Forecast

Thomas Chuffart

thomas.chuffart@univ-fcomte.fr

Introduction

Box-Jenkins Method

- Model **identification** and model **selection** : is the variable stationary ? Is there seasonality ?
- Parameter **estimation** to fit the selected ARIMA model.
- Model Checking by **testing** whether the estimated model **conforms** to the **specifications** of a stationary univariate process.
- **Forecast**.

Introduction

Definition

Given some observations on a, possibly multivariate, variable x, i.e. $x_1, \ldots, x_N$, we want to find a good approximation to the **unknown** observation x_{T+h}.

The information set available at $t = T$ for the forecast necessarily includes the observed time series but it may be much larger in practice.

Introduction

What is a **forecast** ?

- "Anything can be forecast, but not everything can be predicted."
- **Predict** : implies inference from laws of nature, it is a theoretical property
- **Forecast** : more probabilistic, Fore : "in advance", Cast : dice, spells, horoscopes..., it is the exploitation of the theoretical property.
- A forecast is **any statement** about the future.

Introduction

1. Curiosity
2. **Decision making** (VaR, risk measure)
3. Policy evaluation : modified policy may change expectations and thus model behavior (Lucas)
4. **Model evaluation**

Introduction

- Demographic projections (accurate)
- the accuracy of meteorological forecasts (less accurate)
- demand forecasts (less less accurate)
- speculative markets (very less accurate - Sir Granger) ;
- short-run macroeoconomic forecasts

Introduction

Methods for forecasting :

- A crystal ball that can see the future ;
- Model-based forecast.
- Extrapolate from the present information.

Introduction

Methods of Economic forecasting :

- Extrapolation, leading indicators, surveys (**model-free** procedures) ;
- Time series models (**model-based** procedures) ;
- Econometric forecasting models.

Most institutional forecasts use a mix of **subjective** and **objective** elements.

Introduction

Successful forecasting requires that :

- there are stylized facts to be **captured** ;
- these latter one are **informative** about the future ;
- the proposed method captures it and excludes **outliers**.

Two key **assumptions** :

- The econometric model is a **good representation** of the economy ;
- The structure of the economy remains **relatively** constant.

Introduction

If decisions are based on forecasts, they may affect the forecasted variables. Effects can be positive or negative :

- **self-fulfilling** forecasts : a bad growth forecast may cause pessimism and decrease demand ; a high inflation forecast may raise incentives for wage bargaining ;
- **self-defeating** forecasts : a high unemployment forecast may cause active labor market policies ; a high inflation forecast may cause central banks to implement anti-inflationary policies.

A good excuse for **inaccurate** economic forecasts ?

Introduction

It exists two kind of forecasts : **out-of-sample** or **in-sample**

- an *in-sample* forecast observation in the sample using to fit the data. Forecast errors will be residuals, not true prediction errors.
- a true *out-of-sample* forecast x_{T+h} only uses information over the time range $t \leq T$.

Introduction

- *Journal of forecasting* and *International journal of forecasting* ;
- *Time series forecasting*, CHATFIELD (2001)
- *Forecasting economic time series*, CLEMENTS and HENDRY (1998)
- *Evaluating Econometric Forecasts of Economic and Financial Variables*, CLEMENTS (2005)

Evaluation

- Find a forecast subject (Macroeconomic previsions for example), the one in the time series project
- Collect and clean the data (FED `https://fred.stlouisfed.org/`)
- Make a literature review
- Model the data
- Forecast
- About 10 pages with R code
- You will be alone on that

Evaluation

Advices :

- Work on a subject that you like
- Simple and clean is better than complex and noisy
- Show that you understand what you do (no negative Fisher statistic)
- Ask google, stackexchange, books, me ...

Part 1 : Model based methods

- Univariate forecasting methods
- Risk measures

"Forecasting is like driving a car blindfolded with help from someone looking out of the rear window"

Reminders

- A time series can have a **trend** (stochastic or deterministic) and a seasonal pattern. The first task to the econometrician is to **remove** it.
- If futures values can be predicted exactly from past values, then a series said to be **deterministic**.
- A stochastic process is denoted by $(X_1, X_2, \dots)$ or more generally by $\{X_t, t \in T\}$, in discrete time.
- We restrict attention to real-valued series which are **causal** (value at time t depends of past value).

