INDR 450/550

Spring 2022

Lecture 4: Simple forecasts, errors and performance, transformations

Feb. 23, 2022

Fikri Karaesmen

Announcements

- The TA for the course is Tufail Ahmad (tahmad20@ku.edu.tr)
- Blackboard page is active
  - Slides of first week's lectures + additional notes
  - Links to books
- I uploaded the data files and python notebooks that I used for analysis during the lectures on blackboard.
- This week, we'll upload a video on basic analysis of some forecasting models.

Correction

- Last lecture, I said in passing that for an exponential smoothing forecast against a stationary process, the asympotic variance of the forecast is minimized for α=1/2.

  - vi.) Exponential smoothing

$$\hat{y}_{T+1|T} = \alpha y_T + (1-\alpha)\hat{y}_{T|T-1}$$

  - Exponential smoothing is unbiased with asymptotic variance (as $T \to \infty$): $(\alpha\sigma^2)/(2-\alpha)$.

- In fact, the asymptotic variance is an increasing function of α. The value that minimizes the variance is α=0.
- What happens when α=0? Why is this not desirable for most real data?

Poll Exercise

- Consider the following forecast for a stationary demand process:

i) stationary i.i.d model

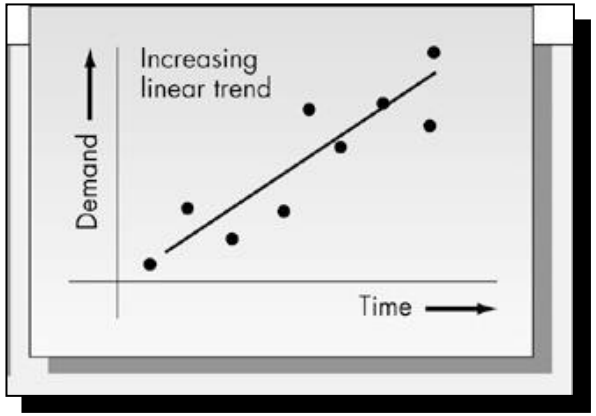
$$Y_t = c + \epsilon_t$$

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha) y_{t-1}$$

This is a weighted average of the two most recent observations.

Answer: $Var(Y_t) = (\alpha^2 + (1-\alpha)^2)\sigma^2$. To minimize the variance we differentiate with respect to α. We can verify that the function is convex wrt to α. The first order condition then yields: α^2=1/2.

Time series: simple forecasts



- Let's now check the simple trend model: $Y_t = bt + c + \epsilon_t$.
- The naive forecast is not unbiased: $\hat{y}_{T+h|T} = y_T$. Taking expectations: $E[\hat{Y}_{T+h}] = b(T+h) + c \neq E[Y_T] = bT + c$.
- Similarly, average, moving average, and exponential smoothing are not unbiased.
- This is expected because to capture the functional form (i.e. slope), we would need to estimate an additional term beyond the 'level' of the series.
- The trend forecast is unbiased:

$$E[\hat{Y}_{T+h|T}] = E\left[Y_T + h\left(\frac{Y_T - Y_1}{T-1}\right)\right] = c + bT + hb = c + (T+h)b.$$

Time series: simple forecasts

- We can of course develop other unbiased estimators. Inspired by the naive method:

$$\hat{y}_{T+h|T} = y_T + h(y_T - y_{T-1})$$

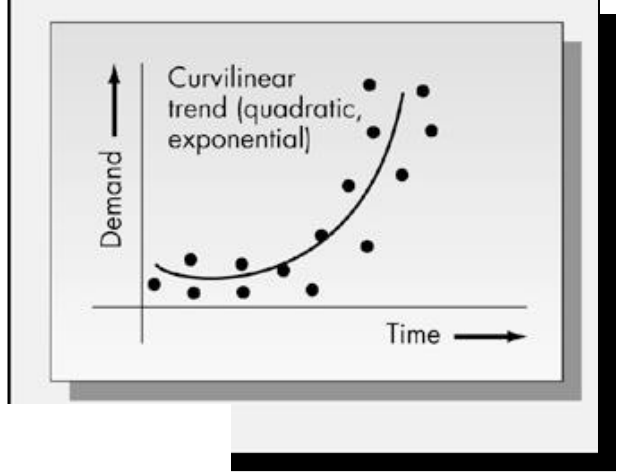
- Inspired by moving averages:

$$\hat{y}_{T+h|T} = y_T + h\left(\frac{(y_T - y_{T-1}) + (y_{T-1} - y_{T-2}) + (y_{T-2} - y_{T-3})}{3}\right)$$

