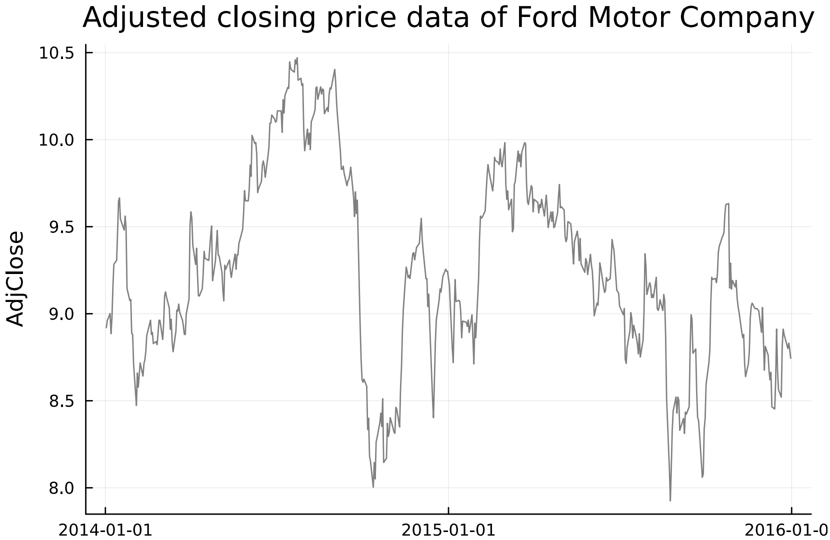
Lecture 1: Time series regression: Basics

Yu Bai

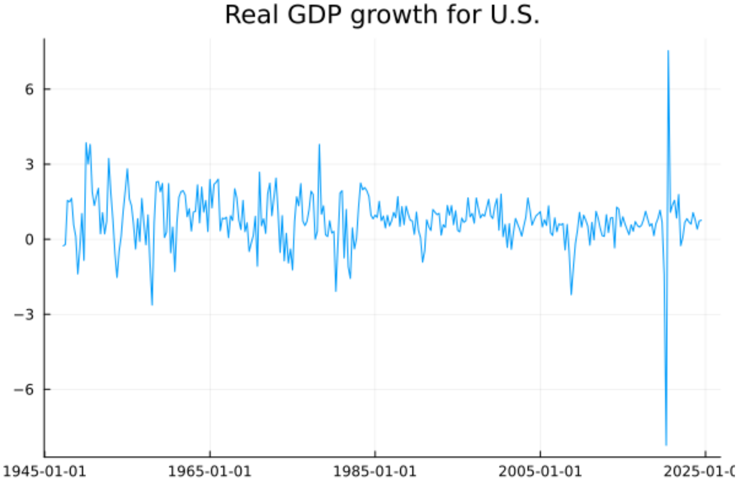
City University of Macau

Time series



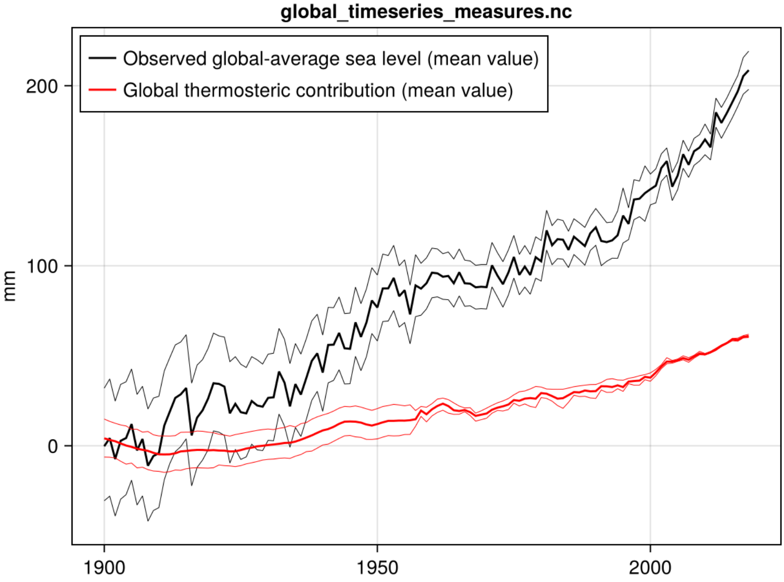
Adjusted closing price data of Ford Motor Company
AdjClose
10.5
10.0
9.5
9.0
8.5
8.0
2014-01-01
2015-01-01
2016-01-0

