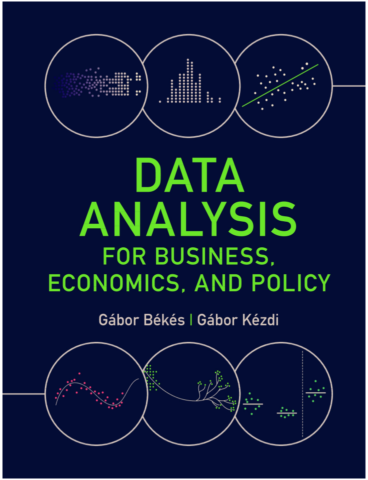
12. Time series regression

Gabor Bekes

Data Analysis 2: Regression analysis

2019

Slideshow for the Békés-Kézdi Data Analysis textbook



- Cambridge University Press, 2021 January
- Available in paperback, hardcover and e-book
- Slideshow be used and modified for educational purposes only
- **gabors-data-analysis.com**
  - Download all data and code
  - Additional material, links to references

Motivation 1

- Investing in a company stock, and you want to know how risky that investment is.
- In finance, a relevant measure of risk relates returns on a company stock to market returns: a company stock is considered risky if it moves in the direction of the market, and the more it moves in that direction the riskier it is.
- Data on daily stock prices for 21 years.
- How to define returns?
- How to assess whether and to what extent returns on the company stock move together with market returns?

Motivation 2

- People heat and cool in most places
- Heating and cooling are potentially important uses of electricity.
- How does weather conditions affect electricity consumption?
- Monthly data on temperature and residential electricity consumption in a hot region.
- What model to estimate, how best define variables?
- How best take into account seasonal patterns?

Time series and time series regressions

- Time series data is somewhat special
- Data preparation is a bit hard, need to make decisions

- Linear regression with time series data.
- Special features of time series regression
- Time series data presents additional opportunities as well as additional challenges to compare variables.
- Three key issues to deal with: trends, seasonality and serial correlation.

Time series regressions: data preparation

- Frequency of time series = time elapsed between observations of a variable
- Frequency may be yearly, monthly, weekly, daily, hourly, etc
- Practical problems with frequency

- Frequency may be irregular with gaps in-between.
- Often: ignore them, think as day1, day2, ...
- Sometimes it matters: weekends in financial markets may bring on news. Can add a dummy.

- Extreme values, spikes

Time series regressions: data preparation

- Regressions: to condition y on values of x in time series data the two variables need to be on the same frequency.
- When the frequency of y and x is different we need to adjust one of them. Most often - aggregating the variable at higher frequency (e.g., from weekly to monthly).
- Flow variables, such as sales, aggregation means adding up;
- Other kinds of variables, such as prices, it is often taking an average or picking one value
  - Stock price varies by transaction (e.g. second). Daily stock price is closing price on a given day.

Time series comparisons - S&P 500 case study

- Daily price of Microsoft stock and value of S&P 500 stock market index
- The data covers 21 years starting with December 31 1997 and ending with December 31 2018.
- Many decisions to make
- Look at data first

Case study: Stock price and stock market index value

Microsoft, daily close price

SP 500 index value, daily close

Time series comparisons - S&P 500 case study

- Daily price of Microsoft stock and value of S&P 500 stock market index
- The data covers 21 years starting with December 31 1997 and ending with December 31 2018.

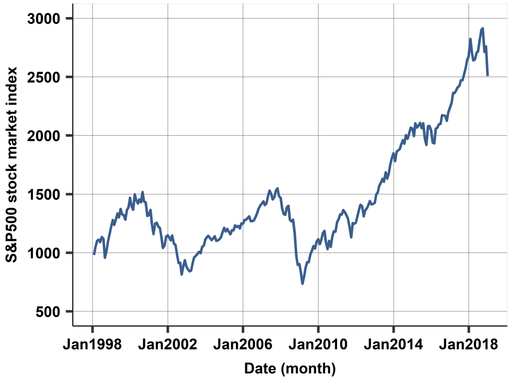
- Key decisions:
- Daily price = closing price
- Gaps will be overlooked
  - Friday-Monday gap ignored
  - Holidays (Christmas, 4 of July (when would be a weekday)
- All values kept, extreme values part of process

Time series comparisons - S&P 500 case study

- In finance, portfolio managers often focus on monthly returns - this is the time horizon for which performance are measured and communicated to clients.
- Hence, we choose monthly returns to analyze.
- Take the last day of each month

Case study: Stock price and stock market index value

Microsoft, monthly price



S&P 500 index value, monthly close

What is special in time series

- Time series regressions is special for several reasons.
- Many aspects of regression analysis remains.
  - Generalization, confidence intervals
  - Time series regression uncover patterns rather than evidence of causality
  - Practical data issues, missing observations, extreme values etc, remain
  - Coefficient interpretation is based on conditional comparison

What is special in time series

- Ordering matters – key difference to cross section
- Complications...
- Trend - variables for later time periods will tend to be higher (lower)
- Seasonality - cyclical component, such 4 seasons, months, - every e.g. December value is expected to be different.
- Time series values are often not independent

