INDR 422/522

Fikri Karaesmen

Spring 2023

Simple time series forecasts

March 7, 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
  - Will be uploading the current slides as we proceed
- Please follow announcements

- Participation taken starting today

Class Exercise from last lecture

Spring 2023 Fikri Karaesmen

INDR 422/522 CLASS EXERCISE, March, 2, 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) $E[X]$
   Solution:

$$\begin{aligned} E[X] &= \sum_x \sum_y x p(x,y) \\ &= 0(p(0,0) + p(0,1)) + 1(p(1,0) + p(1,1)) \\ &= 1/2 \end{aligned}$$

   (b) $E[Y]$

   (c) $E[XY]$
   Solution:

$$\begin{aligned} E[XY] &= \sum_x \sum_y xy p(x,y) \\ &= 1p(1,1) \\ &= 1/2 \end{aligned}$$

   (d) $P(X = 0|Y = 1)$
   Solution:

$$\begin{aligned} P(X = 0|Y = 1) &= \frac{P(X = 0, Y = 1)}{P(Y = 1)} \\ &= \frac{p(0,1)}{p(0,1) + p(1,1)} \\ &= \frac{1/4}{1/4 + 1/2} \\ &= 1/3 \end{aligned}$$

   (e) $E[X|Y = 1]$
   Solution: We found that $P(X = 0|Y = 1) = 1/3$, therefore $P(X = 1|Y = 1) = 2/3$. Then,

$$E[X|Y = 1] = 0(1/3) + 1(2/3) = 2/3.$$

Fitting a probability distribution

- Let us assume that we have an i.i.d sample of observations for **Y** (after some data transformations).
  - Obtaining and i.i.d. sample requires cleaning up many things in practice through data transformations.

- Eventually, we have something that may look like: y_1 = 24, y_2=35, y_3=11, y_4=48,…, y_n=55.
- Or : y_1 = 24.2, y_2=35.4, y_3=11.9, y_4=48.1,…, y_n=55.3.
- We may plot the histogram of the data and explore its shape (monotone, unimodal, multimodal, symmetrical, skewed).
- And take a guess for continuous or a discrete distribution to fit.

Fitting a probability distribution

- Let's assume we have a sample of iid demand observations $d_1, d_2, \ldots d_n$.
- We think that this sample might correspond to a Poisson r.v. with parameter λ:

$$p_D(x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad x = 0, 1, 2, \ldots$$

- Since λ is not known, We look for the value of λ that makes the sample as likely as possible. This is an optimization problem:

$$\max_{\lambda} \Pi_{i=1}^{n} p_D(d_i, \lambda) = \Pi_{i=1}^{n} \frac{\lambda^{d_i} e^{-\lambda}}{d_i!}$$

This approach to find the optimal fit of the parameter through likelihood maximization is called Maximum Likelihood Estimation (MLE).

Fitting a probability distribution (MLE)

- The solution of the above problem:

$$\lambda^* = \arg\max_{\lambda} \Pi_{i=1}^{n} p_D(d_i, \lambda)$$

corresponds to the value that maximizes the likelihood of the sample with respect to a given distribution.

And is called the Maximum Likelihood Estimation (MLE) estimator.

- To solve the optimization problem, we take the logarithm of the likelihood function to convert the product to a sum.

Ex: Poisson (λ), sample $x_1, x_2 \dots x_n$

The likelihood function:

$$L(x_1, x_2 \dots x_n; \lambda) = \frac{\lambda^{x_1} e^{-\lambda}}{x_1!} \cdot \frac{\lambda^{x_2} e^{-\lambda}}{x_2!} \dots \frac{\lambda^{x_n} e^{-\lambda}}{x_n!}$$

We take logs to convert the product to a sum

$$\ell(x_1, x_2 \dots x_n; \lambda) = \log L(x_1, x_2 \dots x_n; \lambda) \quad x_1 \log\lambda - \lambda - \log(x_1!) + x_2 \log\lambda - \lambda - \log(x_2!)$$
$$+ \dots + x_n \log\lambda - \lambda - \log(x_n!)$$

$$\frac{d\ell}{d\lambda} = \frac{\sum x_i}{\lambda} - n \Rightarrow \lambda^* = \frac{\sum x_i}{n}$$

