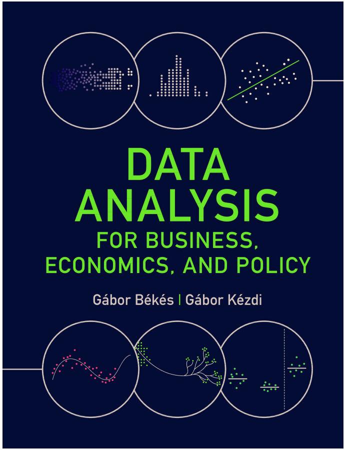
18. Forecasting from times series

Gábor Békés

Data Analysis 3: Prediction

2020

Slideshow for the Békés-Kézdi Data Analysis textbook



- Cambridge University Press, 2021
- **gabors-data-analysis.com**
  - Download all data and code: gabors-data-analysis.com/data-and-code/
- This slideshow is for **Chapter 18**

Forecasting basics

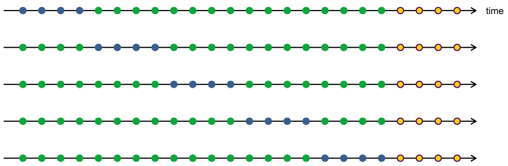
- Forecasting is a special case of prediction.
- Forecasting makes use of time series data on y, and possibly other variables x.
- The original data used for forecasting is a time series from 1 through T, such as $y_1, y_2, ..., y_T$
- The forecast is prepared for time periods after the original data ends, such as $\hat{y}_{T+1}, \hat{y}_{T+2}, ..., \hat{y}_{T+H}$. This is the live time series data.
- Cross-validation with time series is necessary, and it's not trivial

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

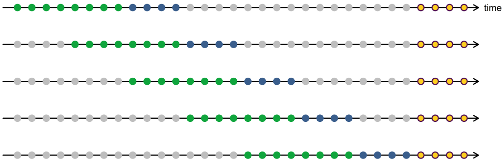
Cross-validation - option 1: test within data

- 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 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

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

ABQ swimming

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

Sample design

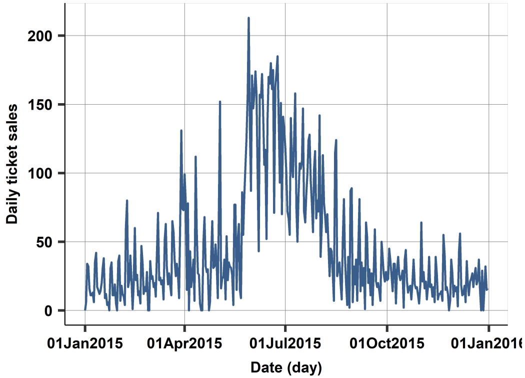
- Sample: Single swimming pool
- Aggregated: number of ticket sales per day
- After some sample design - regular tickets only

Modeling decisions

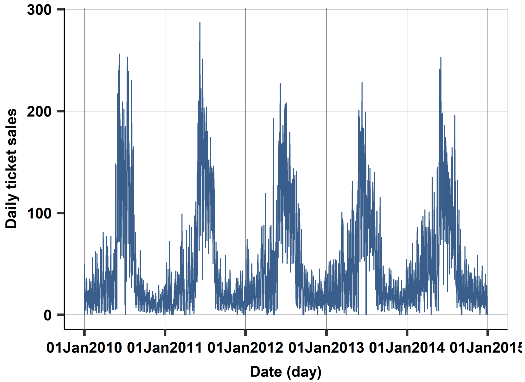
- Trend is simple – use simple linear trend: αt
  - Maybe not really important at all
- Seasonality is important and tricky