What is special in time series: Trend

Define change (or fist difference): $\Delta x_t = x_t - x_{t-1}$

$$\text{Positive trend: } E[\Delta x_t] > 0 \tag{1}$$

$$\text{Negative trend: } E[\Delta x_t] < 0 \tag{2}$$

- A time series variable follows a *positive trend* if its change is positive on average. It follows a *negative trend* if its change is negative on average
- Trend is *linear* if the change is the same on average.
- Trend is *exponential* if the change in the log of the variable is the same on average.

$$\text{Linear trend: } E[\Delta x_t] = constant \tag{3}$$

$$\text{Exponential trend: } E[\Delta ln(x_t)] = constant \tag{4}$$

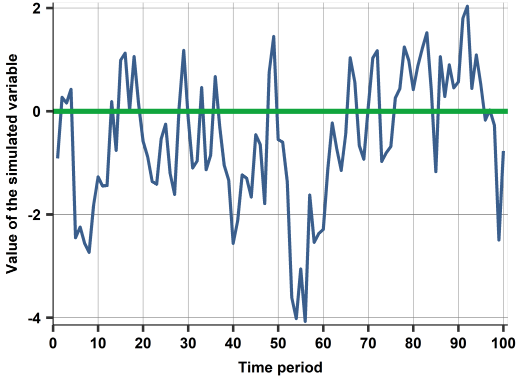
What is special in time series: Seasonality

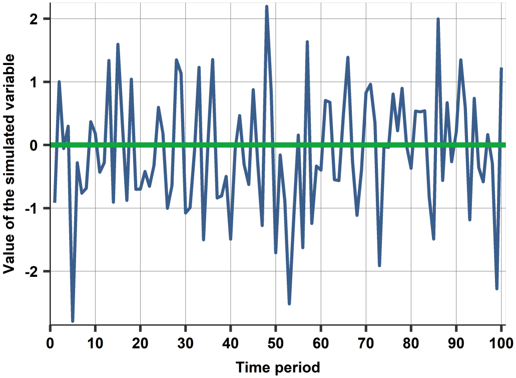
- There is seasonal variation, or simply *seasonality*, in a time series variable if its expected value changes periodically.
- Follows the seasons of the year, days of the week, hours of the day.
- Seasonality may be linear, when the seasonal differences are constant; it may be exponential, if relative differences (that may be approximated by log differences) are constant.
- Important real life phenomenon - many economic activities follow seasonal variation over the year, through the week or day.

What is special in time series: Serial correlation

- Serial correlation means correlation of a variable with its previous values
- The 1st order serial correlation coefficient is defined as $\rho_1 = Corr[x_t, x_{t-1}]$
  - the 2nd order serial correlation coefficient is defined as $\rho_2 = Corr[x_t, x_{t-2}]$;
- For a *positively serially correlated* variable, if its value was above average last time, it is more likely that it is above average this time, too.
- $\rho_1 = 0$ - no serial correlation. "White Noise"
  - Like cross-section, order does not matter.
  - Example?

Two simulated series: rho=0.8 (left), rho=0 (right)





What is special in time series: Stationarity

- Stationarity = a feature of the series itself. Key new concept.
- **Stationary time series** have the same expected value and same distribution, at all times.
- Stationarity means stability (in expectations).

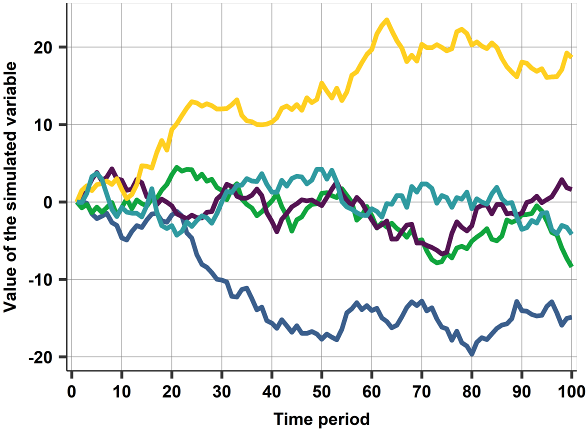
- **Non-stationary** time series are those that are not stable for some reason.
- Trends and seasonality violate stationarity because the expected value is different at different times.
- Unstable patterns also lead to non-stationary series

What is special in time series: Stationarity

- Another example of nonstationary time series is the *random walk*.
- Random walk when $\rho = 1$ – also called a unit root.
- Time series variables that follow random walk change in completely random ways.
- Whatever the previous change was the next one may be anything. Wherever it starts, a random walk variable may end up anywhere after a long time.

What is special in time series: Random walk

- 5 simulated random walk series
- Each random walk series wanders around randomly.
- Further and further away as time passes



What is special in time series: Random walk

- Random walks are impossible to predict
- after a change, they don't revert back to some value or trend line but continue their journey from that point.
- Spread rising from one interval to another

- For stationary series, we need stability of patterns
- Avoid series with random walk when running regressions

What is special in time series: Unit root

- Testing is complicated. FYI
- Phillips-Perron test is based on this mode:

$$x_t = \alpha + \rho x_{t-1} + e_t \tag{5}$$

- This model represents a random walk if $\rho = 1$ (with drift if $\alpha \neq 0$)
- The Phillips-Perron test has hypothesis $H_0 : rho = 1$ against the alternative $H_A : rho < 1$.
- Statistical software calculate the p-value for this test.
- When the p-value is large (e.g., larger than 0.05), we don't reject the null, concluding that the time series variable follows a random walk (perhaps with drift).

