

Basic Regression Analysis with Time Series Data. Wooldridge (2013),
Chapter 10

- Time Series vs. Cross Sectional
- Finite Distributed Lag Models
- Assumptions for Unbiasedness
- Variances of OLS Estimators
- Inference on the long-run propensity
- Trending Time Series
- Seasonality

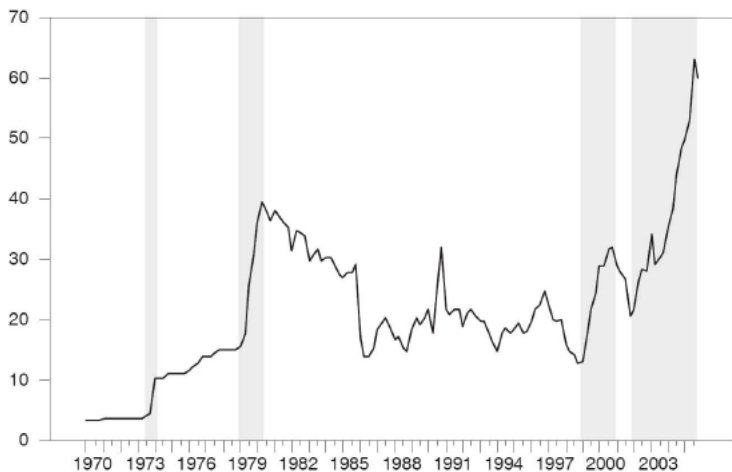
Basic Regression Analysis with Time Series Data

Time Series vs. Cross Sectional

- Time series data has a temporal ordering, unlike cross-section data
- Will need to alter some of our assumptions to take into account that we no longer have a random sample of individuals.
- Instead, we have a sequence of random variables indexed by time
- This type of sequence is known as stochastic (i.e. random) process or time series process

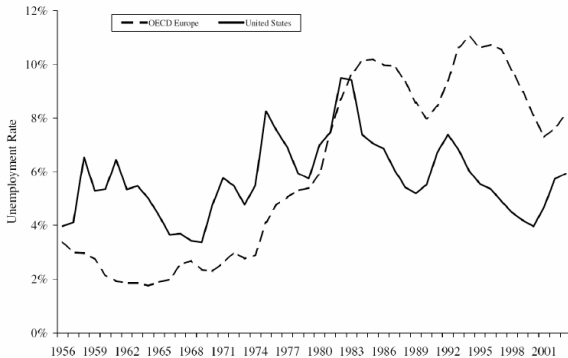
Basic Regression Analysis with Time Series Data

Oil prices (\$ per barrel)



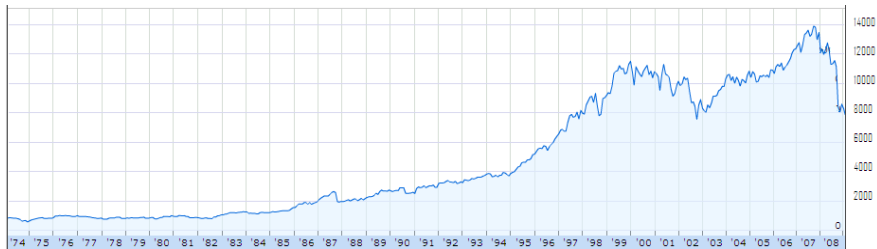
Basic Regression Analysis with Time Series Data

Unemployment rate: U.S. And Europe



Basic Regression Analysis with Time Series Data

Dow Jones Industrial Average



Basic Regression Analysis with Time Series Data

Examples of Time Series Models

- We use the subscript t (e.g. y_t) denote an observation taken at time t .
- A **static model** relates contemporaneous variables:

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

- **Examples:**

- Static Phillips curve:

$$\text{infl}_t = \beta_0 + \beta_1 \text{unemp}_t + u_t$$

- Murders per 1000 people in a given city:

$$\text{mrdrte}_t = \beta_0 + \beta_1 \text{convrte}_t + \beta_2 \text{unemp}_t + \beta_3 \text{yngmle}_t + u_t$$

where: mrdrte_t is the murder rate, convrte_t is the murder conviction rate, unemp_t is the unemployment rate and yngmle_t is the fraction of males aged between 18 and 25 (all for year t).