Daily ticket sales

Daily sales - 1 year

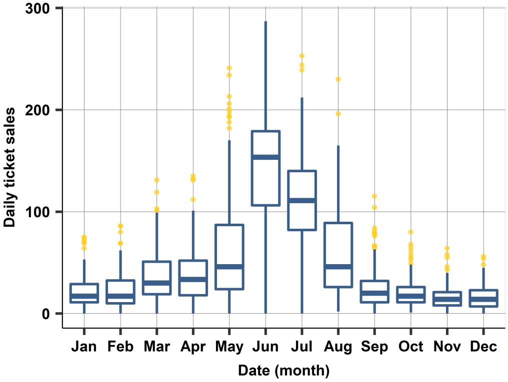


Daily sales - 5 years

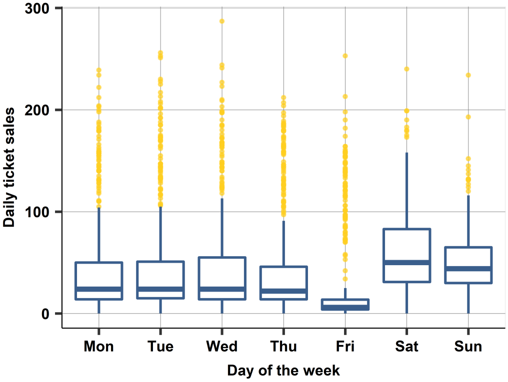


Monthly and daily seasonality in the number of tickets sold

(a) Seasonality by months

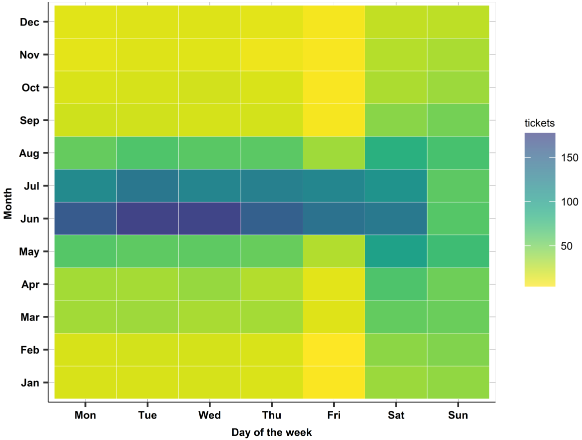


(b) Seasonality by days of the week



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



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

Model features and RMSE

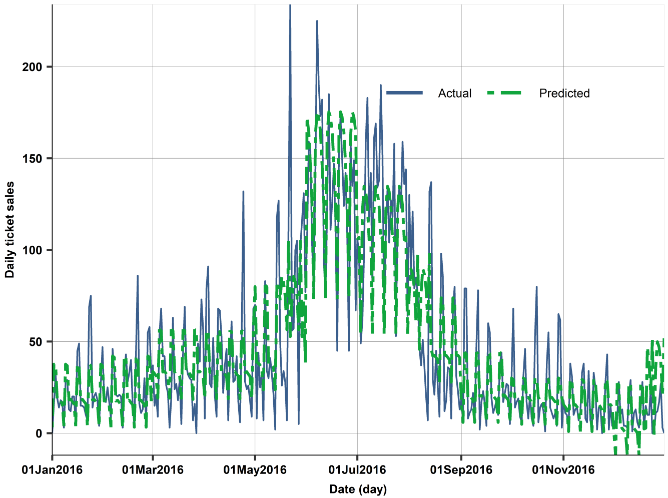
	trend	months	days	holidays	school*days	days*months	RMSE
M1	X	X					32.35
M2	X	X	X				31.45
M3	X	X	X	X			29.46
M4	X	X	X	X	X		27.61
M5	X	X	X	X	X	X	26.90
M6 (log)	X	X	X	X	X		30.99
M7 (Prophet)	X	X	X	X	N/A	N/A	29.47

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).

Modeling steps

- 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

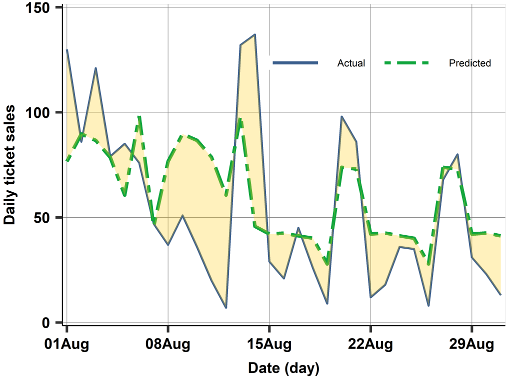
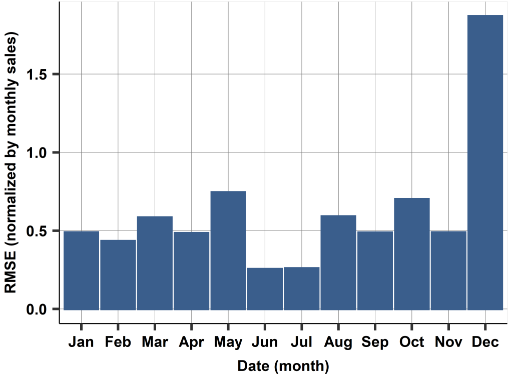


Figure: Monthly RMSE



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-horizon forecasting: what is new?

- Serial correlation
- Model 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

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.
- Can have $\Delta y_t = y_t - y_{t-1}$ as target, too.

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: AR(p)

- AR(p), which predicts y_t using up to p of its own lags:

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 y_{t-1} + \hat{\beta}_2 y_{t-2} + \ldots + \hat{\beta}_p y_{t-p} \tag{6}$$

- 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: ARMA(p,q)

- MA(q) models also capture serial correlation
- Useful when serial correlation drops suddenly to zero
  - AR(p) good when gradual decay

- AR+MA=ARMA – ARMA(p,q) model has $p+q+1$ coefficients to estimate
- ARMA model coefficients cannot be interpreted – too complicated
- ARIMA(p,d,q) – with I(d): whether the ARMA model is written in terms of y itself or its change, Δy.