What is special in time series: Trends and seasonality

- Stationary series are those where the expected value does not change, variance does not change over time: two observations at different points in time have the same mean and variance.
- A series is stationary if all time intervals are similar in this sense.
- We have seen three examples of non–stationarity:
  - Trend - Expected value is different in later time periods than in earlier time periods
  - Seasonality - Expected value is different in periodically recurring time periods
  - Random walk and similar series – Variance keeps increasing over time
- We care about this because regression with time series data variables that are not stationary are likely to give misleading results.

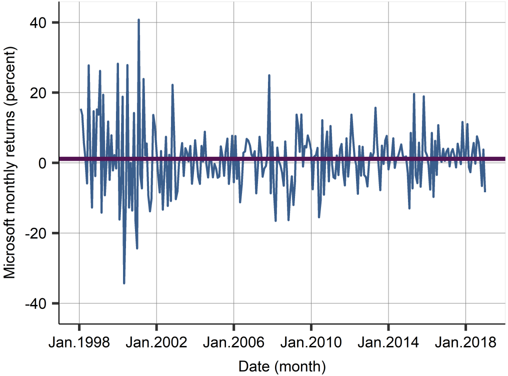
Returns on a company stock and market returns

- Started with looking at prices
- Prices series are random walk
- They have a unit root – using the Phillips-Perron test, we find a very high p-value (and go for random walk if p>0.05), we are very certain that process is random walk–> need action
- Need to use difference (= return)
- A: First difference of log price
- B: Percent change – choose this as more used in Finance

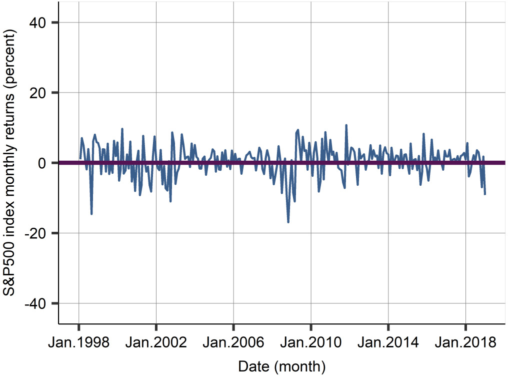
Returns on a company stock and market returns

- Take percent return
- Correlation in time series show visually
- We can estimate the regression formally

Case study: Stock price and stock market index return (pct)

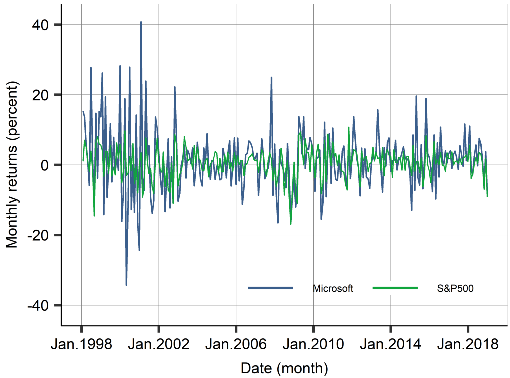


Microsoft, monthly return (pct)

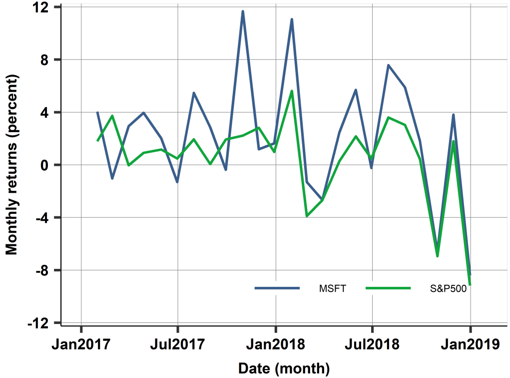


S&P 500 index value, monthly return (pct)

Case study: Stock and market returns over time (pct)



The entire time series, 1998-2018



2017-18

Returns on a company stock and market returns

- Correlation in time series: the price of the Microsoft stock tends to increase when market prices increase, and it tends to decrease when market prices decrease.
- Market changes are smaller
- Focus on two years, we can see it better
- We can estimate the regression formally
  - Monthly
  - Percent return

Returns on a company stock and market returns

$$pctchange(MSFT_t) = \alpha + \beta pctchange(SP500_t) \tag{6}$$

- $\alpha = 0.54; \beta = 1.26$

Returns on a company stock and market returns

$$pctchange(MSFT_t) = \alpha + \beta pctchange(SP500_t) \tag{6}$$

- $\alpha = 0.54; \beta = 1.26$
- Intercept: returns on the Microsoft stock tend to be 0.54 percent when the S%P500 index doesn't change.
- Slope: returns on the Microsoft stock tend to be 1.26% higher when the returns on the S&P500 index are 1% higher.
- The 95% confidence interval is $[1.06, 1.46]$.
- R-squared: 0.36
- First difference of log prices. Estimate is 1.24
- Daily returns (percent), beta is 1.10

Returns on a company stock and market returns

- Slope is actually the well-known "beta" in finance
- Beta - measure of the riskiness of the company stock.
  - Close to one?
  - Greater than one?
  - Positive, less than one?
  - Negative?

Returns on a company stock and market returns

- We have seen challenges that make time series regression more complicated
- Now let's review what we do
- It will turn out to be simple...