Basic Regression Analysis with Time Series Data

Examples of Time Series Models

- A **finite distributed lag** (FDL) model allows one or more variables to affect y with a lag:

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + u_t$$

- **Examples:** Effect of taxes on fertility rates:

$$gfr_t = \beta_0 + \delta_0 pe_t + \delta_1 pe_{t-1} + \delta_2 pe_{t-2} + u_t$$

where: gfr_t is the gross fertility rate for year t and pe_t is the real value of personal tax exemptions for year t .

- More generally, a finite distributed lag model of order q will include q lags of z .

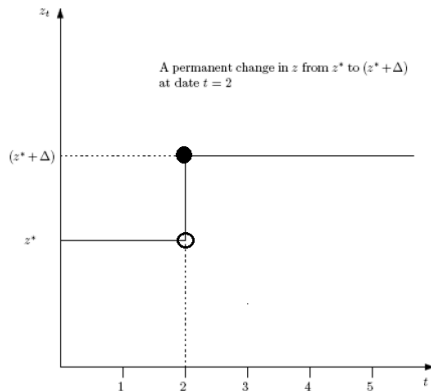
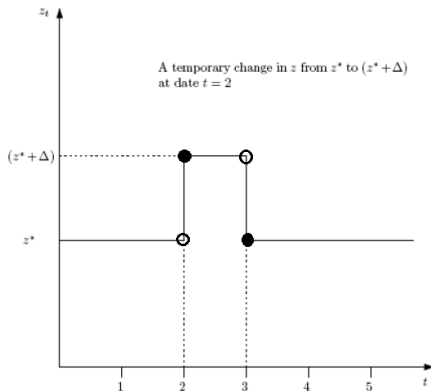
$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \dots + \delta_q z_{t-q} + u_t.$$

Basic Regression Analysis with Time Series Data

Finite Distributed Lag Models

We can consider:

- temporary increases in z
- permanent increases in z



Basic Regression Analysis with Time Series Data

Finite Distributed Lag Models

Consider the model:

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + u_t$$

- Suppose that z increases temporarily from z^* to $(z^* + \Delta)$ at date t but then reverts back to z^* at date $(t + 1)$. Then:

$$y_t = \alpha_0 + \delta_0(z^* + \Delta) + \delta_1 z^* + \delta_2 z^* + u_t.$$

$$y_{t+1} = \alpha_0 + \delta_0 z^* + \delta_1(z^* + \Delta) + \delta_2 z^* + u_{t+1},$$

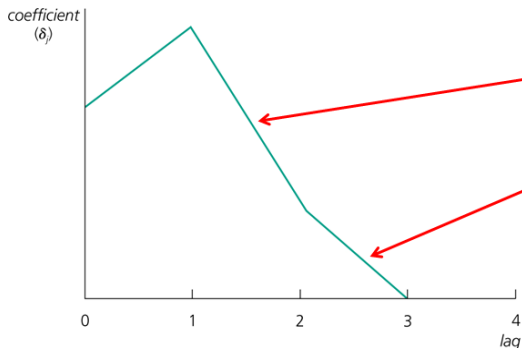
$$y_{t+2} = \alpha_0 + \delta_0 z^* + \delta_1 z^* + \delta_2(z^* + \Delta) + u_{t+2}.$$

- We call δ_0 the impact propensity or impact multiplier – it reflects the immediate change in y when z increases temporarily from z^* to $(z^* + \Delta)$ at date t , with $\Delta = 1$
- $\frac{\partial y_{t+s}}{\partial z_t} = \delta_s, s = 0, 1, 2$ – Effect of a transitory shock: If there is a one time unit change in time t of z_t , the dep. variable in time $t + s$ will change temporarily by the amount indicated by the coefficient of the corresponding lag (z_t)

Basic Regression Analysis with Time Series Data

Finite Distributed Lag Models

Graphical illustration of lagged effects



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).

Basic Regression Analysis with Time Series Data

Finite Distributed Lag Models

Consider the model:

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + u_t$$

Suppose that z increases permanently from z^* to $(z^* + \Delta)$ at date t . Then:

$$\begin{aligned}y_t &= \alpha_0 + \delta_0(z^* + \Delta) + \delta_1 z^* + \delta_2 z^* + u_t. \\y_{t+1} &= \alpha_0 + \delta_0(z^* + \Delta) + \delta_1(z^* + \Delta) + \delta_2 z^* + u_{t+1}, \\y_{t+2} &= \alpha_0 + \delta_0(z^* + \Delta) + \delta_1(z^* + \Delta) + \delta_2(z^* + \Delta) + u_{t+2}.\end{aligned}$$

