INDR 422/522

Fikri Karaesmen

Spring 2023

Simple time series forecasts - 2

March 9, 2023

Reminders

- Course TA’s: Bijan Bibak (bibak20), Mert Gürel (fegurel)
- Blackboard page is becoming active
  - Last Week’s slides
  - Last year’s lecture slides
  - Class Exercise Solutions
  - Will be uploading the current slides as we proceed
- We’ll upload the first lab soon
- Participation taken. Please participate in polls.
- Please follow announcements

Class Exercise from last lecture

CLASS EXERCISE, March 7, 2023

1. Consider two discrete random variables X and Y with the following joint probability mass function: $P(X = 0, Y = 0) = 1/4$, $P(X = 0, Y = 1) = 1/4$, $P(X = 1, Y = 0) = 0$, $P(X = 1, Y = 1) = 1/2$. Find:

   (a) $Var(X)$

   Solution:

   We can use the short cut formula:

   $$Var(X) = E[X^2] - E[X]^2 = 1/2 - (1/2)^2 = 1/4$$

   Note that $Var(X) = E[(X - E[X])^2]$ and is therefore always non-negative.

   (b) $Cov(X, Y)$

   Solution:

   We can use the short cut formula:

   $$Cov(X, Y) = E[XY] - E[X]E[Y] = (1/2) - (1/2)(3/4) = 1/8.$$

   Recall that $Cov(X, Y) = E[(X - E[X])(Y - E[Y])]$. $Cov(X, Y) > 0$ implies that when X takes large (small) values with respect to its mean Y is also likely to take large (small) values with respect to its mean.

   (c) $Corr(X, Y)$

   Solution:

   $$Corr(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

   Since σ_X and σ_Y cannot be negative, $Corr(X, Y)$ has the same sign as $Cov(X, Y)$ but is normalized to the interval $[-1, 1]$.

Class Exercise from last lecture

2. Let $Y_t = \mu + \epsilon_t$ where ϵ_t are i.i.d random variables with mean zero and variance σ^2. Consider the forecast: $\hat{y}_{t+1} = (y_t + y_{t-1})/2$.

   (a) Is this forecast unbiased?

   Solution: Let us check:

   $$E[\hat{Y}_{t+1}] = E[(Y_t + Y_{t-1})/2] = (2\mu)/2 = \mu$$

   We therefore have: $E[\hat{Y}_{t+1}] = E[Y_{t+1}] = \mu$ so the forecast is unbiased.

   (b) Variance of the above forecast?

   Solution: Let us check:

   $$Var[\hat{Y}_{t+1}] = Var[(Y_t + Y_{t-1})/2] = (2\sigma^2)/4 = \sigma^2/2.$$

   (c) Now consider the forecast: $\hat{y}_{t+1} = (3/4)y_t + (1/4)y_{t-1}$. Is this unbiased.

   Solution: We can see that : $E[\hat{Y}_{t+1}] = E[Y_{t+1}] = \mu$ so the forecast is unbiased.

   (d) Variance of the above forecast?

   $$Var[\hat{Y}_{t+1}] = Var[(3/4)Y_t + (1/4)Y_{t-1}] = (9/16)\sigma^2 + (1/16)\sigma^2 = (5/8)\sigma^2.$$

Class Exercise from last lecture (cont.)

- 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.

<u>Comment:</u> $Var(\hat{Y}_t) = (\alpha^2 + (1-\alpha)^2)\sigma^2$. 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

- We have data corresponding to a time series $y_1, y_2, ..y_T$. For our purposes, we can assume that y_t corresponds to demand in period t.The goal is to forecast the demand in period $T+h$ $h = 1, 2, ..$ given the observations.
- Here are some simple ideas: i) average

$$\hat{y}_{T+h|T} = \frac{\sum_{t=1}^{T} y_t}{T}$$

- ii) naive method

$$\hat{y}_{T+h|T} = y_T$$

- iii) seasonal naive method (where m is the seasonal period)

$$\hat{y}_{T+h|T} = y_{T+h-m} \text{ if } T+h-m \leq T.$$

Time series: simple forecasts

- iv.) Drift (trend) estimation

$$\hat{y}_{T+h|T} = y_T + h\left(\frac{y_T - y_1}{T-1}\right)$$

- v) moving average over k periods

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

Time series: simple forecasts

- vi.) Exponential smoothing

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

where $0 \leq \alpha \leq 1$. Note that since $\hat{y}_{T|T-1} = \alpha y_{T-1} + (1-\alpha)\hat{y}_{T-1|T-2}$ we can recursively write:

$$\begin{aligned}
\hat{y}_{T+1|T} &= \alpha y_T + \alpha(1-\alpha)y_{T-1} + (1-\alpha)^2\hat{y}_{T-1|T-2} \\
&= \sum_{t=1}^{T} \alpha(1-\alpha)^{T-t} y_t
\end{aligned}$$