Reminder: estimators and properties

- Let us note that sample based estimators are themselves random variables. Each time we draw a new random sample, we'll get a different value for our estimator.
- Unbiasedness: A desirable property for an estimator is that it does not have a systematic error on the average (in expectation). The sample mean $\bar{X}$ is an unbiased estimator of the population mean since:

$$E[\bar{X}] = \mu.$$

- Note that there are many unbiased estimators: X_1 and $(2X_1 + X_2)/3$ are also unbiased. Since:

$$E[X_1] = E[(2X_1 + X_2)/3] = \mu.$$

Reminder: estimators and properties

- Variance of the Estimator: Among unbiased estimators, it makes sense to prefer one with a lower variance.
- Assuming that our sample has variance σ^2:

$$Var[\bar{X}] = \frac{\sigma^2}{n}$$

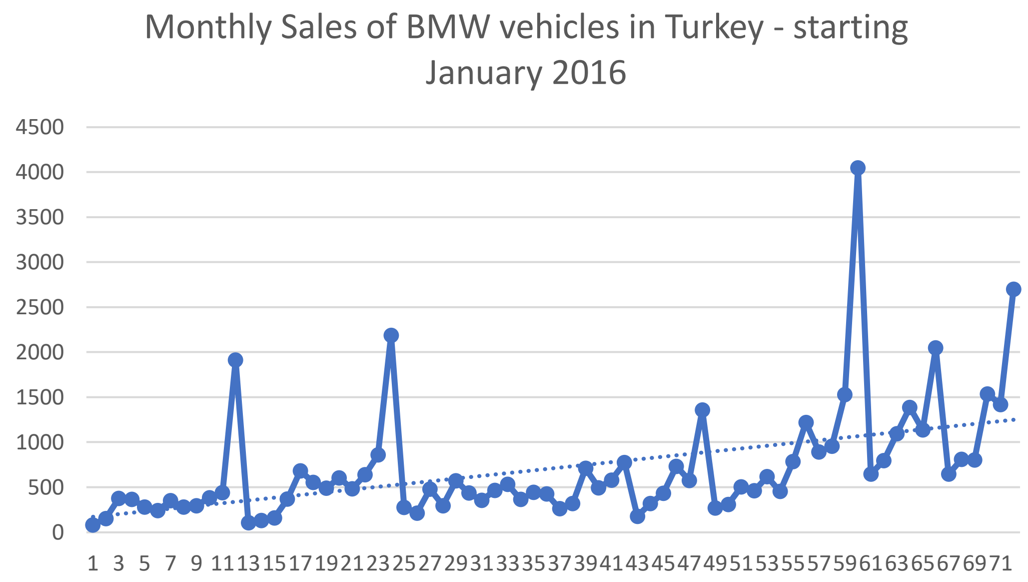
- whereas for the other estimators:

$$Var[X_1] = \sigma^2 \text{ and } Var[(2X_1 + X_2)/3] = \frac{5\sigma^2}{9}.$$

- We will see that for demand forecasting there is a trade-off between responsiveness and low variance.

Time series

- Demand data is typically in the form of a time series: an observation per period
- Note that we typically have sales data which may be different than the actual demand
- We'll forecast future values time series to eventually solve planning problems



Some more real data – Makridakis, Wheelwright, Hyndman (1997)