- Inspired by exponential smoothing:

$$\hat{y}_{T+h|T} = y_T + h\left(\alpha(y_T - y_{T-1}) + (1-\alpha)(\hat{y}_{T-1} - \hat{y}_{T-2})\right)$$

Time series: simple forecasts



$$Y_t = at^2 + bt + c + \epsilon_t$$

- None of the above methods will be unbiased. We need to estimate three coefficients.
- We know that we can estimate the expected value of the first difference by $y_T - y_{T-1}$ (similar to the first derivative)
- Similarly the second difference would be estimated by $(y_T - y_{T-1}) - (y_{T-1} - y_{T-2})$.

Time series: simple forecasts

- We know that we can estimate the expected value of the first difference by $y_T - y_{T-1}$ (similar to the first derivative)
- Similarly the second difference would be estimated by $(y_T - y_{T-1}) - (y_{T-1} - y_{T-2})$.
- From the Taylor expansion of functions we have:

$$f(x+h) = f(x) + hf'(x) + h^2 \frac{f''(x)}{2} + ...$$

We can now infer an unbiased forecast:

$$\hat{y}_{T+h} = y_T + +h(y_T - y_{T-1}) + h^2((y_T - y_{T-1}) - (y_{T-1} - y_{T-2}))/2$$

Time series: simple forecasts

- It is therefore easy to obtain a forecast for $Y_t = f(t) + \epsilon_t$ (especially if $f(t)$ is a polynomial).
- We also know that any component of the estimator can be 'smoothed' to decrease the variance by averaging over multiple observations or by exponential smoothing (taking a weighted average of all past observations).
- This is also the basis for more general methods like double exponential smoothing (where we smooth both the estimator of the level and the trend) or triple exponential smoothing (where we also smooth the estimator of the seasonality coefficient).
- We are able to build our own simple forecasting methods that are unbiased and have low variance while keeping an eye on the responsiveness tradeoff.

Time series: double exponential smoothing

- To emphasize some of the general principles, let us look at the idea behind double exponential smoothing, a popular method in practice.
- Double exponential smoothing method applies for a data series with trend. Recall that a reasonable simple and unbiased forecast for such a series would be:

$$\hat{y}_{T+1|T} = y_T + (y_T - y_{T-1})$$

  where y_T is an estimator for the latest position, $(y_T - y_{T-1})$ is an estimator for the trend .
- Such a simple estimator could be unbiased but it has extremely high variance since the two parameters are each estimated by a single value.

Time series: double exponential smoothing

- It then makes sense to smooth each one of the estimators. We can do this by exponential smoothing:
  i) Instead of y_T we can use $\alpha y_T + (1-\alpha)\hat{y}_T$.
  ii) Instead of $(y_T - y_{T-1})$, we can use $\beta(y_T - y_{T-1}) + (1-\beta)\hat{w}_T$. where $\hat{w}_T$ is a an estimator for $y_T - y_{T-1}$.
- This has the advantage of controlling the variance by the choice of the parameters α and β.
- We need to make sure that we choose the best values of the parameters on a training set and check the error performance on a separate test set.

The same idea extends easily to seasonality estimation (Triple Exponential Smoothing
If we have more parameters then quadruple exponential smoothing

Time series: simple forecasts

- We can fit simple but effective models to time-series data if we know the model that generates the stochastic process Y_t. In practice, no model would be given to us in the form: $Y_t = f(t) + \epsilon_t$. We should then validate the model and asses the error performance from the data.

- The simple models that we looked at are basic but effective. Moving averages and exponential smoothing also take into account the responsiveness by focusing on the most recent observations therefore somehow addressing the changes in the demand process. Therefore, even when they are biased, they are somehow able to follow the trends with a time lag. This explains the popularity of double or triple exponential smoothing methods in practice.

Time series: simple forecasts: error performance

- Let us now look at the errors: the discrepancy between the forecast and the realized demand.
- $e_t = y_t - \hat{y}_t$ is called the model residual. We would like the residuals to have mean zero (unbiasedness), to be small in absolute value (low variance). In addition, it would be great to have them uncorrelated (iid).
- The forecast has room for improvement if the above properties are not true.
- Finally, it's even better if the residuals have constant variance over time and they are normally distributed.