Time series: simple forecasts

- To get some insight, let us consider some models that will generate data. Assume that ϵ_t are iid random variables with mean zero and standard deviation σ.

  i) stationary i.i.d model

  $$Y_t = c + \epsilon_t$$

  ii) stationary seasonal model

  $$Y_t = c_{t(mod\ m)} + \epsilon_t$$

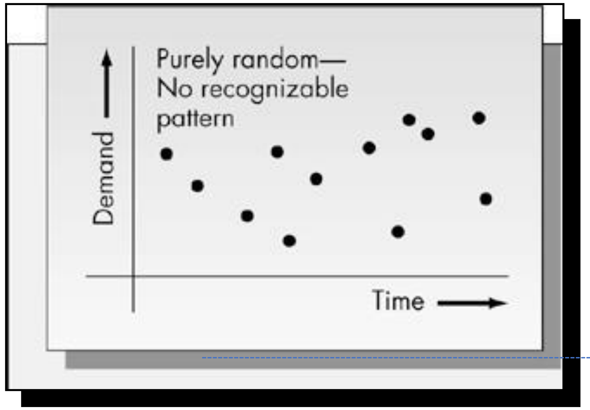
  iii) a model with linear trend

  $$Y_t = bt + c + \epsilon_t$$

  iv) a model with quadratic trend

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

Time series: simple forecasts



- We can now test the properties of the simple estimators:
  i) stationary i.i.d model
  $$Y_t = c + \epsilon_t$$
- The average method and the naive method are both unbiased estimators: $E[\hat{Y}_T] = E[Y_T] = c$.
- The variance of the estimator $Var[\hat{Y}_T]$ is σ^2 for the naive method and σ^2/T for the average method.
- The drift method is also unbiased. The estimator of the drift term is zero in expectation.

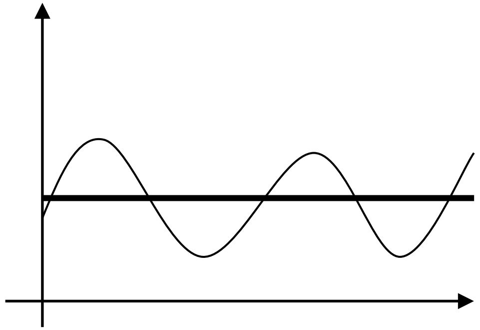
Time series: simple forecasts

- The k-period moving average is unbiased with variance σ^2/k.
- Exponential smoothing is unbiased with asymptotic variance (as $T \to \infty$): $(\alpha\sigma^2)/(2-\alpha)$.
- Note that there are many other unbiased forecasts for a simple stationary series, for instance

$$\hat{y}_{T+h} = y_{T_1},$$
$$\hat{y}_{T+h} = y_T + (y_{T-1} - y_{T-2}),$$
$$\hat{y}_{T+h} = \beta y_T + (1-\beta)y_{T-2} \ (0 \leq \beta \leq 1). \text{ etc.}$$

- These simple models (MA and ES) are basic but effective and frequently used in practice thanks to their responsiveness. Note that MA puts equal weight on the k most recent observations whereas ES puts geometrically decreasing weight on all past observations.

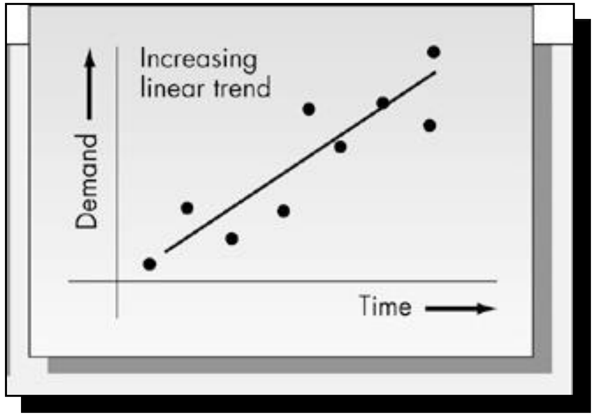
Time series: simple forecasts



- None of the simple forecasts are unbiased for a seasonal series except for the naive seasonal forecast.
- By using the same principles we can build other simple unbiased forecasts.
- For instance, inspired by moving averages we have:
- $$\hat{y}_{T+h|T} = \frac{y_{T+h-m} + y_{T+h-2m} + y_{T+h-3m}}{3} \text{ if } T + h - m \leq T.$$
- And inspired by exponential smoothing we have:

$$\hat{y}_{T+h|T} = \alpha y_{T+h-m} + (1 - \alpha)\hat{y}_{T+h-2m} \text{ if } T + h - m \leq T.$$