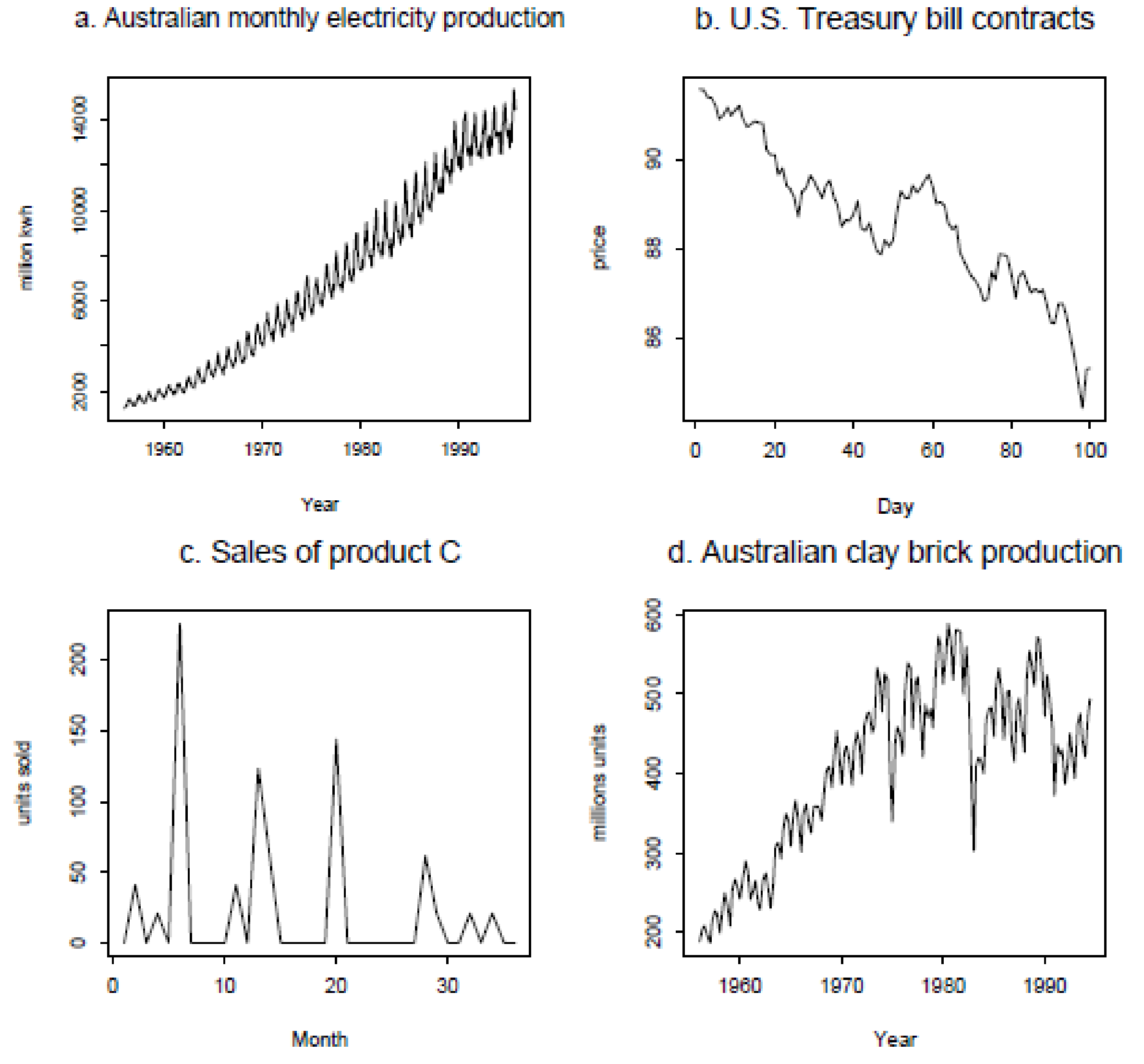


Figure 1-2: *Historical data on four variables for which forecasts might be required.*

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

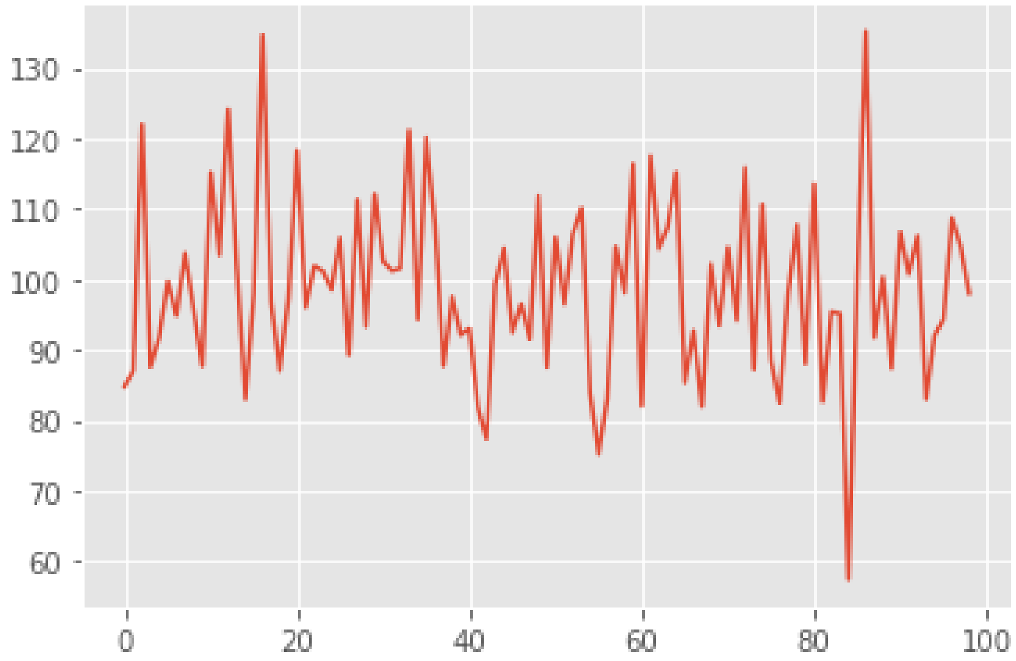
  $$Y_t = at^2 + bt + c + \epsilon_t$$

Time series: simple forecasts

i) stationary i.i.d model

$$Y_t = c + \epsilon_t$$

```
c=100; sigma=15;
y=[0]*101;
for i in range(1,100):
  y[i] =c + sigma*random.normalvariate(0, 1)
```