Time series regressions

- Regression in time series data is defined and estimated the same way as in other data.

$$y_t^E = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots \tag{7}$$

- Interpretations similar to cross-section
- β_0: We expect y to be β_0 when all explanatory variables are zero.
- β_1: Comparing time periods with different x_1 but the same in terms of all other explanatory variables, we expect y to be higher by β_1 when x_1 is higher by one unit.

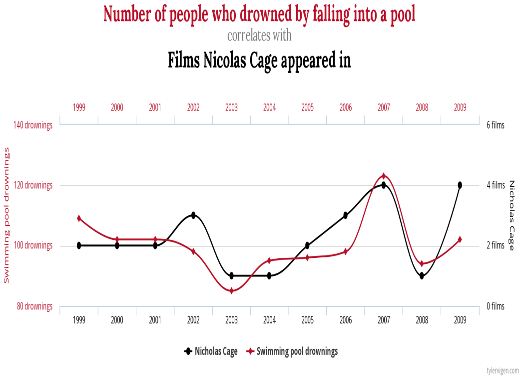
Time series regressions - list of issues

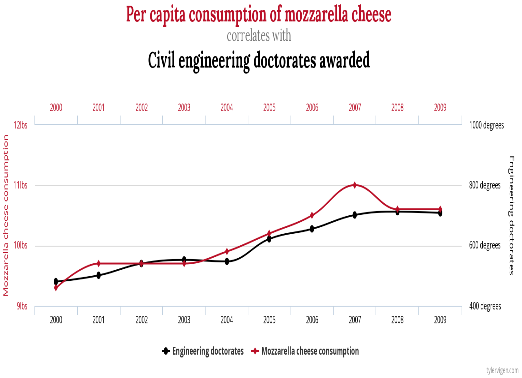
- Handling trend and seasonality
- Checking and dealing with unit roots
- Transforming the series, such as taking first differences
- Dealing with serial correlation (in y_t) – specifying the proper standard errors
- Considering lags

Time series regressions: Trends and seasonality

- Trends, seasonality, and random walks can present serious threats to uncovering meaningful patterns in time series data.
- Example: time series regression in levels $y_t^E = \alpha + \beta x_t$.
- If both y and x have a positive trend, the slope coefficient β will be positive whether the two variables are related or not.
- That is simply because in later time periods both tend to have higher values than in earlier time periods.
- Associations between variables only because of the effect of trends are said to be **spurious correlation**

Correlated time series. But....





These and similar graphs from `http://tylervigen.com/spurious-correlations`

Time series regressions: Trends and seasonality

- Spurious - could be very far fetched reason, randomness
  - especially with small samples!

- One frequent reason: trend and seasonality as confounders.
- Trend/seasonality is a confounder if both y_t and x_t have trend / seasonal variation.
- If not included while they should be - omitted variables

- A trend may capture omitted global tendencies in population growth, economic activity, fashion, technology.
- A seasonality may capture variation in weather, holidays and leisure time, sleeping and eating habits, open and close time of shops, etc.

Time series regressions: Trends and seasonality

- Example, a regression of the price of college education in the U.S. on the GDP of Germany over the past few decades would result in a positive slope coefficient even though that two may not be related in any fundamental way.

Time series regressions: Trends and seasonality

- A good solution to trends is replacing variables in the regression with their first differences
- Variables in differences do not have trends and are therefore more likely to to be stationary.
  - Could be log difference for exponential trends

- A good solution to seasonality is including *binary season variables* in regressions.
  - Look at pattern, figure out if quarters, months, weeks, days of week, etc.
- Another good solution to handle seasonality is working with year-on-year changes instead of first differences.

Time series regressions – first difference

We use the Δ notation to denote a first difference:

$$\Delta y_t = y_t - y_{t-1} \tag{8}$$

A linear regression in differences is the following

$$\Delta y_t^E = \alpha + \beta \Delta x_t \tag{9}$$

- Coefficients same interpretation as before, but use "when"
- α is the average left-hand-side variable when all right-hand-side variables are zero,
- β shows the difference in the average left-hand-side variable for observations with different Δx_t.

Time series regressions – first difference

$$\Delta y_t^E = \alpha + \beta \Delta x_t$$

- Because variables denote changes...
- α is the average change in y when x doesn't change.
- The slope coefficient on Δx_t shows how much more y is expected to change when x changes by one more unit.
- "more" – needed as we expect y to change anyway, by α, when x doesn't change.
  - The slope shows how y is expected to change when x changes, in addition to α.

Practice of time series regressions

- If you think there is a simple stable trend, having levels and a simple trend variable can be a solution. Rarely the case
- For most applications, time series regression involving using differences or log differences.
- Take differences unless you have a good reason not to.
  - One such case is when your variable is already a difference, GDP growth = difference of levels of GDP in percentage

Practice of time series regressions

- Capturing seasonality also important
- Higher frequency – the more important
  - People behave differently on different hours and days
  - Weather varies over months
  - Holidays, ect

- Have seasonal dummies if seasonality is stable. Often good enough
- Pattern may vary over time. If it does, solutions must capture exact pattern – difficult
  - Example?

Time series regressions: Standard errors

- Serial correlation makes the usual standard error estimates wrong.

