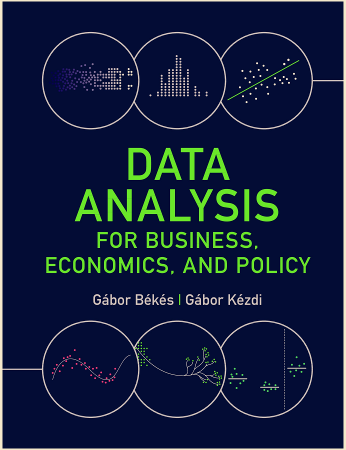
Békés-Kézdi: Data Analysis, Chapter 18: Forecasting from Time Series Data



Data Analysis for Business, Economics, and Policy

Gábor Békés (Central European University)
Gábor Kézdi (University of Michigan)

Cambridge University Press, 2021

gabors-data-analysis.com

Central European University

Any comments or suggestions:
gabors.da.contact@gmail.com

Intro

"It is difficult to make predictions, especially about the future."

(Niels Bohr; possibly an old Danish proverb)

Intro

"It is difficult to make predictions, especially about the future."

- It is fairly easy to make forecasts for variables that have a decently stable pattern over time
  - Weekly FMCG product sales, sensor of gadgets used regularly
- It is very hard to make forecasts of really interesting variables unexpected stuff may happen
  - GDP with Recession, crisis,
  - High tech gadgets with new products
  - Stock market prices

Forecasting basics

- Forecasting is a special case of prediction.
- Forecasting makes use of time series data on y, and possibly other variables x.
- The original data used for forecasting is a time series from 1 through T, such as $y_1, y_2, ..., y_T$
- The forecast is prepared for time periods after the original data ends, such as $\hat{y}_{T+1}, \hat{y}_{T+2}, ..., \hat{y}_{T+H}$. This is the live time series data.

Forecast horizon

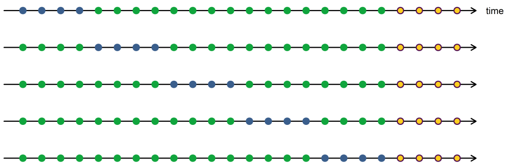
- The length of the live time series data (here H) = the forecast horizon.
- Short-horizon forecasts are carried out for a few observations after the original time series;
  - 5-10 years of monthly data, forecast a 3-12 months ahead
  - 10 years of quarterly data, predict ahead of a few quarters
- Long-horizon forecasts are carried out for many observations.
  - Often: data on activity, operation
  - 5 years of daily data, forecast daily ahead for a year
  - 2 months of hourly activity data, predict weeks ahead

Cross-validation in time series

- Cross-validation with time series is necessary
- It is tricky

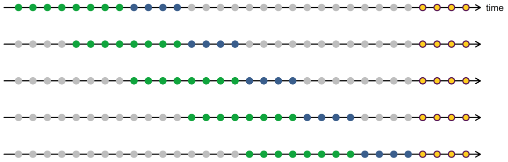
Cross-validation - option 1: test within data

- Forecast period relatively long compared to data.
- example: predict for 1 year, data 6 years
- long run serial correlation, trend less of an issue
- Insert test sets + use all remaining observations for the training sets
- Green: training set, Blue: test set, Yellow: holdout set



Cross-validation - option 2 Rolling window

- Rolling windows. Training set only before the test set
- Forecast period relatively short compared to data.
- example: predict for 12 months, data 15 years
- Serial correlation matters
- Green: training set, Blue: test set, Yellow: holdout set



Cross-validation in time series

Consider a forecast for horizon H (e.g., 12 months, or 365 days).

- For the holdout set, reserve the last H time periods in the original data and use the rest as the work set.
- From the work set, select k test sets as non overlapping complete time series of length H.
- For each test set, select the corresponding training set in one of the following two ways:
  - for long-horizon forecasts that don't use serial correlation, select all other observations, including those after the test set;
  - for short-horizon forecasts that use serial correlation, select the time series preceding the test set, in such a way that all training sets are of equal length.