Notice that:

- δ_0 yields the immediate change in y_t when z^* changes permanently to $(z^* + \Delta)$ at period t with $\Delta = 1$
- $\delta_0 + \delta_1$ yields the change in y_{t+1} when z^* changes permanently to $(z^* + \Delta)$ at period t with $\Delta = 1$.
- $\delta_0 + \delta_1 + \delta_2$ yields the change in y_{t+2} when z^* changes permanently to $(z^* + \Delta)$ at period t with $\Delta = 1$.
- We can call $\delta^* = \delta_0 + \delta_1 + \delta_2$ the long-run propensity (LRP) – it reflects the long-run change in y after a permanent change in z .

Basic Regression Analysis with Time Series Data

Finite Distributed Lag Models

- In the long run and $z_t = z_{t-1} = z_{t-2} = z^* + \Delta$. Define the long run values as $y_t = \tilde{y}^*$ and $\tilde{z}^* = z^* + \Delta$ and notice that in the long run $u_t = 0$. Hence the long run relationship is given by

$$\begin{aligned}\tilde{y}^* &= \alpha_0 + \delta_0 \tilde{z}^* + \delta_1 \tilde{z}^* + \delta_2 \tilde{z}^* \\ &= \alpha_0 + (\delta_0 + \delta_1 + \delta_2) \tilde{z}^*\end{aligned}$$

- Consider the model:

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

In this case the long run propensity is given by $\delta_0 + \delta_1 + \dots + \delta_q$

Basic Regression Analysis with Time Series Data

Finite Distributed Lag Models

Example: Effect of taxes on fertility rates: Data for US (1913-1984):

$$gfr_t = \beta_0 + \delta_0 pe_t + \delta_1 pe_{t-1} + \delta_2 pe_{t-2} + \beta_1 ww2_t + \beta_2 pill_t + u_t$$

where: gfr_t is the gross fertility rate for year t and pe_t is the real value of personal tax exemptions for year t . $ww2_t$ — dummy value that takes value 1 from 1941-1945 (World War II), 0 otherwise. $pill_t$ — dummy variable equal to 1 from 1963 on, when the birth control pill was made available for contraception.

$$\widehat{gfr}_t = 95.87 + 0.073pe_t - 0.0058pe_{t-1} + 0.034pe_{t-2} \\ - 22.13ww2_t - 31.30pill_t,$$

(3.28) (0.126) (0.1557) (0.126)
(10.73) (3.98)

$$n = 70, R^2 = 0.499.$$

Basic Regression Analysis with Time Series Data

Assumptions for Unbiasedness

Assumption (TS.1 - linearity in parameters)

The stochastic process $\{(y_t, x_{t1}, x_{t2}, \dots, x_{tk}); t = 1, 2, \dots, n\}$ follows the linear model:

$$y_t = \beta_0 + \beta_1 x_{t1} + \dots + \beta_k x_{tk} + u_t$$

- Note that some regressors may be lagged values of other regressors.

Assumption (TS.2 - no perfect collinearity)

No regressor is a constant nor a perfect linear combination of the other regressors.

Let X denotes the matrix of all regressors for all time-periods, i.e. $X = [x_{tj}; t = 1, \dots, n \ \& \ j = 1; \dots; k]$.

Assumption (TS.3 - zero conditional mean)

$$E(u_t|X) = 0, t = 1, 2, \dots, n$$

Basic Regression Analysis with Time Series Data

Assumptions for Unbiasedness

- If Assumption TS.3 holds we say that the x 's are *strictly exogenous*
- Note that this assumption implies the error term in any given period is uncorrelated with the explanatory variables in all time periods.
- OLS is unbiased under Assumptions TS.1–TS.3. That is

$$E(\hat{\beta}_j) = \beta_j, j = 0, \dots, k$$

Basic Regression Analysis with Time Series Data

Assumptions for Unbiasedness

- An alternative (and weaker) assumption, more parallel to the cross-sectional case, is

$$E(u_t|x_t) = 0$$

This implies that the current error term, u_t , is uncorrelated with the current regressors, $(x_{t1}, x_{t2}, \dots, x_{tk})$. That is the x 's are *contemporaneously exogenous*.