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:

$$\hat{y}_{T+h|T} = y_T + (y_T - y_{T_1})$$

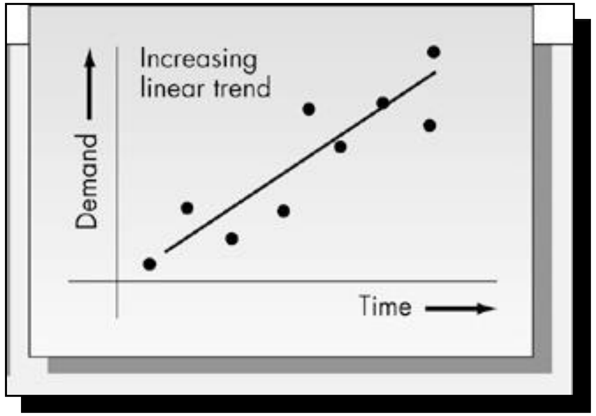
- Inspired by moving averages:

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

- Inspired by exponential smoothing:

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

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:

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

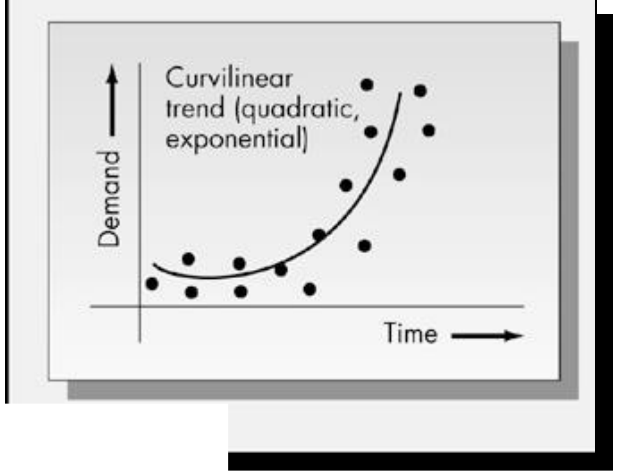
- Inspired by moving averages:

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

- 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:

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

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

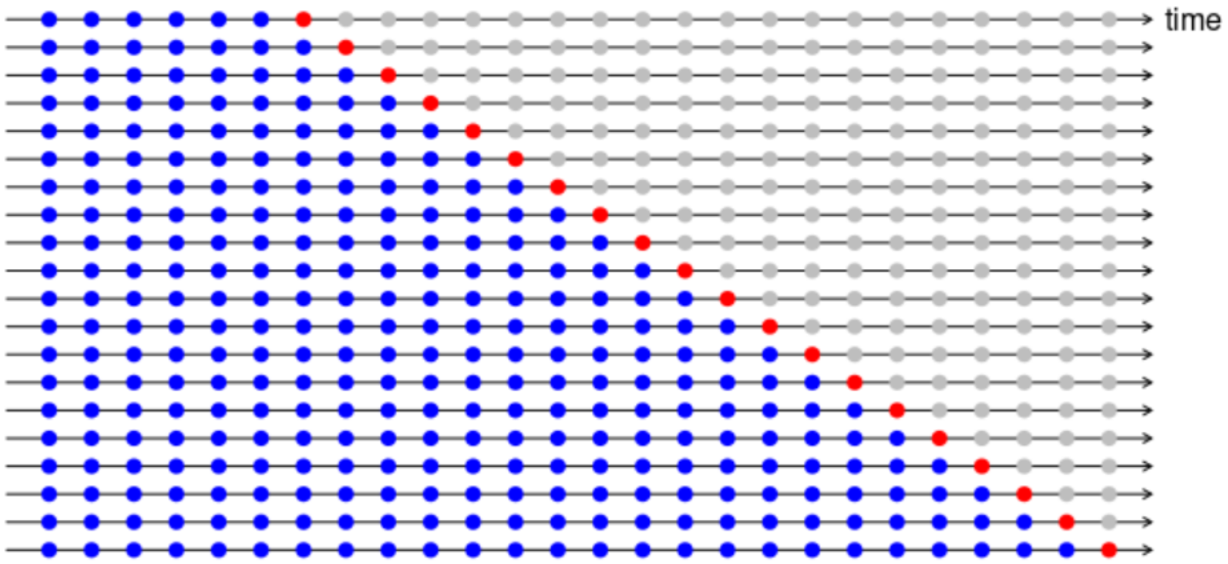
- We use the term error to denote the difference between an observed value and its forecast:

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

  (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.