Long-horizon: Seasonality, trend and predictable events

Long-horizon forecasting: Seasonality and predictable events

- Look for aspect of data that matter for long time
- Focus on predictable aspects of time series
- Trend(s) + Seasonality + Other regular events

- Two options to model trend: estimate average change or trend line

- Seasonality – especially true for forecasts with a daily or higher frequency such as hours or minutes
- Seasonality: model with set of variables (11 months), maybe interactions
- Other regular events - set of binary vars

Long-horizon forecasting: Trends - option 1

- First model - estimate average change

$$\widehat{\Delta y} = \hat{\alpha} \tag{1}$$

- For prediction this means

$$\hat{y}_{T+1} = y_T + \widehat{\Delta y}$$
$$\hat{y}_{T+2} = \hat{y}_{T+1} + \widehat{\Delta y} = y_T + 2 \times \widehat{\Delta y}$$
$$\ldots$$
$$\hat{y}_{T+H} = y_T + H \times \widehat{\Delta y}$$

Long-horizon forecasting: Trends - option 2

- Estimate trend line
- The simplest trend line is linear in time, with an intercept and a slope multiplying the time variable:
$$\hat{y}_t = \hat{\alpha} + \hat{\delta} t \tag{2}$$
- $\hat{\alpha}$ is predicted y when $t = 0$
- $\hat{\delta}$ tells us how much predicted y changes if t is increased by one unit.

Long-horizon forecasting: Trends - compare options

- Difference in models
- Model changes: assume that y continues from the last observation and increase by the same amount each time.
- If last observation unusually large or small y value, a trend modeled as change would continue from that unusual value.

- Model trend line, we assume that y remains close to the trend line. Last unusual observation would not matter for the forecast, because it would be the trend line.
- Neither approach is inherently better than the other

Long-horizon forecasting: Seasonality

- Capture regular fluctuations
- Months, days of the week, hours, combinations

Case study: ABQ swimming

- Swimming pool data
- Albuquerque (ABQ), New Mexico, USA
- Big data, transaction level entry data logged from sales systems
- 1.5m observations

Case study: ABQ swimming

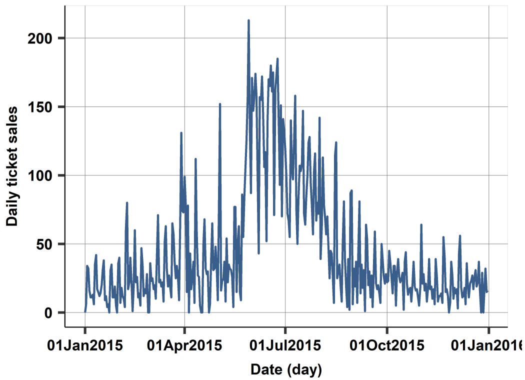
- Sample: Single swimming pool
- Aggregated: number of ticket sales per day
- After some sample design - regular tickets only

Case study: Modelling

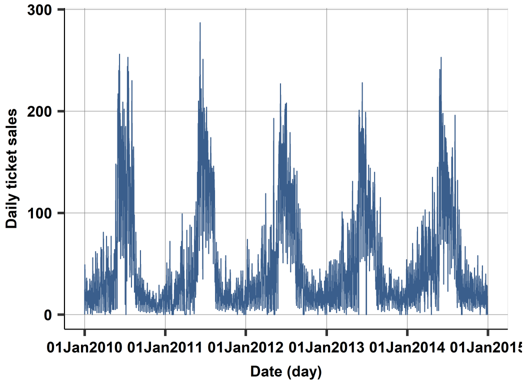
- Trend is simple – use simple linear trend: αt
  - Maybe not really important at all
- Seasonality is important and tricky