Time series



Real GDP growth for U.S.
6
3
0
−3
−6
1945-01-01
1965-01-01
1985-01-01
2005-01-01
2025-01-0

Time series


global_timeseries_measures.nc
Observed global-average sea level (mean value)
Global thermosteric contribution (mean value)
mm
200
100
0
1900
1950
2000

Time series

- Some distinguishing features of time series data:
  - (Chronological) ordering of data matters
  - Dependence across time
  - Dependence on time (seasonal effects, trends)
- A time series data set is a realization of a stochastic process.
  - This **process** is often **dependent** and may be **nonstationary**.
  - Repeated sampling from the population of all its possible realizations induces randomness of estimators and other time series statistics.

Examples of time series regression models

- Static models:

$$y_t = \beta_0 + \beta_1 z_t + u_t, \quad t = 1, 2, \cdots, T.$$

- Implication: a change in z at time t is believed to have an immediate effect on y, so that $\Delta y_t = \beta_1 \Delta z_t$ when $\Delta u_t = 0$.

Example

(Static Phillips curve). The static Phillips curve is given by

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

where inf_t is the annual inflation rate and $unem_t$ is the annual unemployment rate.

Examples of time series regression models

- Finite distributed lag (FDL) models:

$$y_t = \alpha_0 + \sum_{\ell=0}^{q} \delta_q z_{t-q} + u_t, \quad t = 1, 2, \cdots, T.$$

- We allow one (or more) variables to affect y with lags.
- We are often interested in the **long-run multiplier** or the **long-run propensity (LRP)** when using these type of models.

A short example

- $q = 2$
- z increases permanently to $c + 1$ at time t

A short example

- $q = 2$
- z increases permanently to $c + 1$ at time t
- To focus on the ceteris paribus effect of z on y, we set the error term in each time period to zero. Then,

$$
\begin{aligned}
y_{t-1} &= \alpha_0 + \delta_0 c + \delta_1 c + \delta_2 c \\
y_t &= \alpha_0 + \delta_0(c+1) + \delta_1 c + \delta_2 c \\
y_{t+1} &= \alpha_0 + \delta_0(c+1) + \delta_1(c+1) + \delta_2 c \\
y_{t+2} &= \alpha_0 + \delta_0(c+1) + \delta_1(c+1) + \delta_2(c+1),
\end{aligned}
$$

and so on.

A short example

- $q = 2$
- z increases permanently to $c + 1$ at time t
- To focus on the ceteris paribus effect of z on y, we set the error term in each time period to zero. Then,

$$
\begin{aligned}
y_{t-1} &= \alpha_0 + \delta_0 c + \delta_1 c + \delta_2 c \\
y_t &= \alpha_0 + \delta_0(c+1) + \delta_1 c + \delta_2 c \\
y_{t+1} &= \alpha_0 + \delta_0(c+1) + \delta_1(c+1) + \delta_2 c \\
y_{t+2} &= \alpha_0 + \delta_0(c+1) + \delta_1(c+1) + \delta_2(c+1),
\end{aligned}
$$

  and so on.
- With the permanent increase in z, after one period, y has increased by $\delta_0 + \delta_1$, and after two periods, y has increased by $\delta_0 + \delta_1 + \delta_2$.

A short example

- $q = 2$
- z increases permanently to $c + 1$ at time t
- To focus on the ceteris paribus effect of z on y, we set the error term in each time period to zero. Then,

$$
\begin{aligned}
y_{t-1} &= \alpha_0 + \delta_0 c + \delta_1 c + \delta_2 c \\
y_t &= \alpha_0 + \delta_0(c+1) + \delta_1 c + \delta_2 c \\
y_{t+1} &= \alpha_0 + \delta_0(c+1) + \delta_1(c+1) + \delta_2 c \\
y_{t+2} &= \alpha_0 + \delta_0(c+1) + \delta_1(c+1) + \delta_2(c+1),
\end{aligned}
$$

  and so on.
- With the permanent increase in z, after one period, y has increased by $\delta_0 + \delta_1$, and after two periods, y has increased by $\delta_0 + \delta_1 + \delta_2$.
- This shows that the sum of the coefficients on current and lagged z, $\delta_0 + \delta_1 + \delta_2$, is the long-run change in y given a permanent increase in z.