- Contemporaneous exogeneity will only be sufficient for consistency that is

$$\text{plim}(\hat{\beta}_j) = \beta_j, j = 0, \dots, k.$$

- This is fortunate as the current error term, u_t , is often correlated with the future values of the regressors

Basic Regression Analysis with Time Series Data

Assumptions (continued)

- Note we have skipped the assumption of a random sample
- The key impact of the random sample assumption is that each u_i is independent of the regressors of other observations in the sample.
- Our strict exogeneity assumption takes care of it in this case
- Omitted variable bias can be analyzed in the same manner as in the cross-section case.

Basic Regression Analysis with Time Series Data

Variations of OLS Estimators

- Just as in the cross-section case, we need to add an assumption of homoskedasticity in order to be able to derive variances

Assumption (TS.4 - homoskedasticity)

For each $t = 1, 2, \dots, n$:

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

- Thus, the error variance is independent of all the x 's, and it is constant over time

Assumption (TS.5 - No Serial Correlation)

For each $t; s = 1, 2, \dots, n$ such that $t \neq s$:

$$\text{Corr}(u_t, u_s|X) = 0$$

Remark: If Assumption TS.5 does not hold that is if $\text{Corr}(u_t, u_s|X) \neq 0$ we say that the errors suffer from *serial correlation*.

Basic Regression Analysis with Time Series Data

OLS Variances (continued)

- Under Assumptions TS.1–TS.5:
 - the OLS variances in the time-series case are the same as in the cross-section case.

$$\text{Var}(\hat{\beta}_j|X) = \frac{\sigma^2}{SST_j(1 - R_j^2)}$$

where

- $SST_j = \sum_{i=1}^n (x_{tj} - \bar{x})^2$ and
 - R_j^2 is the R-squared from regressing x_{tj} on the other independent variables.
- The estimator of σ^2 is the same:

$$\hat{\sigma}^2 = \frac{SSR}{n - k - 1}$$

- OLS remains Best Linear Unbiased Estimator (Gauss Markov Theorem).
- With the additional assumption of normal errors (**TS.6**), inference is the same as in the cross-sectional case.

Basic Regression Analysis with Time Series Data

Inference on the long-run propensity

Consider the model:

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + u_t$$

- Inference on $\delta^* = \delta_0 + \delta_1 + \delta_2$. Use the t-test

$$t = \frac{\hat{\delta}_0 + \hat{\delta}_1 + \hat{\delta}_2 - \delta^*}{se(\hat{\delta}_0 + \hat{\delta}_1 + \hat{\delta}_2)}$$

where $se(\hat{\delta}_0 + \hat{\delta}_1 + \hat{\delta}_2)$ is the standard error of $\hat{\delta}_0 + \hat{\delta}_1 + \hat{\delta}_2$

- Notice that

$$\begin{aligned} Var(\hat{\delta}_0 + \hat{\delta}_1 + \hat{\delta}_2) &= Var(\hat{\delta}_0) + Var(\hat{\delta}_1) + Var(\hat{\delta}_2) \\ &\quad + 2Cov(\hat{\delta}_0, \hat{\delta}_1) + 2Cov(\hat{\delta}_1, \hat{\delta}_2) + 2Cov(\hat{\delta}_0, \hat{\delta}_2) \end{aligned}$$

- Hence

$$\begin{aligned} se(\hat{\delta}_0 + \hat{\delta}_1 + \hat{\delta}_2)^2 &= se(\hat{\delta}_0)^2 + se(\hat{\delta}_1)^2 + se(\hat{\delta}_2)^2 \\ &\quad + 2s_{01} + 2s_{12} + 2s_{02} \end{aligned}$$

where $se(\hat{\delta}_j)$ is the standard error of $\hat{\delta}_j$ and s_{ij} is an estimator of $Cov(\hat{\delta}_i, \hat{\delta}_j)$.

Basic Regression Analysis with Time Series Data

Inference on the long-run propensity

- Alternative way: Computation of the standard errors of the OLS estimator of δ^* . Notice that $\delta_0 = \delta^* - \delta_1 - \delta_2$, substituting this in

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + u_t$$

we obtain

$$y_t = \alpha_0 + \delta^* z_t + \delta_1 (z_{t-1} - z_t) + \delta_2 (z_{t-2} - z_t) + u_t.$$

- Hence by running the regression of y_t on a intercept, z_t , $z_{t-1} - z_t$ and $z_{t-2} - z_t$ the standard error of the OLS estimator of the slope of z_t is the standard error of the OLS estimator of δ^* .