Case study: Daily ticket sales

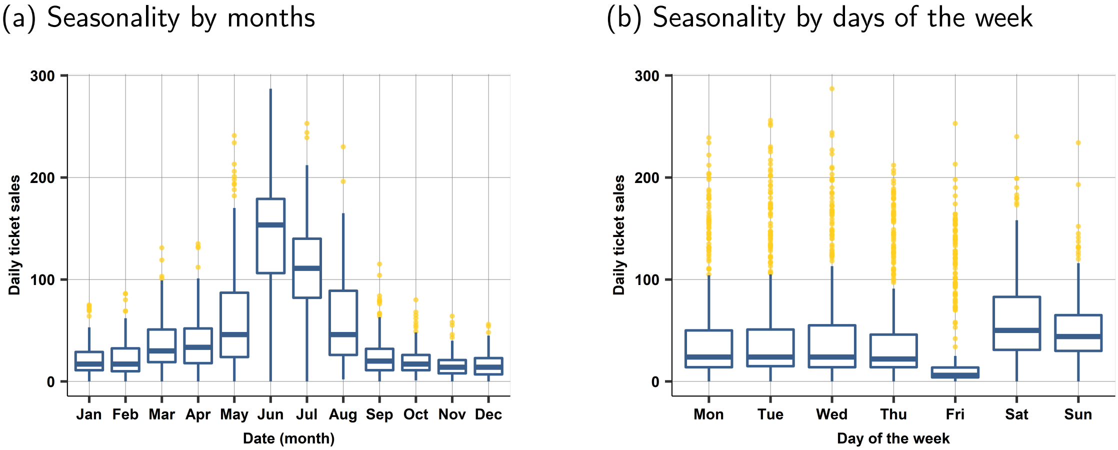
Daily sales - 1 year



Daily sales - 5 years

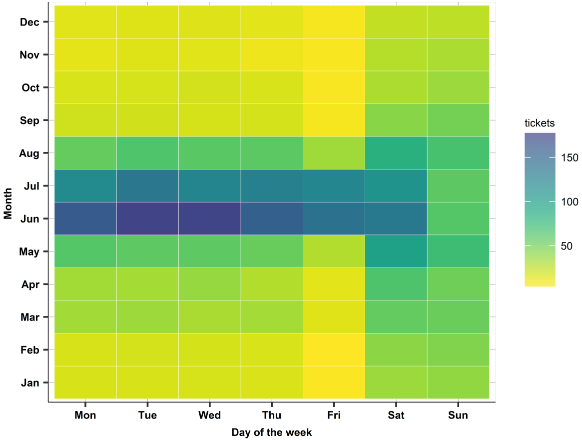


Case study: Monthly and daily seasonality in the number of tickets sold



Case study: Daily ticket sales: A heatmap

- Tool to model seasonality
- Each cell is average sales for a given combination of day and month over years
- Colors help see pattern



Case study: Modeling

- Trend is simple - linear trend
- Seasonality is tricky - need to model and simplify
  - Months
  - Days of the week
  - USA holidays
  - Summer break
  - Interaction of summer break and day of the week
  - Interaction of weekend and month

Case study: Model features and RMSE

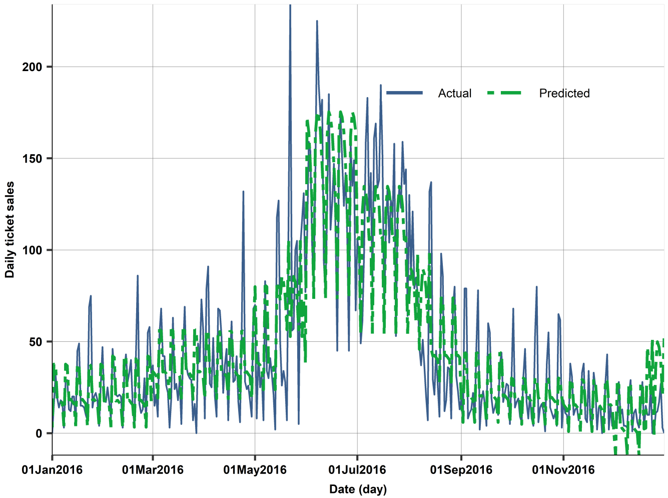
	trend	months	days	holidays	school*days	days*months	RMSE
M1	X	X					32.35
M2	X	X	X				31.45
M3	X	X	X	X			29.46
M4	X	X	X	X	X		27.61
M5	X	X	X	X	X	X	26.90
M6 (log)	X	X	X	X	X		30.99
M7 (Prophet)	X	X	X	X	N/A	N/A	29.47

Note: Trend is linear trend, days is day-of-the-week, holidays: national US holidays, school*days is school holiday (mid-May to mid-August and late December) interacted with days of week. RMSE is cross-validated. Source: swim-transactions dataset. Daily time series, 2010–2016, N=2522 (work set 2010–2015, N=2162).