Time series: simple forecasts: error performance

- To assess the error performance of models, even for simple forecasting methods, we need to use the principle of fitting the model on a (training set) and test its performance on a separate portion of the data called (test set).
- This becomes crucial for more complicated models but for simple forecasts (naive, moving average etc.) there is no need to fit a parametrized model. This makes the separation very easy. We use the data from first T periods to forecast for period $T + h$.

Source: Hydman and Athanasopoulos

Time series: simple forecasts: error performance

- To assess the error performance of models, even for simple forecasting methods, we need to use the principle of fitting the model on a (training set) and test its performance on a separate portion of the data called (test set).
- This becomes crucial for more complicated models but for simple forecasts (naive, moving average etc.) there is no need to fit a parametrized model. This makes the separation very easy. We use the data from first T periods to forecast for period $T + h$.

Source: Hydman and Athanasopoulos

Time series: simple forecasts: error performance

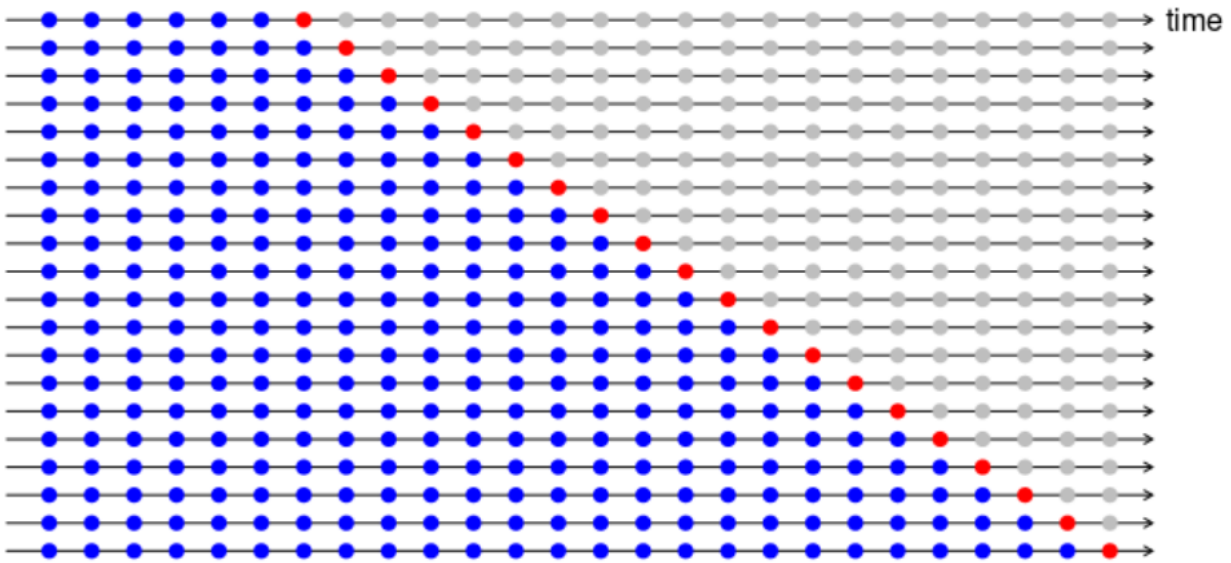
- We use the term error to denote the difference between an observed value and its forecast:

$$e_{T+h} = y_{T+h} - \hat{y}_{T+h|T}$$

  (where the training set is $\{y_1, y_2, ..y_T\}$ and the test set is $\{y_{T+1}, y_{T+2}, ..\}$
- Note that the term residual refers to the training set and the term error refers to the test set.

Time series: cross-validation

Most of the simple methods do not require complicated parametrization. A natural cross-validation is then to compute an estimator for period t using data available up to period t-1, and then compute one for t+1, Using data available upto period t etc.