Time series regressions: Standard errors

- Serial correlation makes the usual standard error estimates wrong.
- When the dependent variable is serially correlated - heteroskedasticity robust SE is wrong - sometimes very wrong, with a large bias.
  - More precisely it is serial correlation in residuals, but think about is as serial correlation in y_t is okay
- Use new SE - the **Newey-West** SE
  - procedure incorporates the structure of serial correlation of the regression residuals
  - Fine if heteroskedasticity as well
  - Need to specify lags. If enough data, frequency and seasonality should help, Months - 12 should be good
  - An alternative solution is to have lagged dependent variable in the regression

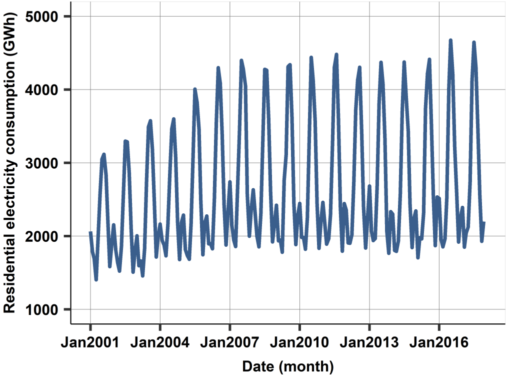
Electricity consumption and temperature

- Monthly weather and electricity data for Phoenix, Arizona
- January 2001 and ends in Dec 2017– 204 month
- The weather data includes "cooling degree days" and "heating degree days" per month.
- Cooling degree days and heating degree days are daily temperatures transformed in a simple way and then added up within a month.

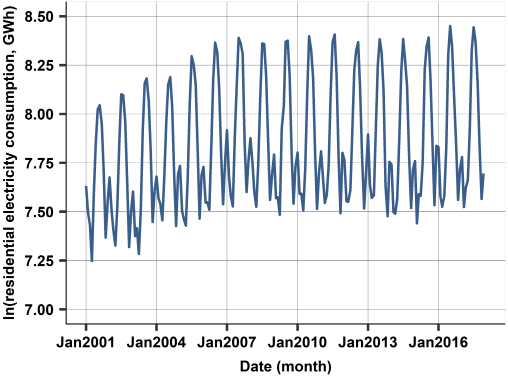
Electricity consumption and temperature

- The cooling degree days measure takes the average temperature within each day, subtracts a reference temperature (65F, or 18C), and adds up these daily values.
- If the average temperature in a day is, say, 75F (24C), the cooling degree is 10F (6C). This would be the value for one day.
- Then we would calculate the corresponding values for each of the days in the month and add them up.
  - Days when the average temperature is below 65F have zero values.
- For heating degree days it's the opposite: zero for days with 65F or warmer, and the difference between the daily average temperature and 65F when lower.
  - For example, with 45F (7C), the heating degree is 20F (11C).

Electricity consumption and temperature

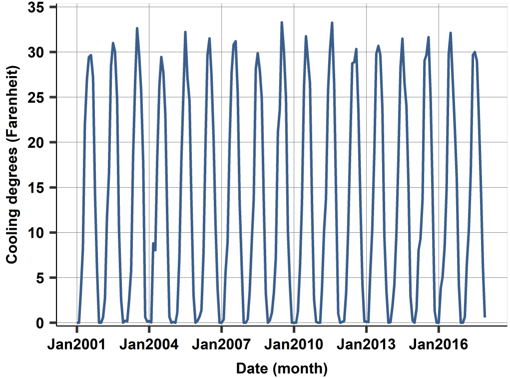


Electricity consumption

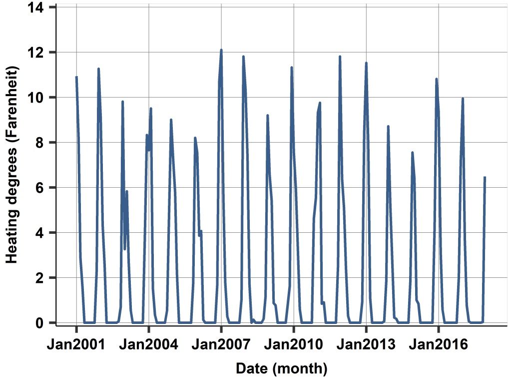


Log electricity consumption

Electricity consumption and temperature



Average cooling degrees



Average heating degrees

Electricity consumption and temperature

- No unit root.
- There is a trend in electricity for sure, exponential –> log difference
- For easier interpretation, take FD of cooling days and heating days.
- Natural question: How much does electricity consumption change when temperature changes?

- In this example, taking first difference does not make a huge difference, would not be a mistake to keep in levels
  - Another option could be to take 12-month difference

- Add monthly dummies, January (December to January) is reference
- Newey-West standard errors in parentheses; ** $p < 0.01$, * $p < 0.05$

Electricity consumption and temperature

VARIABLES	(1) ΔlnQ	(2) ΔlnQ
ΔCD	0.031**	0.017**
	(0.001)	(0.002)
ΔHD	0.037**	0.014**
	(0.003)	(0.003)
month = 2, February		-0.274**
month = 3, March		-0.122**
....		
month = 7, July		0.058**
month = 8, August		-0.085**
month = 9, September		-0.176**
....		
month = 12, December		0.067**
Constant	0.001	0.092**
	(0.002)	(0.013)
Observations	203	203

Electricity consumption and temperature

- In months when cooling degrees increase by one degree and heating degrees do not change, electricity consumption increases by 3.1 percent, on average.
    - When heating degrees increase by one degree and cooling degrees do not change, electricity consumption increases by 3.7 percent, on average.
- Monthly dummies matter, reduce slope coefficient estimates
- How to think about monthly dummies?
- Monthly dummies may be interpreted. Not easy.