Case study: Modeling

- We build a set of models.
- The winner has all these ingredients:
  - Months, days of the week, USA holidays
  - Interaction of summer break and day of the week
  - Interaction of weekend and month
- Tried level and log, level is better in terms of CV RMSE
- Took best model, re-estimated on full work and predicted for holdout

Compared actual vs predicted on holdout set (2016)



Diagnostics - holdout set (2016)

Figure: Actual v predicted for August

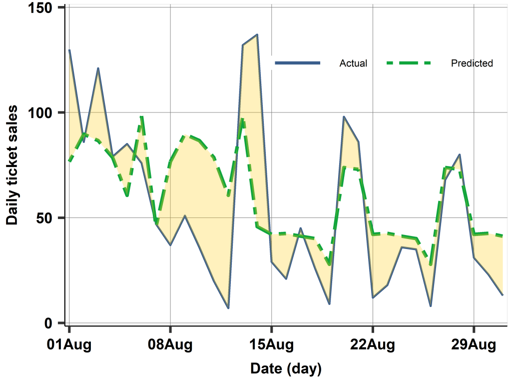
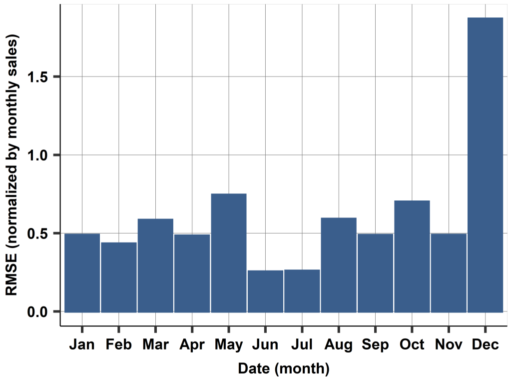


Figure: Monthly RMSE



Diagnostics vs model building

- Here we used diagnostics to learn about what to expect, strength and weaknesses of the *already selected* model.
- This means, no going back to drawing board.
- But, we could have said, lets run some checks on the training set and maybe alter the model accordingly.
  - As we normally do with scatterplots, lowess, tabulations, etc
  - Here: something weird happening in December.
- Act, build a new model and test, etc.

An algorithm to build a models: Prophet

- Another option is an algo built by Facebook folks
- Prophet is a forecasting procedure algorithm - https://facebook.github.io/prophet/
- Trends + seasonality + change in trends + add-on for special events (holidays)
- Find functional form flexibly, try out many different combinations
  - In a smart way

Case study: Using Prophet

- Prophet
  - Trends + seasonality + special events
- In our case study as good as simple model, not as good as best model
  - But fairly close
  - And automatic..

Short run horizon, serial correlation and ARIMA

Short-horizon forecasting: content

We'll also cover 18.U1, 18.U2

Short-horizon forecasting: what is new?

- Serial correlation is essential
- Modeling *how* a shock fades away
- **Autoregressive models– AR models**, capture the patterns of serial correlation – y at time t is regressed on its lags, that is its past values, $t-1$, $t-2$, etc.
- The simplest includes one lag only, AR(1):

$$y_t{}^E = \beta_0 + \beta_1 y_{t-1} \tag{3}$$

- Interested in estimating β_1 or as serial correlation coefficient is called, p.
  - $p = 1$ is random walk, $p = 0$ is white noise
    - May review Ch 12.3 and 12.6 – 12.8

Short-horizon forecasting: AR(1)

- One-period-ahead forecast from an AR(1)

$$\hat{y}_{T+1} = \hat{\beta}_0 + \hat{\beta}_1 y_T \tag{4}$$

- Forecasting to $T+2$ would need y_{T+1} in the formula. – need to use its predicted value, $\hat{y}_{T+1}$:

$$\hat{y}_{T+2} = \hat{\beta}_0 + \hat{\beta}_1 \hat{y}_{T+1} = \hat{\beta}_0 + \hat{\beta}_1(\hat{\beta}_0 + \hat{\beta}_1 y_T) = \hat{\beta}_0(1 + \hat{\beta}_1) + \hat{\beta}_1^2 y_T \tag{5}$$

- As β_1 is less than one (in absolute value), its square is smaller, and higher powers are even smaller – practically zero after a while.

Short-horizon forecasting: ARIMA

- ARIMA(p,d,q) models that are generalizations of the AR(1) model
- can approximate any pattern of serial correlation.
- ARIMA models are put together from three parts: AR(p), I(d) and MA(q).