A short example

- $q = 2$
- z increases permanently to $c + 1$ at time t
- To focus on the ceteris paribus effect of z on y, we set the error term in each time period to zero. Then,

$$
\begin{aligned}
y_{t-1} &= \alpha_0 + \delta_0 c + \delta_1 c + \delta_2 c \\
y_t &= \alpha_0 + \delta_0(c+1) + \delta_1 c + \delta_2 c \\
y_{t+1} &= \alpha_0 + \delta_0(c+1) + \delta_1(c+1) + \delta_2 c \\
y_{t+2} &= \alpha_0 + \delta_0(c+1) + \delta_1(c+1) + \delta_2(c+1),
\end{aligned}
$$

  and so on.
- With the permanent increase in z, after one period, y has increased by $\delta_0 + \delta_1$, and after two periods, y has increased by $\delta_0 + \delta_1 + \delta_2$.
- This shows that the sum of the coefficients on current and lagged z, $\delta_0 + \delta_1 + \delta_2$, is the long-run change in y given a permanent increase in z.

Examples of time series regression models

- For the FDL model with q lags, we can define the **LRP** accordingly by

$$\text{LRP} = \delta_0 + \delta_1 + \cdots + \delta_q.$$

Example

Consider the model

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

where gfr_t is the general fertility rate and pe_t is the real dollar value of the personal tax exemption. δ_0 measures the immediate change in fertility due to a one dollar increase in pe. If pe permanently increases by one dollar, then, after two years, gfr will have changed by $\delta_0 + \delta_1 + \delta_2$, and no further change afterwards.

Something more...

- Can have multiple explanatory variables in both static models and FDL models.
- Needs to have a propert treatment on the initial observations, **BUT**
  - Do not worry, as regression packages automatically keep track of these things.

Finite Sample Properties of OLS under Classical Assumptions

Unbiasedness of OLS

Assumption TS.1: Linear in parameters

The stochastic process $\{(\mathbf{X}_t, y_t)_t\}$ follows the linear model

$$y_t = \beta_0 + \mathbf{X}_t'\boldsymbol{\beta} + u_t,$$

where $(u_t)_t$ is the sequence of errors or disturbances.

Assumption TS.2: No perfect collinearity

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

Unbiasedness of OLS

Assumption TS.3: Zero conditional mean

For each t, the expected value of the error u_t, given the explanatory variables for all time periods $\mathbf{X}$, is zero:

$$\mathbb{E}\left(u_t|\mathbf{X}\right) = 0.$$

- No **random sampling** in TS context
- Assumption TS.3 implies that u_t must be uncorrelated with x_{sj} when $s = t$ and even $s \neq t$.

Going further questions 1

In the FDL model $y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + u_t$, what do we need to assume about the sequence $\{z_0, z_1, \cdots, z_T\}$ in order for Assumption TS.3 to hold?

A short example

- Consider the model

$$mrdrte_t = \beta_0 + \beta_1 polpc_t + u_t,$$

where $mrdrte_t$ is a city's murder rate and $polpc_t$ is police officers per capita.

A short example

- Consider the model
$$mrdrte_t = \beta_0 + \beta_1 polpc_t + u_t,$$
where $mrdrte_t$ is a city's murder rate and $polpc_t$ is police officers per capita.
- Suppose that Assumption TS.3 is satisfied:
    - It may be reasonable to assume that u_t is uncorrelated with $polpc_t$ and even with past values of $polpc_t$.

A short example

- Consider the model

$$mrdrte_t = \beta_0 + \beta_1 polpc_t + u_t,$$

  where $mrdrte_t$ is a city's murder rate and $polpc_t$ is police officers per capita.
- Suppose that Assumption TS.3 is satisfied:
  - It may be reasonable to assume that u_t is uncorrelated with $polpc_t$ and even with past values of $polpc_t$.
  - But suppose that the city adjusts the size of its police force based on past values of the murder rate.

A short example

- Consider the model

$$mrdrte_t = \beta_0 + \beta_1 polpc_t + u_t,$$

  where $mrdrte_t$ is a city's murder rate and $polpc_t$ is police officers per capita.
- Suppose that Assumption TS.3 is satisfied:
  - It may be reasonable to assume that u_t is uncorrelated with $polpc_t$ and even with past values of $polpc_t$.
  - But suppose that the city adjusts the size of its police force based on past values of the murder rate.
- This means that, say, $polpc_{t+1}$ might be correlated with u_t (because a higher u_t leads to a higher $mrdrte_t$).