Electricity consumption and temperature

- The reference month is January;
- constant (when cooling and heating degrees stay the same), electricity consumption increases by about 9% from December to January.
- The other season coefficients compare to this change.
- February – the January to February change is 28 percentage points lower than in the reference month, December to January.
- That was +9%, so electricity consumption decreases by about 19% on average to February from January when cooling and heating degrees stay the same.

Time series regressions: changes and lags

- Useful tool, potential causal scenario where changes take an impact in several periods later

$$\Delta y_t^E = \alpha + \beta_0 \Delta x_t + \beta_1 \Delta x_{t-1} + \beta_2 \Delta x_{t-2} \tag{10}$$

- Coefficients – how y is expected to change after a one-time change in x, i.e., when x changes in one time period *but not afterwards*.
- β_0 shows the contemporaneous association: what to expect in the same time period.
- β_1 shows the once-lagged association: what to expect in the next time period.

Time series regressions: Lags

$$\Delta y_t^E = \alpha + \beta_0 \Delta x_t + \beta_1 \Delta x_{t-1} + \beta_2 \Delta x_{t-2}$$

- β_0 = how many units more y is expected to change within the same time period when x changes by one more unit (and it didn't change in the previous two time periods).
- β_1 = how much more y is expected to change *in the next time period* after x changed by one more unit – provided that it didn't change at other times.
- Cumulative effect?

Time series regressions: Lags

$$\Delta y_t^E = \alpha + \beta_0 \Delta x_t + \beta_1 \Delta x_{t-1} + \beta_2 \Delta x_{t-2}$$

- $\beta_0 =$ how many units more y is expected to change within the same time period when x changes by one more unit (and it didn't change in the previous two time periods).
- $\beta_1 =$ how much more y is expected to change *in the next time period* after x changed by one more unit – provided that it didn't change at other times.

- Cumulative effect?

$$\beta_{cumul} = \beta_0 + \beta_1 + \beta_2 \tag{11}$$

Time series regressions: Lags

- To get a SE on the cumulative effect, do a trick and transformation, and estimate a different model

$$\Delta y_t^E = \alpha + \beta_{cumul} \Delta x_{t-2} + \delta_0 \Delta(\Delta x_t) + \delta_1 \Delta(\Delta x_{t-1}) \tag{12}$$

- the β_{cumul} in this regression is exactly the same as $\beta_0 + \beta_1 + \beta_2$ in the previous regression.
  - Other two right-hand-side variables strange and we do not care
- Typically estimate both. One with lags to see patterns. One with cumulative second to test the cumulative value.

Time series regressions: choosing lags

- Lag selection is a practical question
- Think about theory, domain knowledge. This may drive your call.
- Try out a few lags. Few depends on the size of your dataset.
  - Few dozen observations - need to be picky
  - 10-20 years of monthly data, can try all months
- watch for seasonality. Often need lags to capture 12 months, 4 quarters, etc.
- Try a few versions. Choose based on coefficient significance.

Electricity consumption and temperature

- Go back to model
- Add 2 lags - for both cooling and heating days
- And keep monthly dummies

Electricity consumption and temperature

VARIABLES	(1) $\Delta \ln Q$	(2) $\Delta \ln Q$
ΔCD	0.020**	
	(0.002)	
ΔCD 1st lag	0.006**	
	(0.002)	
ΔCD 2nd lag	0.001	
	(0.002)	
ΔHD	0.019**	
	(0.003)	
ΔHD 1st lag	0.011**	
	(0.003)	
ΔHD 2nd lag	0.000	
	(0.003)	
ΔCD cumulative coeff		0.027**
		(0.005)
ΔHD cumulative coeff		0.030**
		(0.007)

Electricity consumption and temperature

- Interestingly evidence of lagged effect
- Cumulative effect is now slightly larger.

- Not straightforward answer why
  - People take time to react to weather change
  - Or captures some correlated other variable

- Overall: Temperature is strongly associated with residential electricity consumption in Arizona.
- Even when seasonality is captured

Weather and electricity

How is residential electricity consumption related to weather in Arizona?

Files:

- ch12_arizona_electricity (.do / .R)

Weather and electricity

Data:

- Residential electricity consumption in Arizona, US (US Energy Information Administration (EIA))
  - monthly data
  - state-level
  - 2001-2018
- Temperature data (National Oceanic and Atmospheric Administration (NOAA))
  - monthly data
  - weather station level (100 stations in Arizona, picked one: Phoenix Airport, which is close to most population)
  - 1989-2018 in total, but coverage varies a lot by station

Data preparation

Cross-sectional unit:

- Discrepancy between level of aggregation in two datasets:
  - electricity: Arizona state
  - temperature: Phoenix Airport
- 60% of state population lives in Phoenix metropolitan area
- 2nd and 3rd largest cities are also located close to Phoenix

Frequency:

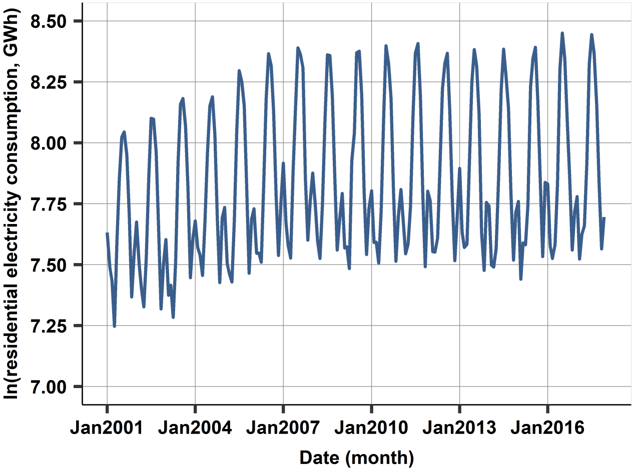
- Everything is at monthly level
- Combined data covers January 2001 - December 2017 (204 months)

Measure of heating / cooling degree days

We need a measure of how hot or cold days are and how likely it is that people use electricity for heating / cooling.

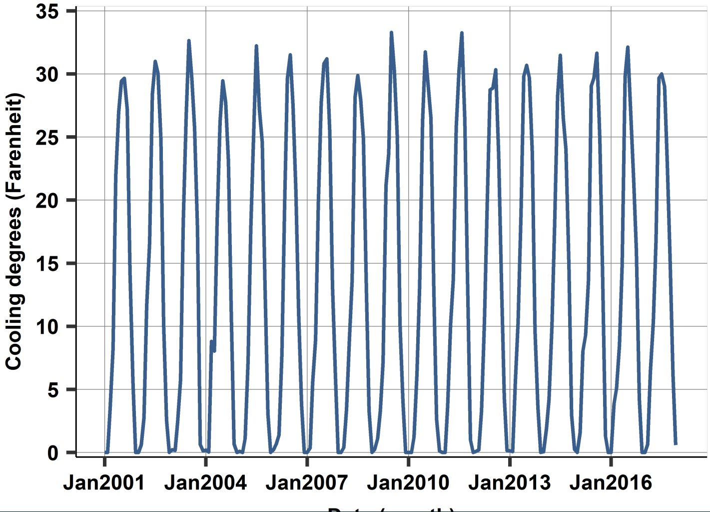
- reference temperature: 18C (65F)
- cooling degree days:
  - take average temperature each day,
  - subtract reference temperature,
  - calculate average of all these within a month (count below-18C as 0)
  - e.g. avg. temp. in a day is 24C (75F), then the cooling degree is 6C (10F)
- heating degree days:
  - take average temperature each day,
  - subtract FROM reference temperature,
  - calculate average of all these within a month (count above-18C as 0)
  - e.g. avg. temp. in a day is 7C (45F), then the heating degree is 11C (20F)

Residential electricity consumption in Arizona

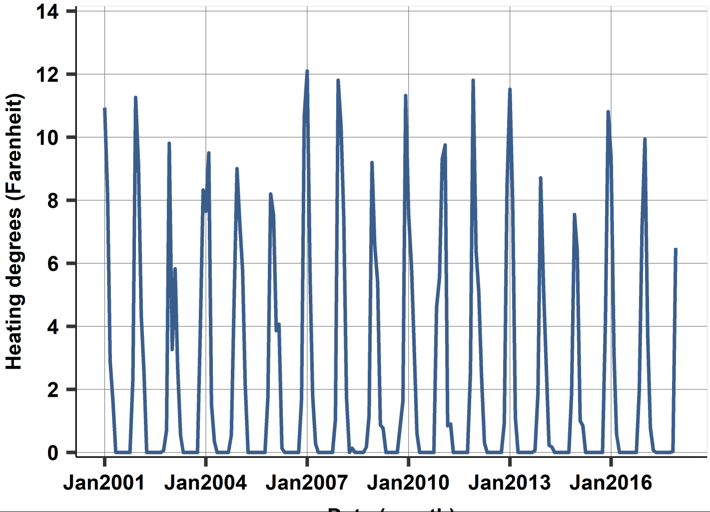


- We use logs
  - for easier interpretation
  - for statistical reasons: in level terms, the variance was increasing over time
- Patterns:
  - upward trend until 2008
  - strong seasonality (highest consumption is in summertime, there is a smaller peak during winter)

Cooling degree days (F)



Heating degree days (F)



Time-series Regression Models

$$lnQ_t^E = \beta_0 + \beta_1 CD_t + \beta_2 HD_t \tag{13}$$

$$\Delta lnQ_t^E = \gamma_0 + \gamma_1 \Delta CD_t + \gamma_2 \Delta HD_t \tag{14}$$

Results of time-series regressions

Results look similar in the two models, but interpretation is different.
Model in levels:

- Average residential electricity consumption in Arizona **is 3.2 percent higher in months with one higher** cooling degree in Phoenix (in Fahrenheit), comparing months with the same heating degrees.
- Average electricity consumption **is 4.5 percent higher in months with one higher** heating degree, comparing months with the same cooling degrees.

Model in differences:

- In months, when cooling degrees **increase by one degree** and heating degrees do not change, electricity consumption **increases by 3.1 percent more**, on average.
- When heating degrees **increase by one degree** and cooling degrees do not change, electricity consumption **increases by 3.7 percent more**, on average.