Source: Hydman and Athanasopoulos

Time series: simple forecasts: error performance

- Absolute error measures:

$$1.\ \text{MAE} = \sum_{t=1}^{T} \frac{|e_t|}{T}$$

$$2.\ \text{MSE} = \sum_{t=1}^{T} \frac{|e_t^2|}{T}$$

$$3.\ \text{RMSE} = \sqrt{\text{MSE}}$$

- Percentage error measure:

$$4.\ \text{MAPE} = \sum_{t=1}^{T} \frac{|p_t|}{T}$$

where $p_t = 100 e_t / y_t$

Time series: simple forecasts: error performance

- Scaled error measure: we express the error in comparison to simple but reasonable benchmark, for instance the naive forecast $\hat{y}_T = y_{T-1}$. Let $e'(t)$ denote the error of benchmark forecast in period T and MAE(benchmark) its mean absolute error. We define

$$q_t = \frac{|e_t|}{\text{MAE(benchmark)}}$$

- and the Mean Absolute Scaled Error is then:

$$5.\ \text{MASE} = \sum_{t=1}^{T} \frac{q_t}{T}$$

Scaled Errors

- We'll look at more sophisticated methods with more parameters. But it's not always the case that having more parameters lead to a better forecast.
- We should always compare performance with respect to simple benchmarks (i.e. Moving averages, double exponential smoothing etc).
- If the benchmarks perform better than the sophisticated forecast, we should use the benchmark.

Time series: simple forecasts: error distribution

- For planning purposes (inventory control, capacity setting, staffing etc), we have to use the probability distribution of the forecast and not only its point estimate (for the mean).
- Let $Y(t) = f(t) + \epsilon_t$. If the distribution of ϵ_t is known then we know the distribution of the forecast $\hat{Y}_t$.
- If all is well (i.e. our forecast is unbiased) and we are lucky, ϵ_t is independent and normally distributed with variance σ_2. If so our forecast is also normally distributed with mean $\mu = f(t)$ and variance σ^2.
- We are then able to write a nice prediction interval around our point forecast $\hat{y}_t$:

$$\hat{y}_t \pm 1.96\sigma$$

This is a 95% prediction interval: $1.96 = \Phi^{-1}(0.975)$.

Time series: simple forecasts: error distribution

- In practice, we don't know the true model that generates the demand but we have observed the residuals. We can use them to estimate the forecast variance $\hat{\sigma}$.
- Note that

$$e_t = y_t - \hat{y}_t$$

Therefore,

$$Var(E_t) = Var(Y_t - \hat{Y}_t) = Var(Y_t) + Var(\hat{Y}_t) > Var(Y_t)$$

- But we make efforts to make the $Var(\hat{Y}_t)$ relatively small. For instance, if we take as a test model $Y_t = c + \epsilon_t$ and as an unbiased forecast a 5-period moving average, we saw that $Var(\hat{Y}_t) = \sigma_2/5$. Then $Var(E_t) = \sigma^2 + \sigma_2/5$. So, $\sigma_E \approx 1.1\sigma$.

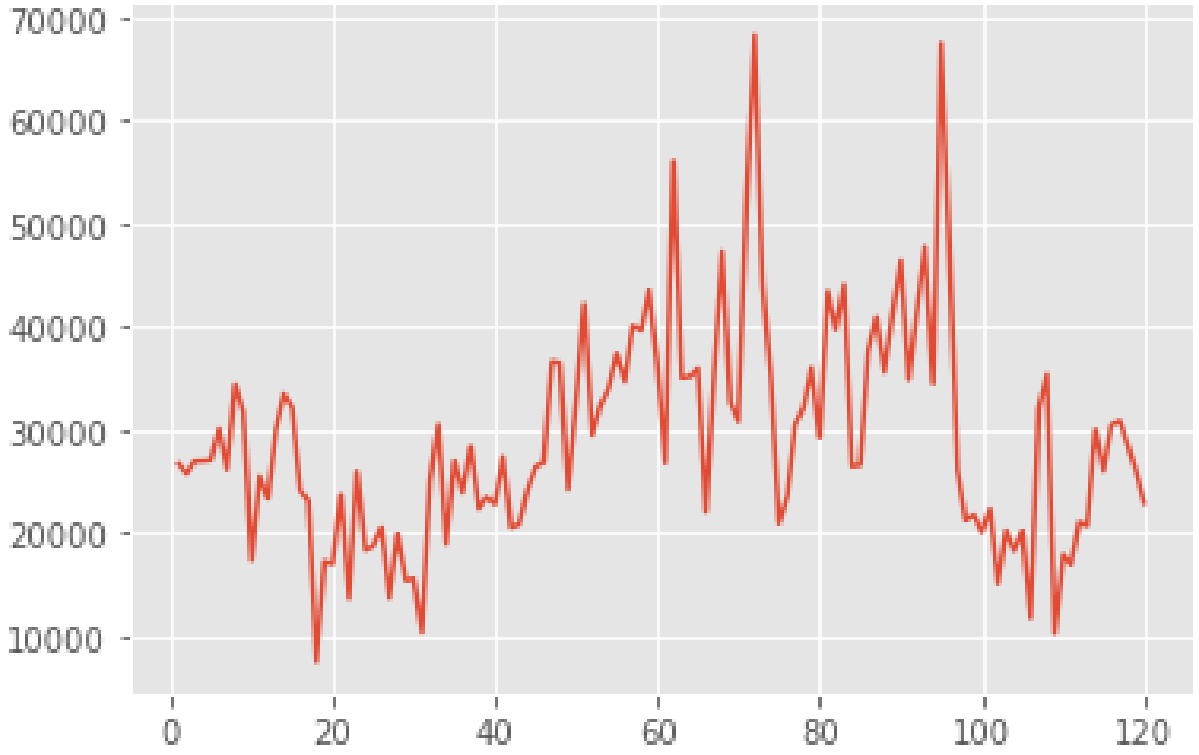
Time series: simple forecasts: error distribution