Short-horizon forecasting: The innovation e_t

- Novelty: idea of innovation = difference in y from the value that could be expected beforehand.

$$e_t = y_t - y_t^E \tag{6}$$

- Content: new in y_t on top of what was expected of it.
  - innovation: y^E is based on previous y observations.
  - Today's y value = what we expected of it based on what we knew yesterday + plus what's new today: the innovation.

$$y_t = y_t^E + e_t$$

  - Note: Innovation is formally like prediction error. But now, past values used for prediction.

Short-horizon forecasting: ARIMA(p,0,0)

- Start: generalize AR(1): AR(p) or ARIMA(p,0,0)
- AR(p), which predicts y_t using up to p of its own lags:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + ... + \beta_p y_{t-p} + e_t \tag{7}$$

- Flexible way to model how shocks fade away
- As we forecast further ahead – use the predicted values in the formula,
  - additional error in the forecasts due to larger estimation error and model error.

Short-horizon forecasting: ARIMA(p,1,0)

- Now, look at first differences: ARIMA(p,1,0)
- AR(p), which predicts y_t using up to p of its own lags:

$$\Delta y_t = \beta_0 + \beta_1 \Delta y_{t-1} + \beta_2 \Delta y_{t-2} + ... + \beta_p \Delta y_{t-p} + e_t \tag{8}$$

- Similar, but interpretation different here
  - β_0 captures the trend,
  - e_t captures how the *change* in y is different from expected based on AR(p) part.

Short-horizon forecasting: ARIMA(0,0,1)

- MA(1) model is a linear combination of two consecutive innovations:

$$y_t = e_t + \theta e_{t-1} \tag{9}$$

- Not a regression model: can't interpret θ as a coefficient.
- θ = moving average coefficient, $|\theta| < 1$.
  - the current value of y is the linear combination of this period's innovation and last period's innovation
  - θ is what multiplies last period's innovation.
  - Estimated via a algo called maximum likelihood (not OLS!)
- It's a linear combination, not a simple average
  - coefficients on e_t and e_{t-1} don't add up to one

Short-horizon forecasting: ARIMA(0,0,q)

- The MA(q) model (an ARIMA(0,0,q)) is a generalization of the MA(1):

$$y_t = e_t + \theta_1 e_{t-1} + ... + \theta_q e_{t-q} \tag{10}$$

- Similarly to an AR model, we can specify an MA model for the changes in y, which would be an ARIMA(0,1,q).

Short-horizon forecasting: MA vs AR

- MA models of serial correlation.
- MA(1) – first order serial correlation = $Corr(y_t, y_{t-1} = \theta/(1+\theta^2)$.
  - Has θ and nothing else, monotonic function.
  - First order serial correlation is stronger, the larger θ is (in absolute value).
- second order serial correlation is zero, and so are all other serial correlations: $Corr(y_t, y_{t-2}) = 0$, $Corr(y_t, y_{t-3}) = 0$, etc.
  - Unlike AR(1) model: serial correlations decrease but don't drop to zero.
- AR and MA models capture different patterns of serial correlation.
  - Combination offers flexibility

Short-horizon forecasting: ARIMA(p,d,q)

- AR(p) and an MA(q) yield an ARIMA(p,0,q) model:

$$y_t = \beta_0 + \beta_1 y_{t-1} + ... + \beta_p y_{t-p} + e_t + \theta_1 e_{t-1} + ... + \theta_q e_{t-q} \tag{11}$$

- The analogous model when y is in changes is an ARIMA(p,1,q):

$$\Delta y_t = \beta_0 + \beta_1 \Delta y_{t-1} + ... + \beta_p \Delta y_{t-p} + e_t + \theta_1 e_{t-1} + ... + \theta_q e_{t-q} \tag{12}$$