A short example

- Consider the model
$$mrdrte_t = \beta_0 + \beta_1 polpc_t + u_t,$$
where $mrdrte_t$ is a city's murder rate and $polpc_t$ is police officers per capita.
- Suppose that Assumption TS.3 is satisfied:
    - It may be reasonable to assume that u_t is uncorrelated with $polpc_t$ and even with past values of $polpc_t$.
    - But suppose that the city adjusts the size of its police force based on past values of the murder rate.
- This means that, say, $polpc_{t+1}$ might be correlated with u_t (because a higher u_t leads to a higher $mrdrte_t$).
- If this is the case, Assumption TS.3 is generally violated.

Unbiasedness of OLS

- Assumptions TS. 1-3 are enough to guarantee the unbiasedness of OLS.

Theorem

Under Assumptions TS.1, TS.2, and TS.3, the OLS estimators are unbiased conditional on $\mathbf{X}$*, and therefore unconditionally as well when the expectations exist:* $\mathbb{E}\left(\hat{\beta}_{nj}\right) = \beta_j$*, where* $j = 0, 1, \cdots, k$*.*

- The proof is essentially the same as that for cross-sectional data, so we omit it.

The Variances of the OLS Estimators and the Gauss-Markov Theorem

Assumption TS.4: Homoskedasticity

Conditional on $\mathbf{X}$, the variance of u_t is the same for all t: $\mathbb{V}\left(u_t|\mathbf{X}\right) = \mathbb{V}(u_t) = \sigma^2$, $t = 1, 2, \cdots, T$.

Assumption TS.5: No Serial Correlation

Conditional on $\mathbf{X}$, the errors in two different time periods are uncorrelated: $\operatorname{corr}\left(u_t u_s|\mathbf{X}\right) = 0$, for all $t \neq s$.

The Variances of the OLS Estimators and the Gauss-Markov Theorem

Assumption TS.4: Homoskedasticity

Conditional on $\mathbf{X}$, the variance of u_t is the same for all t: $\mathbb{V}(u_t|\mathbf{X}) = \mathbb{V}(u_t) = \sigma^2$, $t = 1, 2, \cdots, T$.

Assumption TS.5: No Serial Correlation

Conditional on $\mathbf{X}$, the errors in two different time periods are uncorrelated: $\text{corr}(u_t u_s|\mathbf{X}) = 0$, for all $t \neq s$.

- No such Assumption TS.5 in the cross-sectional setting due to **random sampling**.

Example

Consider an equation for determining three-month T-bill rates $i3_t$ based on the inflation rate inf_t and the federal deficit as a percentage of gross domestic product def_t:

$$i3_t = \beta_0 + \beta_1 inf_t + \beta_2 def_t + u_t.$$

- Policy regime changes are known to affect the variability of interest rates.
- If interest rates are unexpectedly high for this period, then they are likely to be above average (for the given levels of inflation and deficits) for the next period.

Thus, Assumptions TS. 4-5 are unlikely to be satisfied in this example.

The Variances of the OLS Estimators and the Gauss-Markov Theorem

Theorem: OLS Sampling Variances

Under the time series Gauss-Markov Assumptions TS.1 through TS.5, the variance of $\hat{\beta}_{nj}$, conditional on $\mathbf{X}$, is

$$\mathbb{V}\left(\hat{\beta}_{nj}|\mathbf{X}\right) = \frac{\sigma^2}{\mathrm{SST}_j\left(1 - R_j^2\right)}, \;\; j = 1, 2, \cdots, k,$$

where SST_j is the total sum of squares of x_{tj} and R_j^2 is the R-squared from the regression of x_j on the other independent variables.

The Variances of the OLS Estimators and the Gauss-Markov Theorem

Theorem: Unbiased estimation of σ^2

Under Assumptions TS.1 through TS.5, the estimator

$$\hat{\sigma}^2 = \frac{\text{SSR}}{df},$$

is an unbiased estimator of σ^2, where $df = T - k - 1$.

Theorem: Gauss-Markov Theorem

Under Assumptions TS.1 through TS.5, the OLS estimators are the best linear unbiased estimators conditional on $\mathbf{X}$.

- "Best" implies that among all linear estimators, $\hat{\boldsymbol{\beta}}_n$ has the smallest variance.
- This implies that the matrix

$$\text{Var}\left(\tilde{\boldsymbol{\beta}}_n|\mathbf{X}\right) - \text{Var}\left(\hat{\boldsymbol{\beta}}_n|\mathbf{X}\right)$$

  is p.s.d.
- OLS yields the smallest variance. In particular,

$$\text{Var}\left(\hat{\beta}_{nj}|\mathbf{X}\right) \leq \text{Var}\left(\tilde{\beta}_{nj}|\mathbf{X}\right),$$

  for any other linear, unbiased estimator of β_j.