- We therefore proceed with the distribution of the residuals to approximate distribution of the forecast itself (i.e. the distribution of the forecast is a point estimator plus a random variable represented by the residuals).
- If the residuals are normally distributed, we have approximate prediction intervals given by:

$$\hat{y}_t \pm 1.96\sigma_e$$

- If the residuals are not normally distributed, we need to more work (use MLE to fit a distribution or use the sample as is).
- Residuals are critical for planning purposes. Please note that if we need prediction intervals or a probability distribution for a multi-step ahead forecast, we need to consider multi-step ahead residuals. Typically, the variance of the residuals increase as the horizon increases.

Example: monthly dishwasher sales in Turkey (1998-2008)



Example: one-month ahead forecasts

In [41]:

	Method	MAE	MAPE	RMSE
0	Naive	8222.813924	0.270693	11358.847893
1	MA	7235.245994	0.262230	10137.141796
2	Trend	11252.552095	0.407301	15770.146811
3	ES	7061.946191	0.264597	9433.026540

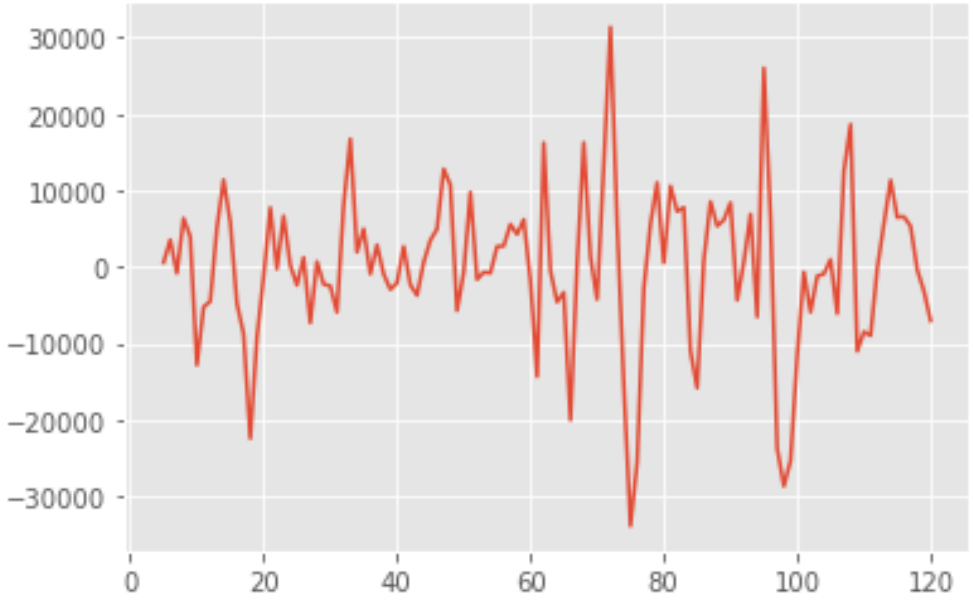
The MA-forecast for month 119 is 29961. Note that the prediction interval is: (29961-1.96*10000, 29961+1.96*10000).
The realized sales for month 119 is: 25997

Example: one-month ahead forecasts (MA)

Sales (orange) vs. forecasts (blue)

The forecast curve is much smoother than the actual sales.

Errors



Note that the average error is close to zero. That's a good sign. But there are huge Fluctuations.