Handling trend and seasonality

If both LHS and RHS variables have trend and/or seasonality we might see a spurious relationship (e.g. age of US spending on science and technology and number of divorces).

→ We need to get rid of them.

- Trend appears only in LHS variable, so it does not matter,
- but seasonality is true for all variables in the model, it might drive our results

Solution: we include binary variables for every months

Regression results

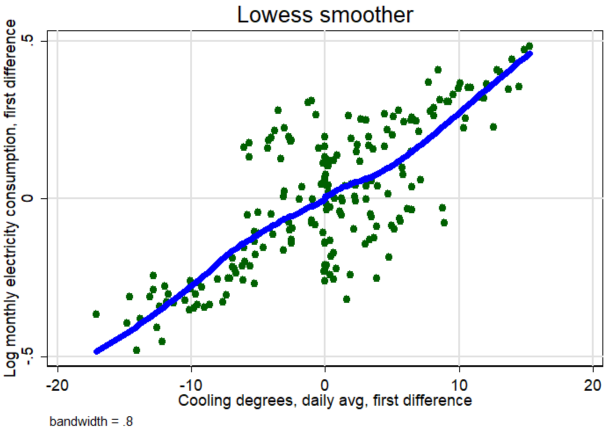
- Comparing the same months across years in Arizona, electricity consumption **is 1.8 percent higher** when cooling degrees **are one degree Fahrenheit higher** (and heating degrees are the same).
- Comparing the same months across years in Arizona, electricity consumption **is 1.5 percent higher** when heating degrees **are one degree Fahrenheit higher** (and cooling degrees are the same).
- Model in differences has very similar results, so it seems potential trends don't matter here
- Coefficient estimates are substantially lower than in the original model:
  - Part of the association is attributable to months as opposed to temperature itself.
  - The part that months take away from the temperature coefficients is larger in winter months.

Can we talk about causality here?

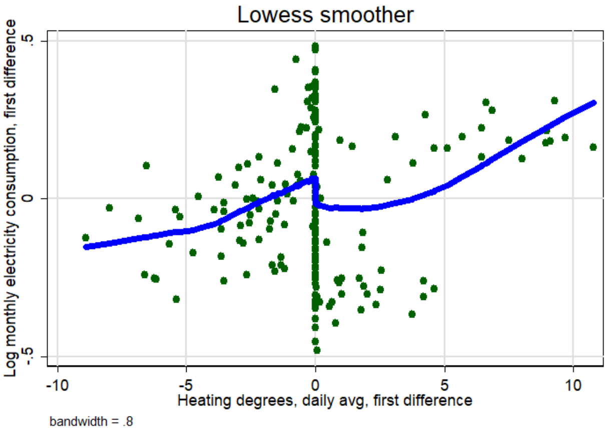
- Does electricity use affect temperature?
  - Not really plausible in the short run
- Is there a third variable that explains both?
  - E.g. daylight, different activities by season (more time at home during holidays)
  - mostly captured by months dummies

We can be kind of convinced that comparing electricity usage in the same months across years will capture the causal effect of temperature.

Potential nonlinearities



Potential nonlinearities



Correct Standard Errors

Serial correlation makes the usual standard error estimates wrong.

Two strategies to get correct standard error estimates:

- Newey-West standard errors (include a full period of seasonal variation)
- include lags of dependent variable

We do both in our example.

Estimation results with corrected S.E.

- The Newey-West standard error estimates are slightly larger for the regression in levels than the simple standard error estimates were. For the regression in differences they are practically the same.
  - The reason is that in the level-regression there is serial correlation, whereas in the diff-regression we don't see serial correlation.
- In level model there is a big difference in S.E. in Newey West and in lag model, whereas in the diff model, they are the same.
  - This is also due to the presence of serial correlation in the level model.

Estimate cumulative effects

$$\Delta y_t^E = \alpha + \beta_0 \Delta x_t + \beta_1 \Delta x_{t-1} + \beta_2 \Delta x_{t-2} \tag{15}$$

$$\Delta y_t^E = \alpha + \beta_{cumul} \Delta x_{t-2} + \delta_0 \Delta(\Delta x_t) + \delta_1 \Delta(\Delta x_{t-1}) \tag{16}$$

Main lessons learnt

- Temperature explains a large part of electricity consumption, i.e. hotter than average summers and cooler than average winters lead to substantially higher electricity consumption.
  - Months matter on their own right as well.
- We had to deal with the strong seasonality in both electricity consumption and temperature.
  - We included month binary variables, and the estimated coefficients became smaller (about half the original for cooling degree days, and about one third the original value for heating degree days)
- If there is serial correlation in the dependent variable, we need to adjust standard error estimation.
  - Most general solution is to use Newey-West standard errors.
  - We saw it does matter, when we have serial correlation.

Time series regressions: Summary of the process

- Decide on frequency; deal with gaps if necessary.
- Plot the series. Identify features and issues.
- Handle trends by transforming variables (Often: first difference).
- Specify regression that handles seasonality, usually by including season dummies.
- Include or don't include lags of the right-hand-side variable(s).
- Handle serial correlation.
- Interpret coefficients in a way that pays attention to potential trend and seasonality.

- Time series econometrics very complicated beyond this
- But: These steps often good enough