Basic Regression Analysis with Time Series Data

Inference on the long-run propensity

Example (cont): Effect of taxes on fertility rates: Data for US (1913-1984):

$$\widehat{gfr}_t = 95.87 + 0.073pe_t - 0.0058pe_{t-1} + 0.034pe_{t-2} - 22.13ww2_t - 31.30pill_t,$$

$(3.28) \quad (0.126) \quad (0.1557) \quad (0.126)$
 $(10.73) \quad (3.98)$

$$n = 70, R^2 = 0.499.$$

(Standard errors are in parentheses)

The estimated variance-covariance matrix of the OLS estimator is

	Intercept	pe	pe(-1)	pe(-2)	pill	ww2
Intercept	10.771					
pe	-0.093	0.016				
pe(-1)	0.018	-0.011	0.024			
pe(-2)	0.002	-0.003	-0.013	0.016		
pill	-1.529	-0.006	-0.002	-0.030	15.853	
ww2	2.642	-0.793	-0.122	0.810	5.711	115.180

Test the hypothesis that the long-run multiplier of gfr with respect to pe is 0.05.

Basic Regression Analysis with Time Series Data

Trending Time Series

- Economic time series often have a trend.
- Just because 2 series are trending together, we can't assume that the relation is causal.
- The phenomenon of finding of a relationship between 2 or more trending variables because each is growing over time is an example of **spurious regression problem**.
- Often, both will be trending because of other unobserved factors
- Even if those factors are unobserved, we can control for them by directly controlling for the trend

Basic Regression Analysis with Time Series Data

Trends (continued)

- One possibility is a linear trend, which can be modeled as

$$\begin{aligned}y_t &= \alpha_0 + \alpha_1 t + e_t, \\t &= 1, 2, \dots\end{aligned}$$

Ignoring the error term α_1 gives the growth of y_t in absolute terms:

$$\Delta y_t = y_t - y_{t-1} = \alpha_1$$

- Another possibility is an exponential trend, which can be modeled as

$$\begin{aligned}\log(y_t) &= \alpha_0 + \alpha_1 t + e_t, \\t &= 1, 2, \dots\end{aligned}$$

Ignoring the error term α_1 gives growth in proportional terms

$$\begin{aligned}\Delta \log(y_t) &= \log(y_t) - \log(y_{t-1}) = \alpha_1. \\ \Delta \log(y_t) &\simeq \frac{y_t - y_{t-1}}{y_{t-1}}\end{aligned}$$

Basic Regression Analysis with Time Series Data

Trends (continued)

- Another possibility is a quadratic trend, which can be modeled as

$$\begin{aligned}y_t &= \alpha_0 + \alpha_1 t + \alpha_2 t^2 + e_t, \\t &= 1, 2, \dots\end{aligned}$$

Notice that

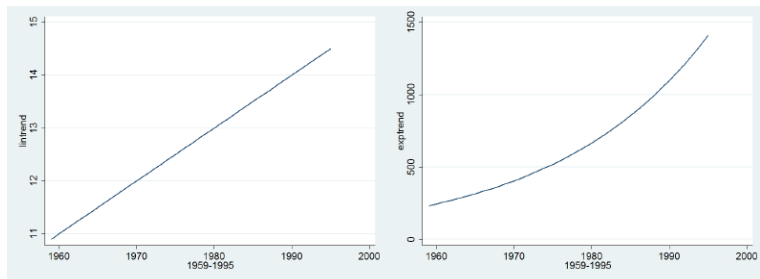
$$\frac{\partial y_t}{\partial t} = \alpha_1 + 2\alpha_2 t$$

so we allow the slope to change with t .

Basic Regression Analysis with Time Series Data

Trends (continued)

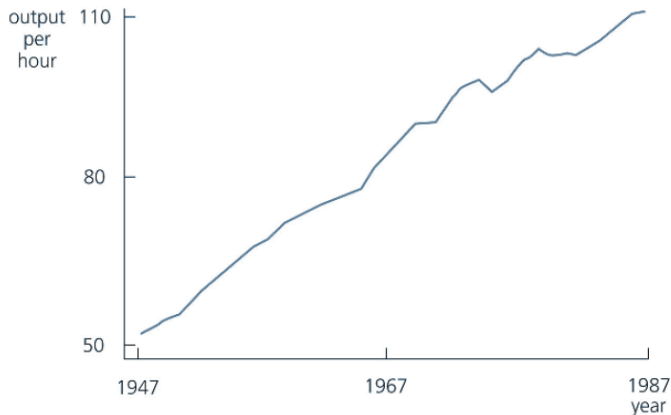
Figure: Linear and Exponential Trends



Basic Regression Analysis with Time Series Data

Trends (continued)

