Basic Regression Analysis with Time Series Data

Chapter 10

Wooldridge: Introductory Econometrics:
A Modern Approach, 5e

with some adjustments by me

Analyzing Time Series: Basic Regression Analysis

- **<u>CLM assumptions for Time Series data</u>**
  - TS.1 Linear in parameters
  - TS.2 No perfect collinearity
  - TS.3 Zero conditional mean
    - strict exogeneity assumption
  - TS.4 Homoscedasticity
  - TS.5 No serial correlation
  - TS.6 Normality

Analyzing Time Series: Basic Regression Analysis

- **Implications ...**
  - When assumptions TS.1 – TS.6 hold, everything we have learned about estimation and inference for cross-sectional regressions applies directly to time series regressions.
  - However, for time series data, the classical linear model (CLM) assumptions are more restrictive than for cross-sectional data, in particular:
    - the strict exogeneity assumption, and
    - the no serial correlation assumption
  - Moreover, finite sample properties. As with cross-section data, large sample properties are less restrictive (*even in time series ...*)

    ➔ *Chapter 11*

Analyzing Time Series: Basic Regression Analysis

- **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

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

  - The expected negative relation is not found. Instead, a positive but insignificant relation is found (t-stat 1.62 implies p-value 0.11)
  - Why …..?

Analyzing Time Series: Basic Regression Analysis

- **Discussion of CLM assumptions**

TS.1: A linear relationship might be restrictive, but it should be a good approximation.

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

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

TS.2: Perfect collinearity is not a problem as long as unemployment varies over time.

Analyzing Time Series: Basic Regression Analysis

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

TS.3: $E(u_t|unem_1, \ldots, unem_n) = 0$ 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

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

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

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

Analyzing Time Series: Basic Regression Analysis

- *10.4* **Functional Form and Dummy Variables**

  - Logarithmic transformation has the usual elasticity interpretation. For example, in the FDL model, we can define *short-run elasticity* (corresponding to *impact propensity*) and *long-run elasticity* (corresponding to *long-run propensity* LRP) by using logarithmic functional forms.

  - Dummy variables are often used to isolate certain periods that may be systematically different from other periods.

Analyzing Time Series: Basic Regression Analysis

- **Example**

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
  - The effect of tax exemption is significant both statistically and economically (\$12 tax exemption→one more baby per 1,000 women).

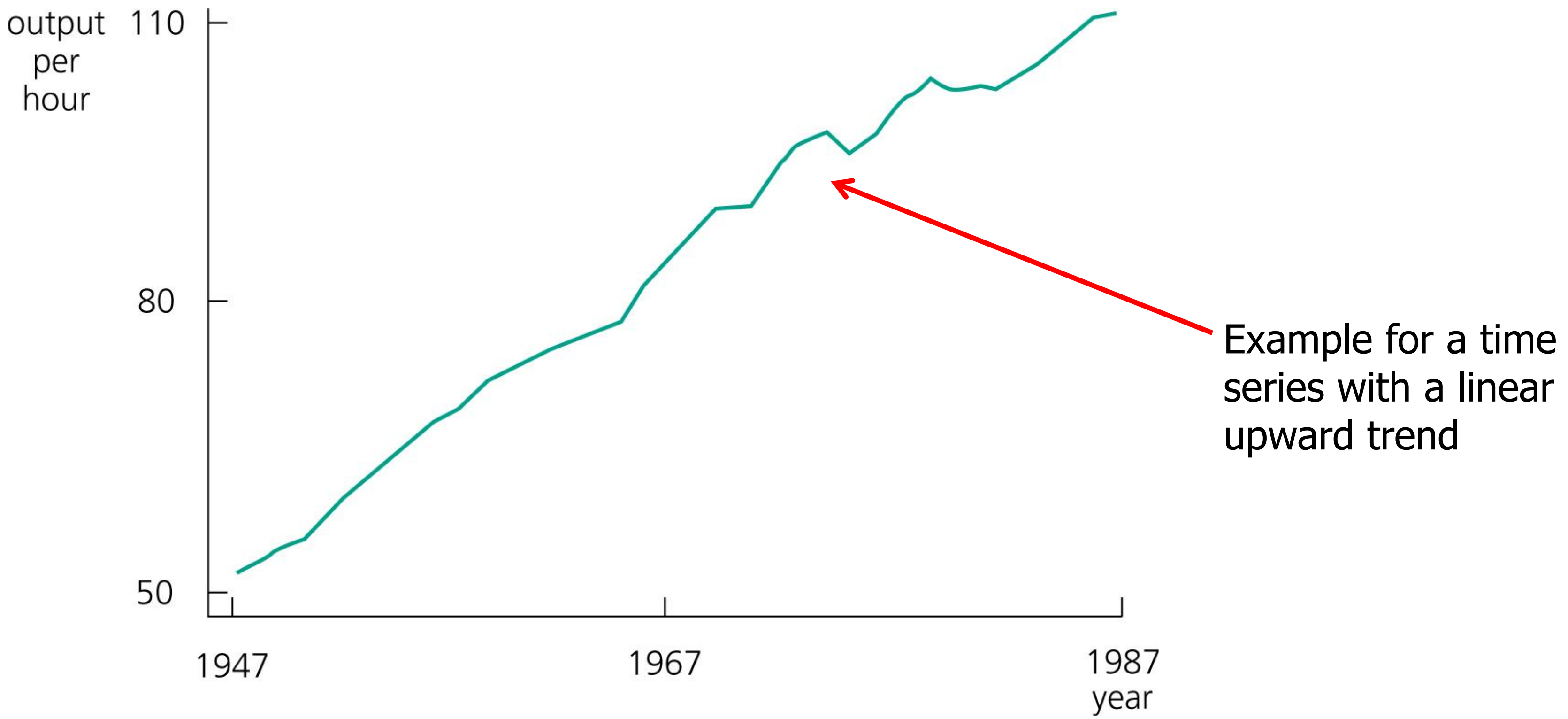
Analyzing Time Series: Basic Regression Analysis