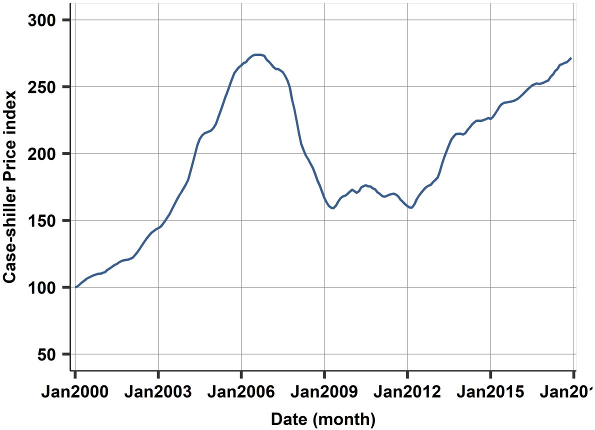
- For more, see 18.U1.

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
- Note: Large econometric literature on time series - lot more about time series models one can learn.

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 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

Target variable

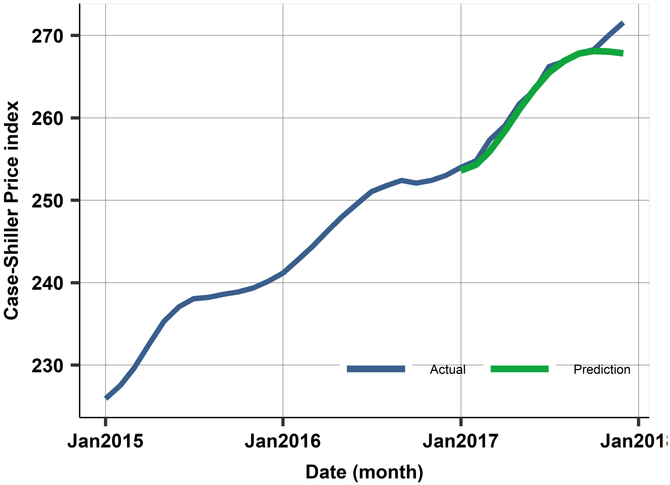
- What should be the target variable?
  - The price index
  - The log of the price index
  - First difference
  - We'll try out, and pick via cross-validation
- The model should include seasonal dummies (could be more complicated)
- The model may include a linear trend or capture it with Δy as target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

Case- Shiller home price index - prediction from ARIMA models

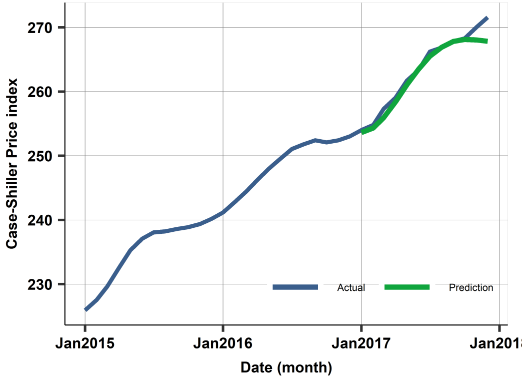
Table: Models and CV RMSE

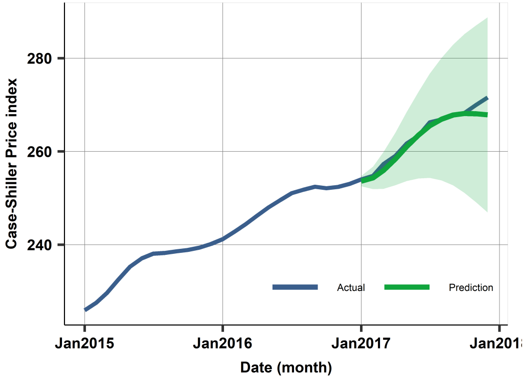
id	target	ARIMA	trend	season	AR	I	MA	RMSE
M1	p	NO	X	X				31.9
M2	p	YES			1	1	2	9.5
M3	p	YES		X	1	1	0	4.1
M4	p	YES	X	X	2	0	0	2.3
M5	dp	NO	X	X				18.8
M6	lnp	YES		X	0	2	0	7.2