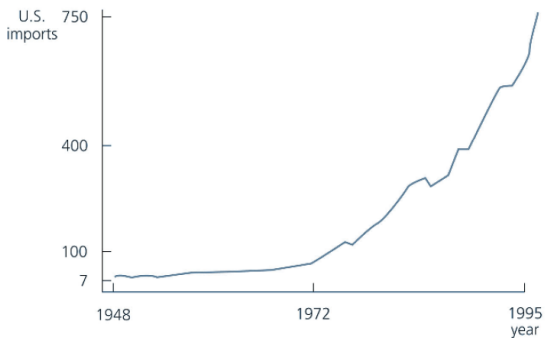
Output per labor hour in the United States during the years 1947–1987; 1977 = 100.



Basic Regression Analysis with Time Series Data

Trends (continued)

Nominal U.S. imports during the years 1948–1995 (in billions of U.S. dollars).



Basic Regression Analysis with Time Series Data

Trends (continued)

Example: Housing investment and prices

Per capita housing investment

Housing price index

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

(.043) (.382)

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

It looks as if investment and prices are positively related

$$\widehat{\log(invpc)} = - .913 + .381 \log(price) + .0098 t$$

(.136) (.679) (.0035)

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

There is no significant relationship between price and investment anymore

Basic Regression Analysis with Time Series Data

Trends (continued)

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
- If only some of the independent variables have trends; their effect on the dep. var. may only be visible after a trend has been subtracted

Basic Regression Analysis with Time Series Data

Detrending

- Adding a linear trend term to a regression is the same thing as using “detrended” series in a regression
- Detrending a series involves regressing each variable in the model on an intercept and the time trend t .
- The residuals of this regression form the detrended series.
- Basically, the trend has been partialled out.

Basic Regression Analysis with Time Series Data

Detrending (continued)

- An advantage to actually detrending the data (vs. adding a trend) involves the calculation of goodness of fit.
- Time-series regressions tend to have very high R^2 .
- Does this mean that we learn more about factors affecting y from time series data?
- No. The R^2 and adjusted R^2 can artificially be high when y is trending.
- The R^2 from a regression on detrended data better reflects how well the x_t 's explain y_t .
- No need to include the intercept in the model with detrended data, but including an intercept affects nothing: the intercept will be estimated to be zero.

Basic Regression Analysis with Time Series Data

Trends (continued)

Example: (Estimation of the Phillips curve)

- Consider the regression of inflation rate (*infl*) on unemployment rate (*ur*) and a trend. 1959 to 1995 (37 obs.), US data.

$$\begin{aligned}\widehat{infl} &= 1.49974 + 0.4016 ur + 0.03666 t, \\ R^2 &= 0.0736, \bar{R}^2 = 0.0191\end{aligned}$$

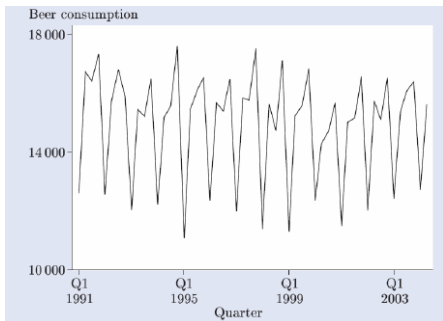
- Consider the regression of the detrended inflation rate (*infdet*) on detrended unemployment rate (*urdet*). 1959 to 1995, US data.

$$\begin{aligned}\widehat{infdet} &= 0.4016 urdet \\ R^2 &= 0.0301, \bar{R}^2 = 0.0301\end{aligned}$$

Basic Regression Analysis with Time Series Data

Seasonality

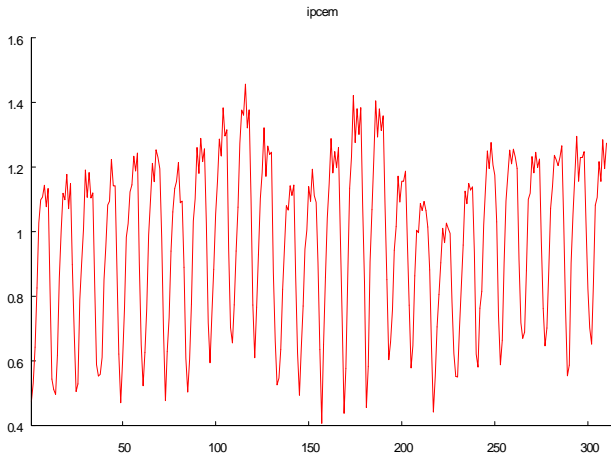
- Often time-series data exhibits some periodicity, referred to as seasonality.
- **Example:** Beer consumption increases in the Summer.