Short-horizon forecasting: ARIMA

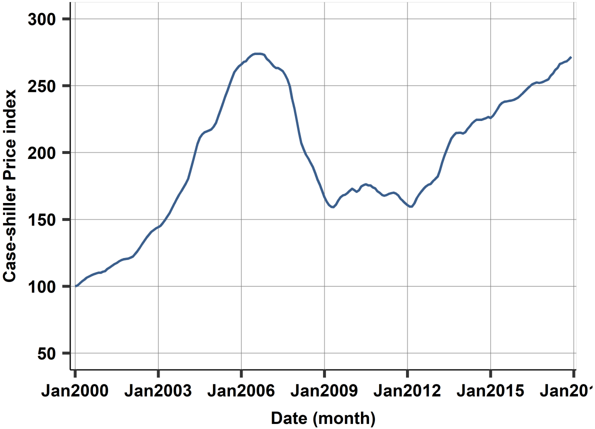
- How to choose (p,d,q)?
- Whichever works best in a cross-validated exercise!
- Try out a few and pick the one that works best
- "auto-arima" - an algo that tries out many options
- keep it simple, $d = 0, 1$ and $p = 0, 1, 2$ and $q = 0, 1, 2$ rarely more

Short-horizon forecasting: ARIMA

- There is a lot more about time series models one can learn.

Case study: Case- Shiller home price index

- Case-Shiller home price index, Los Angeles
- Monthly index of home prices
- Data available: fred.stlouisfed.org
- Use 18 years of monthly data



Case study: Case Shiller home price data

- 18 years of data 2000-2017
- work: 2000-2016, holdout is 2017
- cross-validate with rolling window, 4-fold
  - train is 2000-2012, test is 2013
  - ...
  - train is 2003-2015, test is 2016
- We'll predict 12 months ahead
  - RMSE - symmetric and quadratic loss
  - Assume getting index right matters exactly the same

Case study: target variable

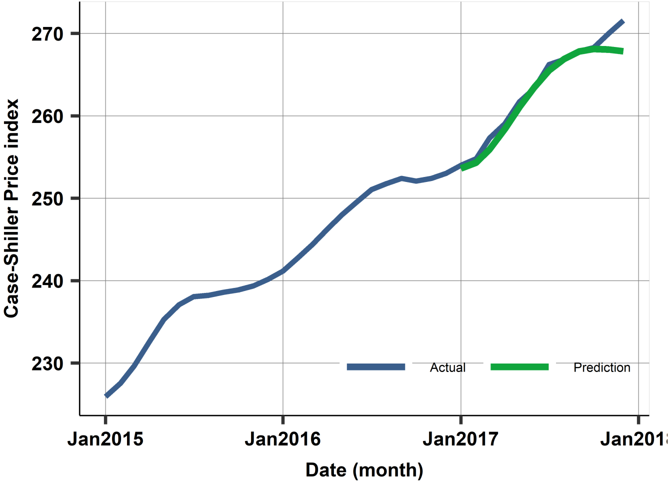
- What should be the target variable?
  - The price index
  - The log of the price index
  - First difference
  - We'll try out, and pick via cross-validation
- The model should include seasonal dummies (could be more complicated)
- The model may include a linear trend or capture it with Δy as target
- The model can have any form of ARIMA

Case study: Case- Shiller home price index - prediction from ARIMA models

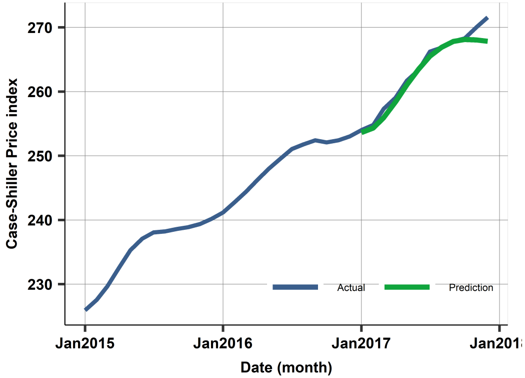
Table: Models and CV RMSE

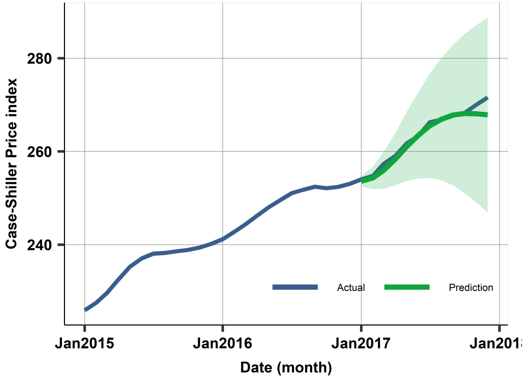
id	target	ARIMA	trend	season	AR	I	MA	RMSE
M1	p	NO	X	X				31.9
M2	p	YES			1	1	2	9.5
M3	p	YES		X	1	1	0	4.1
M4	p	YES	X	X	2	0	0	2.3
M5	dp	NO	X	X				18.8
M6	lnp	YES		X	0	2	0	7.2