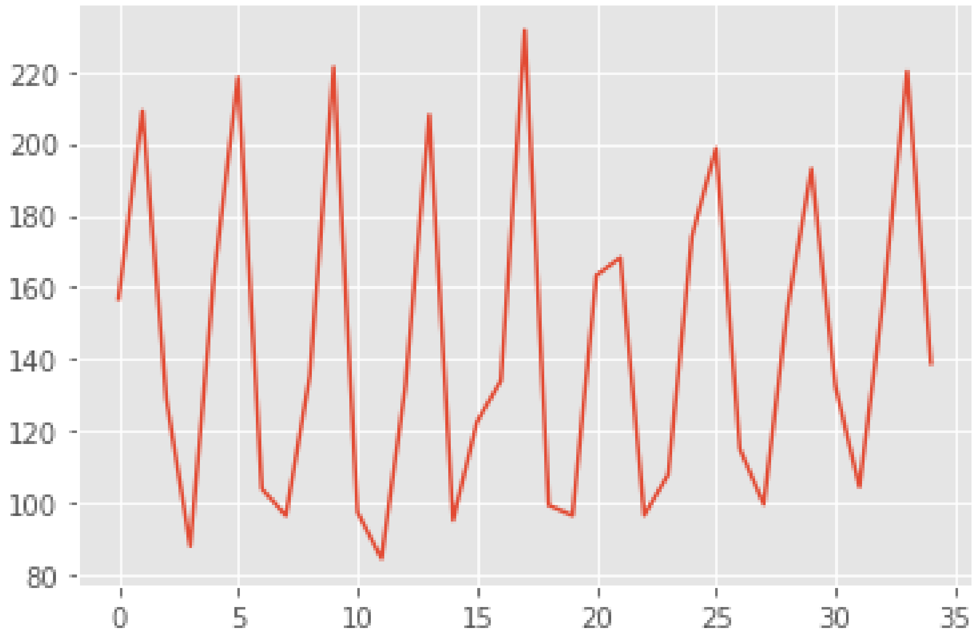


Time series: simple forecasts

ii) stationary seasonal model

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

```
ssc=[0]∗4;
ssc[0]=100; ssc[1]=150; ssc[2]=200; ssc[3]=120; sigma=15; b=2;
yss=[0]∗101;
for i in range(1,100):
 yss[i] =ssc[i % 4] + sigma∗random.normalvariate(0, 1)
```

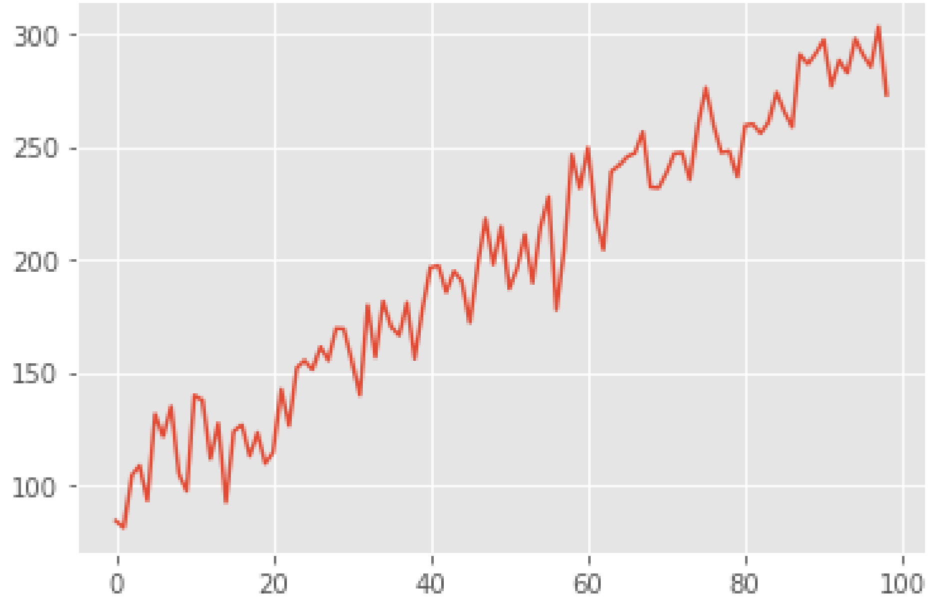


Time series: simple forecasts

iii) a model with linear trend

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

```
c=100; sigma=15; b=2;
ytr=[0]*101;
for i in range(1,100):
 ytr[i] =c + b*i+ sigma*random.normalvariate(0, 1)
```