- *10.5* **Time series with trends**

  - Many economic time series have a common tendency of growing over time
  - Ignoring that time series are trending can lead us to falsely conclude that changes in one variable are *caused* by changes in another variable

Analyzing Time Series: Basic Regression Analysis

- **Example: labor productivity**



Analyzing Time Series: Basic Regression Analysis

- **Modelling a linear time trend**
  - Straightforward option …

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

  - … assuming $\{e_t\}$ an i.i.d. sequence with $E(e_t)=0$ and $var(e_t)= \sigma^2$
  - Then:

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

Abstracting from random deviations, the dependent variable increases by <u>a constant amount</u> 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

Analyzing Time Series: Basic Regression Analysis

- **Modelling an exponential time trend**

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

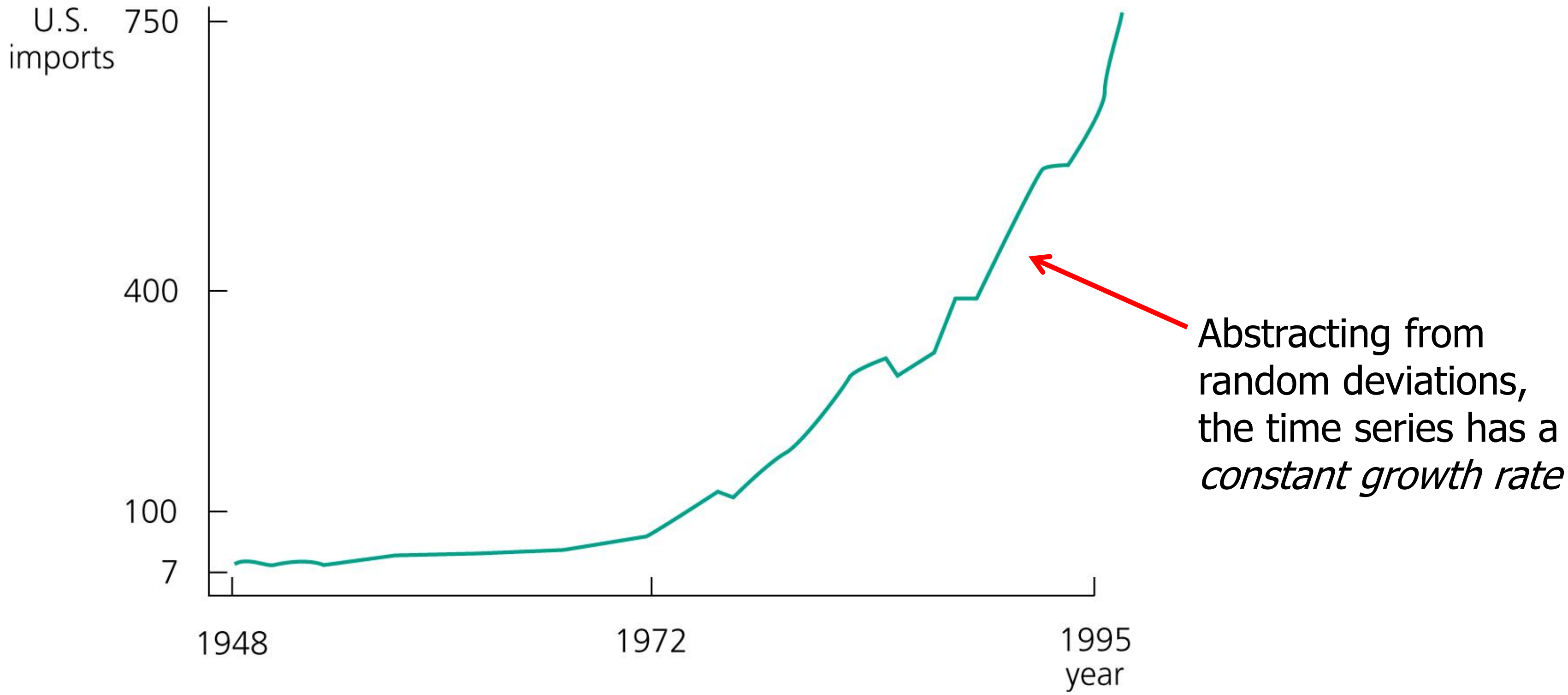
$$(\partial y_t / y_t)/\partial t = \alpha_1$$