Prediction with best model M4



Prediction with best model M4: Uncertainty





VAR: vector autoregressions

- VAR models are a set of time series regressions with more than one variable
- VAR models have a regression for each variable; the lagged values of all variables are entered on the right-hand side of each equation, with the same number of lags everywhere
- A VAR(1) model with two variables y and x is
  - ${y_t}^E = \beta_{10} + \beta_{11} y_{t-1} + \beta_{12} x_{t-1}$
  - ${x_t}^E = \beta_{20} + \beta_{21} y_{t-1} + \beta_{22} x_{t-1}$
- As with any time series regression, VAR models can have their variables in levels (logs) or changes (log changes) and can include trend lines and season dummies
- We can use x variables to help predict y with the help of VAR models. For time periods further ahead, VAR models use predicted values of the x variables when predicting y

VAR forecast

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

$$\hat{y}_{T+1} = \hat{\beta}_{10} + \hat{\beta}_{11} y_T + \hat{\beta}_{12} x_T \tag{7}$$

- 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

$$\hat{y}_{T+2} = \hat{\beta}_{10} + \hat{\beta}_{11} \hat{y}_{T+1} + \hat{\beta}_{12} \hat{x}_{T+1} \tag{8}$$

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.

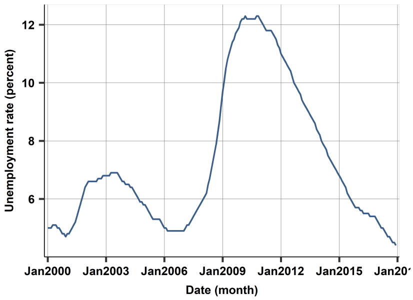
VAR characteristics

There are four important characteristics of a VAR:

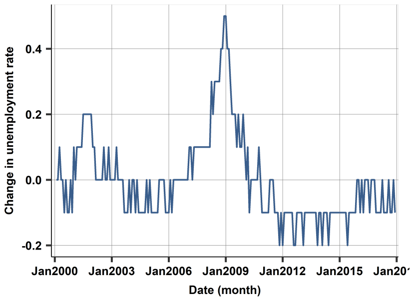
- A VAR has a regression for each of the variables.
- The right-hand side of each equation has all variables.
- Right-hand-side variables are in lags only.
- All right-hand-side variables in all regressions have the same number of lags

Additional predictors 1

Unemployment rate

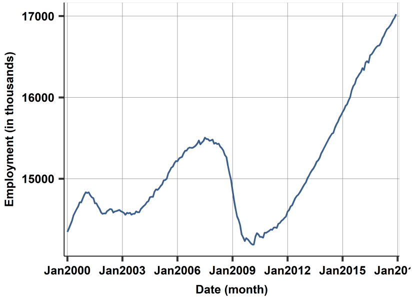


Change in unemployment rate

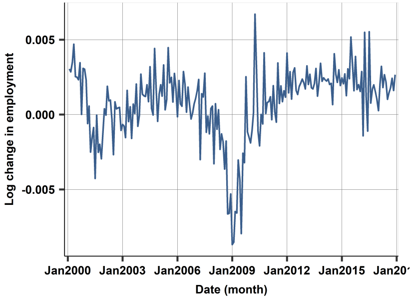


Additional predictors 2

ln(Employment)



Change in ln(Employment)



Case- Shiller home price index - Model selection 2

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

Run the VAR model and compare to previous results. VAR with and without seasonality.

Case- Shiller home price index - VAR

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

External validity

- External validity is about the stability of patterns in the data
- Such as trends, seasonality
- Big threat in time series forecasting
- Look across years to see stability
  - Four rolling windows, test sets were 2013,14,15,16

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

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

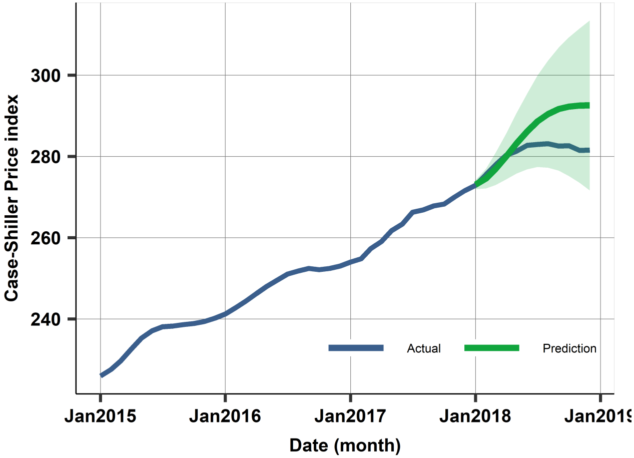
Four test set (in work set) with rolling window CV. RMSE in each test set for each model. VAR with and without seasonality.

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

Prediction with best model M4 for 2018



Main takeaways

- Forecasts use time series data to predict y for one or more time periods ahead
  - For long-horizon forecasts, trend and seasonality are the most important features
  - For short-horizon forecasts, serial correlation can be important, too
  - When using x variables to help forecast y, we need to forecast the values of x, too, and use those in forecasting y