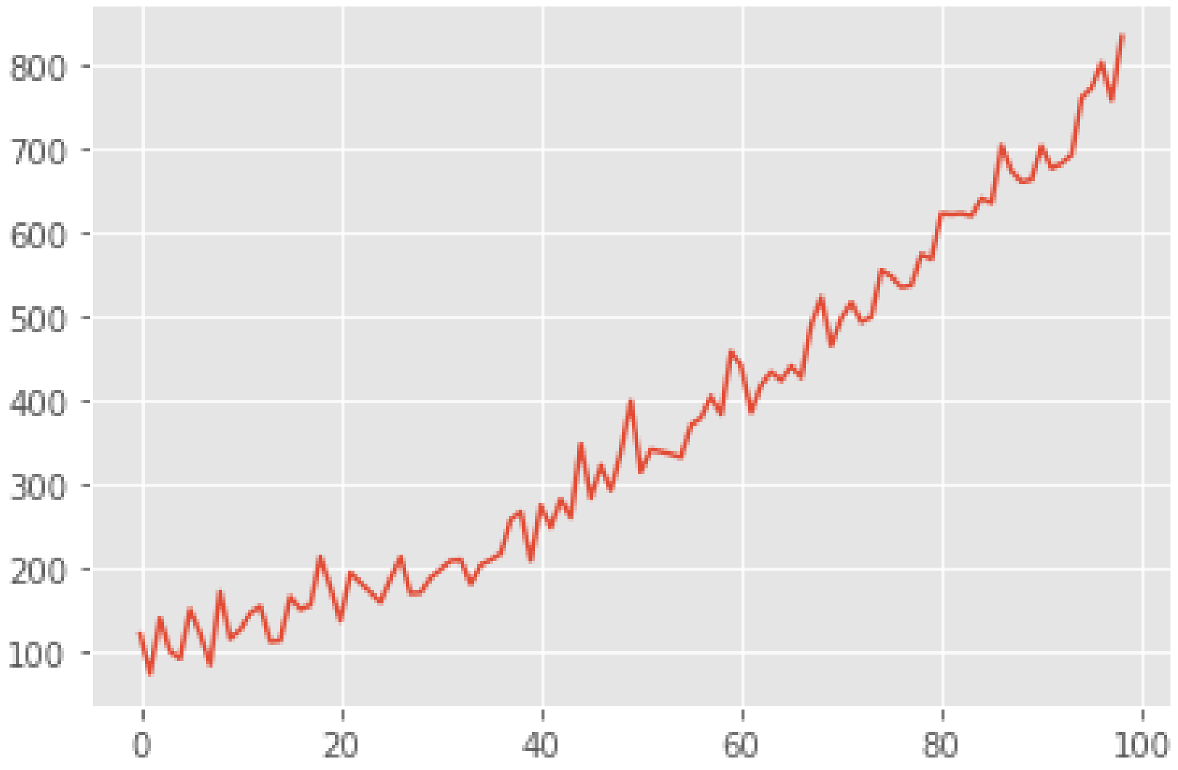


Time series: simple forecasts

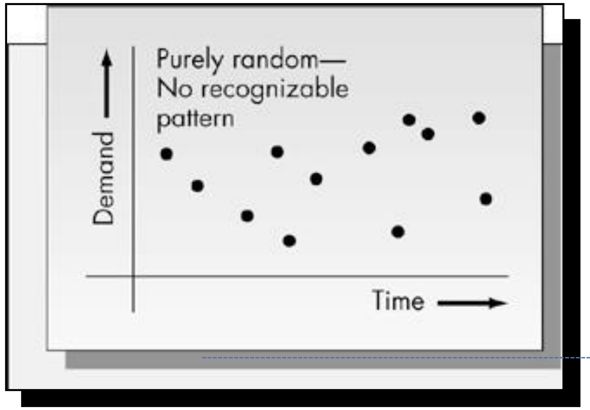
iv) a model with quadratic trend

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

```
c=100; sigma=25; b=2; a=0.05;
yqr=[0]*101;
for i in range(1,100):
 yqr[i] =c + b*i+ a*i*i+ sigma*random.normalvariate(0, 1)
```



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.