared | = | 0.3408 |
| Adj R-squared | = | 0.3070 |
| Root MSE | = | .14364 |

| linvpc | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| lprice | -.3809612 | .6788352 | -0.56 | 0.578 | -1.754035 | .9921125 |
| t | .0098287 | .0035122 | 2.80 | 0.008 | .0027246 | .0169328 |
| _cons | -.9130595 | .1356133 | -6.73 | 0.000 | -1.187363 | -.6387557 |



ISBN 9781408093757

# Modelling seasonality in time series

- **A simple method is to include a set of seasonal dummies:**

$$y_t = \beta_0 + \delta_1 \boxed{feb_t} + \delta_2 \boxed{mar_t} + \delta_3 \boxed{apr_t} + \cdots + \delta_{11} \boxed{dec_t}$$

$$+ \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + u_t$$

=1 if obs. from December
=0 otherwise

- **Similar remarks apply as in the case of deterministic time trends**
  - The regression coefficients on the explanatory variables can be seen as the result of first deseasonalizing the dep. and the explanat. variables
  - An R-squared that is based on first deseasonalizing the dep. var. may better reflect the explanatory power of the explanatory variables


ISBN 9781408093757

# Example: Antidumping Filings and Chemical Imports

- Krupp and Pollard (1996) analyzed the effects of antidumping filings by U.S. chemical industries on imports of various chemicals. We focus here on one industrial chemical, barium chloride, a cleaning agent used in various chemical processes and in gasoline production. The data are contained in the file BARIUM.RAW.
- Three dummy variables: *befile6* is equal to 1 during the six months before filing, *affile6* indicates the six months after filing, and *afdec6* denotes the six months after the positive decision. The dependent variable is the volume of imports of barium chloride from China, *chnimp*, which we use in logarithmic form. We include as explanatory variables, all in logarithmic form, an index of chemical production, *chempi* (to control for overall demand for barium chloride), the volume of gasoline production, *gas* (another demand variable), and an exchange rate index, *rtwex*, which measures the strength of the dollar against several other currencies.



ISBN 9781408093757

# Example: Antidumping Filings and Chemical Imports

reg lchnimp lchempi lgas lrtwex befile6 affile6 afdec6

Using monthly data from February 1978 through December 1988 gives the following:

$$\widehat{\log(chnimp)} = \underset{(21.05)}{-17.80} + \underset{(.48)}{3.12} \log(chempi) + \underset{(.907)}{.196} \log(gas)$$
$$+ \underset{(.400)}{.983} \log(rtwex) + \underset{(.261)}{.060}\, befile6 - \underset{(.264)}{.032}\, affile6 - \underset{(.286)}{.565}\, afdec6$$
$$n = 131,\ R^2 = .305,\ \bar{R}^2 = .271.$$


ISBN 9781408093757

# Example: Antidumping Filings and Chemical Imports

reg lchnimp lchempi lgas lrtwex befile6 affile6 afdec6 feb mar apr may jun jul aug sep oct nov dec

test feb mar apr may jun jul aug sep oct nov dec

```
. test feb mar apr may jun jul aug sep oct nov dec

 ( 1)  feb = 0
 ( 2)  mar = 0
 ( 3)  apr = 0
 ( 4)  may = 0
 ( 5)  jun = 0
 ( 6)  jul = 0
 ( 7)  aug = 0
 ( 8)  sep = 0
 ( 9)  oct = 0
 (10)  nov = 0
 (11)  dec = 0

       F( 11,   113) =    0.86
            Prob > F =    0.5852
```

How about using dummies for quarters? Are the results the same?



ISBN 9781408093757

# Reference

Chapter 10: Basic Regression Analysis with Time Series Data.

In: Wooldridge: Introduction to Econometrics: Europe, Middle East and Africa Edition



ISBN 9781408093757