Basic Regression Analysis with Time Series Data

Seasonality

- **Example:** industrial production index, cement (*ipcem*) US data (1964-1989)



Basic Regression Analysis with Time Series Data

Seasonality

- Seasonality can be dealt with by adding a set of seasonal dummies

$$\begin{aligned} ipcem_t = & \beta_0 + \delta_1 Jan_t + \delta_2 Feb_t + \delta_3 Mar_t + \\ & \delta_4 Apr_t + \delta_5 May_t + \delta_6 Jun_t + \delta_7 Jul_t \\ & + \delta_8 Aug_t + \delta_9 Sep_t + \delta_{10} Oct_t + \delta_{11} Nov_t + \\ & \beta_1 x_{1t} + \dots + \beta_k x_{kt} + u_t. \end{aligned}$$

where Jan_t Feb_t Mar_t Apr_t May_t Jun_t Aug_t Sep_t Oct_t Nov_t and Monthly dummy variables.

- As with trends, the series can be seasonally adjusted before running the regression.

Basic Regression Analysis with Time Series Data

Seasonality

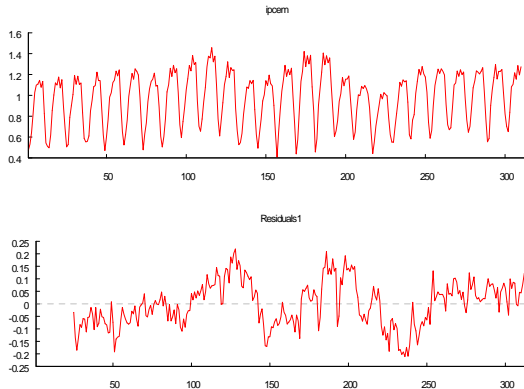
- Running the regression of $ipcem_t$ on an intercept and Jan_t Feb_t Mar_t Apr_t May_t Jun_t Aug_t Sep_t Oct_t Nov_t we have

$$\begin{aligned}\widehat{ipcem}_t &= 0.7025 - 0.15679 Jan_t - 0.07071 Feb_t + 0.10069 Mar_t + \\ &0.28953 Apr_t + 0.38327 May_t + 0.51013 Jun_t \\ &+ 0.45528 Jul_t + 0.53523 Aug_t + 0.48085 Sep_t \\ &+ 0.50122 Oct_t + 0.24739 Nov_t \\ n &= 312, R^2 = 0.8766, \bar{R}^2 = 0.8719\end{aligned}$$

Basic Regression Analysis with Time Series Data

Seasonality

Let Residuals1 be the residuals of the above model. These residuals can be seen as one possible ways of computing the industrial production index of cement after removing seasonality (seasonally adjusted series).



Basic Regression Analysis with Time Series Data

Seasonality

One could also include a trend in the model.

$$\begin{aligned} ipcem_t = & \beta_0 + \delta_0 t + \delta_1 Jan_t + \delta_2 Feb_t + \delta_3 Mar_t + \\ & \delta_4 Apr_t + \delta_5 May_t + \delta_6 Jun_t + \delta_7 Jul_t \\ & + \delta_8 Aug_t + \delta_9 Sep_t + \delta_{10} Oct_t + \delta_{11} Nov_t + \\ & \beta_1 x_{1t} + \dots + \beta_k x_{kt} + u_t. \end{aligned}$$

$$\begin{aligned} \widehat{ipcem}_t = & 0.66693 + 0.00024 t - 1556 Jan_t - 0.06976 Feb_t + 0.1014 Mar_t + \\ & 0.29 Apr_t + 0.3835 May_t + 0.51013 Jun_t + 0.45504 Jul_t \\ & + 0.53476 Aug_t + 0.48014 Sep_t + 0.50027 Oct_t + 0.24763 Nov_t, \\ n = & 312, R^2 = 0.8831, \bar{R}^2 = 0.8782 \end{aligned}$$