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

- More complex trending mechanisms … *e.g*. quadratic trends … but *keep it simple*

Analyzing Time Series: Basic Regression Analysis

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



Analyzing Time Series: Basic Regression Analysis

- **Using trending variables in regression analysis**
  - Hence, accounting for a trend is rather straightforward, …
  - … and does not necessarily violate the CLM assumptions.
  - However, trends in *y* may be correlated with trends in explanatory and/or unobserved factors (errors).
  - If trending variables are regressed on each other, a spurious relationshp may arise if the variables are driven by a common trend
  - In this case, it is important to include a trend in the regression

Analyzing Time Series: Basic Regression Analysis

- **Include trend in the regression**

Suppose a static model:

$$y_t = \beta_0 + \beta_1 x_t + \beta_3 t + u_t.$$

If x_t also includes a trend, then x_t is correlated with t. The regression

$$y_t = \beta_0 + \beta_1 x_t + u_t$$

would have an omitted variable bias. This is why the spurious regression problem appears.

If x_t has a trend but y_t does not, then β_1 in the latter eq. tends to be <u>insignificant</u>. This is because the trending in x_t might be too dominating and obscure any partial effect it might have on y_t.

Analyzing Time Series: Basic Regression Analysis

- **Example: Housing investment and prices**

Per capita housing investment

Housing price index

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

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

It looks as if investment and prices are positively related

  - The elasticity of per capita investment with respect to price is very large …
  - … but we should be careful because both *invpc* and *price* have upward trends (easily revealed when regressing each on *t*)

Analyzing Time Series: Basic Regression Analysis

- **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 including a time trend, the estimated price elasticity is negative[1] and insignificant ...
  - … while the time trend is significant, and implies an approximate 1% increase in *invpc* per year

➔ *The results without trend show a spurious relationship*

[1] Yes, negative. Wooldridge says -0.381 in the book, not +0.381

Analyzing Time Series: Basic Regression Analysis

- **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': detrending

Analyzing Time Series: Basic Regression Analysis

- **A Detrending interpretation of regressions with a time trend**
  - Including a time trend in a regression model vs. detrending the original data series before using them in regression analysis.
  - When we regress y_t on x_{t1}, x_{t2}, and t, we obtain the fitted equation

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_{t1} + \hat{\beta}_2 x_{t2} + \hat{\beta}_3 t.$$

  - The same parameter estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ can be obtained by linearly detrending each of the variables y_t, x_{t1}, and x_{t2}, and next run a regression using the detrended variables.

Analyzing Time Series: Basic Regression Analysis

- **Detrending procedure**
  - For each of the variables y_t, x_{t1}, and x_{t2}, run an OLS regression on a constant and a time trend, e.g.,

$$y_t = \alpha_0 + \alpha_1 t + e_t$$

  - Save the residuals $\hat{e}_t$ as $\hat{e}_t = \ddot{y}_t$ with $\ddot{y}_t = y_t - \hat{\alpha}_0 - \hat{\alpha}_1 t.$
  - Next, run the regression

$$\ddot{y}_t \text{ on } \ddot{x}_{t1}, \ddot{x}_{t2}$$

  - This regression gives exactly the same estimates $\hat{\beta}_1$ and $\hat{\beta}_2$.

Analyzing Time Series: Basic Regression Analysis

- **A Detrending interpretation of regressions with a time trend**
  - Hence, 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 means that the estimates of primary interest can be interpreted as coming from a regression without a time trend, but where we first detrend the dependent variable and all other independent variables
  - Due to this interpretation, it may be a good idea to include a trend in the regression if any independent variable is trending, even if y_t is not.

Analyzing Time Series: Basic Regression Analysis

- **Computing R-squared when the dependent variable is trending**
  - R-squareds in time series regressions are often very high, especially compared with typical R-squareds for cross-sectional data.
  - However, that does not mean that we learn more about factors affecting y from time series data.
    - Time series data often come in aggregate form, that are often easier to explain than outcomes on individuals, families, or firms, which is often the nature of cross-sectional data
    - Moreover, the usual and adjusted R-squareds for time series regressions can be artificially high when the dependent variable is trending

Analyzing Time Series: Basic Regression Analysis

- **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 *net of the effect of the time trend*

Analyzing Time Series: Basic Regression Analysis

- **<u>Seasonality</u>**

  - If a time series is observed at monthly or quarterly intervals (or even weekly or daily), it may exhibit seasonality
    - Weather, Holidays, Agricultural seasons, …

  - One way to model this phenomenon is to allow the expected value of the series, y_t, to be different in each month/quarter/…

Analyzing Time Series: Basic Regression Analysis

<u>Modelling seasonality in time series</u>

- **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