Inference under the Classical Linear Model Assumptions

Assumption TS.6: Normality

The errors u_t are independent of $\mathbf{X}$ and are independently and identically distributed as $\mathcal{N}(0, \sigma^2)$.

Theorem: Normal sampling distribution

Under Assumptions TS.1 through TS.6, the CLM assumptions for time series, the OLS estimators are normally distributed, conditional on $\mathbf{X}$. Further, under the null hypothesis, each t-statistic has a t distribution, and each F-statistic has an F distribution. The usual construction of confidence intervals is also valid.

Since

$$\hat{\boldsymbol{\beta}}_n = \boldsymbol{\beta} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{u}\right),$$

we have (with Assumption TS.6)

$$\hat{\boldsymbol{\beta}}_n | \mathbf{X} \sim \mathcal{N}\left(\boldsymbol{\beta}, \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right).$$

The hypothesis testing of the null

$$\mathcal{H}_0 : \beta_j = 0$$

can be based on the following t-statistic

$$\frac{\hat{\beta}_j}{\mathrm{se}\left(\hat{\beta}_{nj}\right)} \sim t_{n-k-1}.$$

The joint null

$$\mathcal{H}_0 : \mathbf{R}\boldsymbol{\beta} = 0$$

can also be based on the following F-statistic

$$\frac{\left(\mathrm{SSR}_r - \mathrm{SSR}_{ur}\right)/q}{\mathrm{SSR}_{ur}/\left(n-k-1\right)} \sim F(q, n-k-1).$$

Inference under the Classical Linear Model Assumptions

- If Assumptions TS.1 - TS.6 are satisfied, everything we have learned about estimation and inference for cross-sectional regressions applies directly to time series regressions.
- Restrictive as they may sound? Of course!
- Nevertheless, the CLM framework is a good starting point for many applications.

Functional form, dummy variables, and index numbers

Functional form

- All of the functional forms we learned about in earlier chapters can be used in time series regressions.
- The most important of these is the natural logarithm: time series regressions with constant percentage effects appear often in applied work.

Example

Consider a simple model

$$\log y_t = \beta_0 + \beta_1 d_t + u_t,$$

where d_t is a dummy variable. Simple algebra gives

$$\beta_1 = \log\left(\frac{y_{1t}}{y_{0t}}\right) = \log\left\{1 + \frac{y_{1t} - y_{0t}}{y_{0t}}\right\} = \log\{1 + \Delta\% y_p\} \approx \Delta\% y_p.$$

The exact percentage change is given by

$$\frac{y_{1t} - y_{0t}}{y_{0t}} = \exp(\beta_1) - 1.$$

Dummy variables

- Binary explanatory variables are the key component in what is called an **event study**.
- A simple version of an equation used for event studies is

$$R_t^f = \beta_0 + \beta_1 R_t^m + \beta_2 d_t + u_t,$$

where
  - R_t^f: stock return for firm f during period t
  - R_t^m: market return (usually computed for a broad stock market index)
  - d_t: equals to 1 if the event occurred

Example

Consider an airline firm, d_t might denote whether the airline experienced a publicized accident or near accident during time t.

Going further questions 2

Suppose that we would like to quantify the impact of releasing earning report on the stock market performance for firm f. How would you design such an event study?

Index number

- The index of industrial production (IIP) is a measure of production across a broad range of industries, and, as such, its magnitude in a particular year has no quantitative meaning.
- In order to interpret the magnitude of the IIP, we must know the **base period** and the **base value**.
- To change the base period for any index number, we can use the formula

$$newindex_t = 100(oldindex_t / oldindex_{newbase}).$$

Example

With base year 1987, the IIP in 1992 is 107.7; if we change the base year to 1982, the IIP in 1992 becomes $100(107.7/81.9) = 131.5$ (because the IIP in 1982 was 81.9).

Nominal and real economic variables

- Another important example of an index number is a price index, such as the **CPI**.
- The price indexes are necessary for turning a time series measured in *nominal dollars* (or *current dollars*) into *real dollars* (or *constant dollars*).

Example

Suppose that average weekly hours worked are related to the real wage as

$$\log(hours) = \beta_0 + \beta_1 \log(w/p) + u.$$

If we compare with specification using nominal wage

$$\log(hours) = \beta_0 + \beta_1 \log(w) + \beta_2 \log(p) + u,$$

this implies that economic theory imposes a restriction $\beta_2 = -\beta_1$ on the above model.