Case study: Prediction with best model M4



Case study: Prediction with best model M4: Uncertainty





VAR

- Better forecasts with the help of other variables, at least for short forecast horizons.
- Need forecasts of the x variable as well – we need a model.
- Vector autoregression (VAR), is a method that incorporates other variables in time series regressions and can use those other variables for forecasting y.
  - A set of time series regressions.

VAR: simplest model

- The simplest VAR model has y and one x variable, and it includes one lag of each = VAR(1) model.
- Data have all variables at the same frequency
- VAT(1) = set of two TS regressions:

$$\begin{aligned} {y_t}^E &= \beta_{10} + \beta_{11} y_{t-1} + \beta_{12} x_{t-1} \\ {x_t}^E &= \beta_{20} + \beta_{21} y_{t-1} + \beta_{22} x_{t-1} \end{aligned} \tag{13}$$

VAR forecast

- One-period-ahead forecast for y, only need estimates from the first one:

$$\hat{y}_{T+1} = \hat{\beta}_{10} + \hat{\beta}_{11} y_T + \hat{\beta}_{12} x_T \tag{14}$$

- For forecasting y further ahead, we do need all coefficient estimates.
- Such forecasts use forecast values of x as well as y. A two-period-ahead forecast of y from a VAR(1) is

$$\hat{y}_{T+2} = \hat{\beta}_{10} + \hat{\beta}_{11} \hat{y}_{T+1} + \hat{\beta}_{12} \hat{x}_{T+1} \tag{15}$$

where $\hat{y}_{T+1} = \hat{\beta}_{10} + \hat{\beta}_{11} y_T + \hat{\beta}_{12} x_T$, and $\hat{x}_{T+1} = \hat{\beta}_{20} + \hat{\beta}_{21} y_T + \hat{\beta}_{22} x_T$. Forecasts for $T+3$, $T+4$, etc., are analogous.

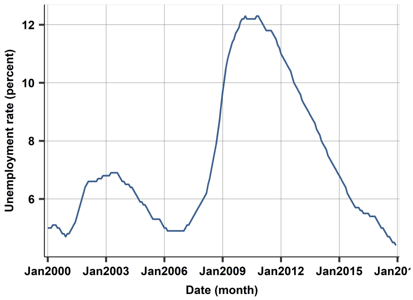
VAR characteristics

There are four important characteristics of a VAR:

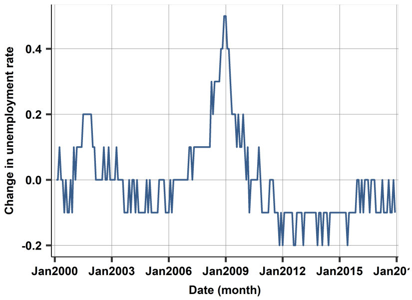
- A VAR has a regression for each of the variables.
- The right-hand side of each equation has all variables.
- Right-hand-side variables are in lags only.
- All right-hand-side variables in all regressions have the same number of lags

Case study: Additional predictors 1

Unemployment rate

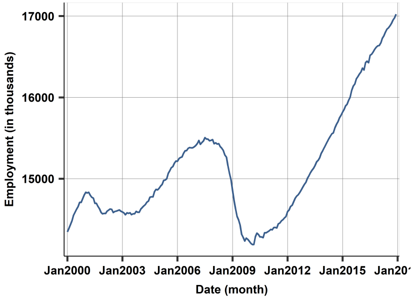


Change in unemployment rate

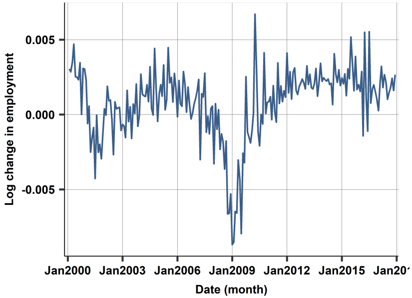


Case study: Additional predictors 2

ln(Employment)



Change in ln(Employment)



Case study: Case- Shiller home price index - Model selection 2

Run the VAR model and compare to previous results.