Trends and seasonality

Trending time series models

- Linear trend:

$$y_t = \alpha_0 + \alpha_1 t + e_t, \quad t = 1, 2, \cdots$$

- Exponential trend:

$$\log y_t = \alpha_0 + \alpha_1 t + e_t, \quad t = 1, 2, \cdots$$

  - How do we interpret β_1?

$$\Delta \log(y_t) \approx (y_t - y_{t-1}) / y_{t-1} = \beta_1,$$

    where the final equality is achieved by setting $\Delta e_t = 0$.

- Quadratic trend:

$$y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + e_t, \quad t = 1, 2, \cdots,$$

  - The approximate slope (holding e_t is fixed) is

$$\frac{\Delta y_t}{\Delta t} \approx \alpha_1 + 2\alpha_2 t.$$

Using Trending Variables in Regression Analysis

- Suppose that y_t has some trending pattern and consider the following model specification:

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + u_t.$$

- The trending pattern is now absorbed into the error term u_t, making Assumptions TS. 4-5 violated.
- OLS estimates $\hat{\beta}_{1T}$ and $\hat{\beta}_{2T}$ are likely to be biased.
- It is thus recommended to add a trending term if variables are themselves trending:

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_3 t + u_t.$$

Going further questions 3

Explain how to interpret β_1 and β_2 in the above specification based on F-W-L theorem.

Computing R-Squared When the Dependent Variable Is Trending

- Recall the definition of R^2:

$$R^2 = 1 - \left(\hat{\sigma}_u^2 / \hat{\sigma}_y^2\right),$$

  where
  - $\hat{\sigma}_u^2$: an unbiased estimator of error variance;
  - $\hat{\sigma}_y^2 = \mathrm{SST}/(T-1)$, $\mathrm{SST} = \sum_{t=1}^{T}(y_t - \overline{y})^2$.
- However, when y_t is trending, $\hat{\sigma}_y^2$ is no longer an unbiased or consistent estimator of σ_y^2.
- We can first obtain residuals $\ddot{y}_t$ from the regression of y_t on t. Then, we compute R^2 based on

$$R^2 = 1 - \frac{\hat{\sigma}_u^2}{\sum_{t=1}^{T} \ddot{y}_t^2}.$$

- Reminder: In computing the R-squared form of an F-statistic for testing multiple hypotheses, we just use the usual R-squareds without any detrending.

Seasonality

- Time series may exhibit seasonality.

Example

- Housing starts are generally higher in June than in January due to weather.
- Retail sales in the fourth quarter are typically higher than in the previous three quarters because of the Christmas holiday.

- Series that do display seasonal patterns are often **seasonally adjusted** before they are reported for public use.

Seasonality

- A simple method for dealing with seasonality is to add seasonal dummies in regression models:

$$y_t = \beta_0 + \delta_1 feb_t + \delta_2 mar_t + \cdots + \delta_{11} dec_t + \mathbf{x}_t' \boldsymbol{\beta} + u_t,$$

  where feb_t, mar_t, $\cdots$, and dec_t are monthly dummy variables.

- No seasonality implies that $\delta_1 = \delta_2 = \cdots = \delta_{11} = 0$, which can be easily tested via an F-test.

Going further questions 4

In the above equation, what is the intercept for March? Explain why seasonal dummy variables satisfy the strict exogeneity assumption.

- Can also **deseasonalizing** the data as we do for detrending.

Problems

Problem 1

Decide if you agree or disagree with each of the following statements and give a brief explanation of your decision:

1. Like cross-sectional observations, we can assume that most time series observations are independently distributed.
2. The OLS estimator in a time series regression is unbiased under the first three Gauss-Markov assumptions.
3. A trending variable cannot be used as the dependent variable in multiple regression analysis.
4. Seasonality is not an issue when using annual time series observations.

Problems

Problem 2

We say that the explanatory variables $\mathbf{x}_t = (x_{t1}, \cdots, x_{tk})$ are said to be *sequentially exogenous* (sometimes called *weakly exogenous*) if

$$\mathbf{E}\left(u_t | \mathbf{x}_t, \mathbf{x}_{t-1}, \cdots, \mathbf{x}_1\right) = 0, \quad t = 1, 2, \cdots,$$

so that the errors are unpredictable given current and all *past* values of the explanatory variables.

1. Explain why sequential exogeneity is implied by strict exogeneity.
2. Explain why contemporaneous exogeneity is implied by sequential exogeneity.

Assignment

- Problems: 2,5
- Computer Exercises: C2,C9