Table: Models and CV RMSE

id	target	ARIMA	trend	season	AR	I	MA	RMSE
M1	p	NO	X	X				31.9
M2	p	YES			1	1	2	9.5
M3	p	YES		X	1	1	0	4.1
M4	p	YES	X	X	2	0	0	2.3
M5	dp	NO	X	X				18.8
M6	lnp	YES		X	0	2	0	7.2
M7a	dp	**VAR**						7.8
M7b	dp	**VAR**		X				4.5

Case study: Case- Shiller home price index - VAR

- In this case study, VAR did not improve on ARIMA.

External validity in time series

- External validity is about the stability of patterns in the data
- Such as trends, seasonality

- Stationarity is what we look for:
  - Distribution of the target, predictors is stable over time
  - Correlation patterns also stable over time

- External validity is massive risk with time series by design: predict for future

Case study: External validity

- Look across years to see stability
- Four rolling windows, test sets were 2013,14,15,16

Case study: Case- Shiller home price index - model fit on test sets

Four test set (in work set) with rolling window CV. RMSE in each test set for each model.

Table: Models and CV RMSE

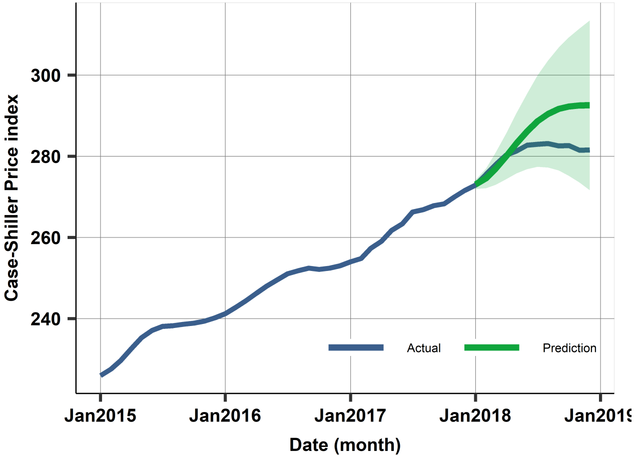
	Fold1	Fold2	Fold3	Fold4	Average
M1	14.90	17.58	34.44	48.58	31.9
M2	14.83	8.39	6.23	5.52	9.5
M3	6.68	1.39	3.29	3.22	4.1
M4	2.22	1.96	2.88	1.20	2.2
M5	33.94	9.79	10.44	7.39	18.8
M6	2.49	4.95	9.22	9.54	7.2
M7a (VAR)	13.30	5.85	3.52	4.28	7.8
M7b (VAR)	5.24	2.51	5.18	4.75	4.5

Case study: External validity 2

- External validity is about the stability of patterns in the data
- Such as trends, seasonality

- First version of this case study a year ago
- Updated more recently, now with 2018 data
- Keep best model of M4. Repeat exercise with training= 2000-2017, holdout=2018, see what happens

Case study: Prediction with best model M4 for 2018



Summary

- Time series prediction is both simple and very hard
- Simple as some basic models work okay
  - Model trend as first difference or linear trend
  - Model seasonality, regular events
  - Some basic method of capturing serial correlation

Summary

- Time series prediction is both simple and very hard
- Simple as some basic models work okay
  - Model trend as first difference or linear trend
  - Model seasonality, regular events
  - Some basic method of capturing serial correlation

- Time series prediction model building is also very hard
  - Getting seasonality, holidays, changing patterns right
  - Getting target variable and ARIMA(p,d,q) selection needs competing models

Summary

- Time series prediction is both simple and very hard
- Simple as some basic models work okay
  - Model trend as first difference or linear trend
  - Model seasonality, regular events
  - Some basic method of capturing serial correlation

- Time series prediction model building is also very hard
  - Getting seasonality, holidays, changing patterns right
  - Getting target variable and ARIMA(p,d,q) selection needs competing models

- Most importantly: external validity is a huge problem
  - Stability may easily break down, and there is nothing we can do.