Reminders

- The observed value at time t, namely x_t, as an observation on an underlying random variable, X_t
- The process should be stationary process before estimating and forecasting.
- The properties of the underlying model do not change through time.
- A stochastic process is said to be second-order stationary if its first two moments are finite and do not change through time

Reminders

Example

- The purely random process : a sequence of uncorrelated, identically distributed random variables (White noise).
- Autoregressive processes : $X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t$
- Movering Average processes : $X_t = \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t$

Reminders

The ARIMA class of models is an important forecasting tool :

$$\Phi(L)(1-L)^d X_t = \Theta(L)\varepsilon_t$$

with

- $\phi(L)$ is a polynomial in L of order p. It has a unique stationary causal solution ig the roots of $\phi(x) = 0$ lie outside the unique circle.
- $\theta(L)$ is a polynomial in L of order q. To ensure invertibility, the roots of $\theta(x) = 0$ must lie outside the unit circle.
- $(1-L)^d X_t$ is the order d *differencing* transformation of apply to X_t to handle non-stationary series.

Reminders

There is **no reason** why real-life generating processes should all be **linear** :

- **asymmetric** phenomena may arise with economic series : different behavior in recession ?
- many financial time series show periods of stability, **followed** by unstable periods with high volatility.
- We will see latter the usefulness of GARCH models in **risk measures**.

The prediction problem

Definition

Let denote any univariate forecast of X_{T+h} by $\hat{x}_T(h)$ with $\hat{x}_T(h) = g(x_T, x_{T-1}, \ldots)$.

- Which specification for g and which properties ?
- How to define a good forecast ?

The prediction problem

- The model-based methods **assume** a DGP (Data Generating Process)
- "All models are wrong, some are useful", Box
- *useful* may refer to forecasting, *model* refers to the function g
- The **true model class** does not guarantee the best forecasting performance.

The prediction problem

Definition

Forecasts are evaluated with some loss functions defined by :

$$L(e) = f(\text{observed value} - \text{forecast})$$

with e the forecast error.

- It computes the loss associated with a forecast error of size e.
- $L(0) = 0$ and $L(e)$ is a continuous function which increases with $|e|$

The prediction problem

Example

- Quadratic loss function : $L(e) = ce^2$
- Absolute loss function : $L(e) = c|e|$

The best forecast minimize the loss function. The choice of loss function has to be set carefully (Patton 2011 - JoE)

The prediction problem

The Mean Square Error (MSE) is a special case of a quadratic loss function :

$$MSE: \; E[(x_{T+h} - x_T(h))^2]$$

- What is the solution of the minimization of the MSE ?

ARMA forecasting

- To compute **forecast**, we choose the expected value of X_{T+h} conditional on the unknown true DGP and the information available at time T
- $x_T(h) = E[x_{T+h}|I_T, M_0]$
- In practice, we use the **fitted** model.

ARMA forecasting

Example with a simple AR(1) model

$$X_t = \phi X_{t-1} + \varepsilon_t$$

- $E[X_{T+1}|I_T] = \phi X_T$
- $E[X_{T+2}|I_T] = \phi X_T(1) = \phi\left[\phi X_T\right]$
- $E[X_{T+3}|I_T] = \phi X_T(2) = \phi\left[\phi\left[\phi X_T\right]\right]$
- ...
- $E[X_{T+h}|I_T] = \phi^h X_T$

Do the same for $X_t = \mu + \phi X_{t-1} + \varepsilon_t$.

ARMA forecasting

Example with a simple MA(2) model

$$X_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \varepsilon_t$$

- $E[X_{T+1}|I_T] = E\left[\theta_1 \varepsilon_T + \theta_2 \varepsilon_{T-1} + \varepsilon_{T+1}|I_T\right] = \theta_1 \varepsilon_T + \theta_2 \varepsilon_{T-1}$
- $E[X_{T+2}|I_T] = E\left[\theta_1 \varepsilon_{T+1} + \theta_2 \varepsilon_T + \varepsilon_{T+2}|I_T\right] = \theta_2 \varepsilon_T$
- $E[X_{T+3}|I_T] = ?$
- ...

Do we know $\{\varepsilon_t\}$?

ARMA forecasting

No, so we use the $AR(\infty)$ representation of the $MA(q)$ process :

- $X_t = \Theta(B)(L)\varepsilon_t \Leftrightarrow \Theta^{-1}(L)X_t = \varepsilon_t$
- $X_t = \sum_{k=1}^{\infty} a_k X_{t-k} + \varepsilon_t$
- so $X_{t+h} = \sum_{k=1}^{\infty} a_k X_{t+h-k} + \varepsilon_{t+h}$
- thus for large T :
  $x_T(h) = \sum_{k=1}^{h-1} a_k x_T(h-k) + \sum_{k=h}^{T+h} a_k X_{T+h-k}$

ARMA forecasting

For the AR(2) process above, the forecast error is :

- $e_{T+1} = X_{T+1} - X_T(1) = \phi_1 X_T + \phi_2 X_{T-1} + \varepsilon_{T+1} - \phi_1 X_T - \phi_2 X_{T-1} = \varepsilon_{T+1}$
- $e_{T+2} = X_{T+2} - X_T(2) = \phi_1 X_{T+1} + \phi_2 X_T + \varepsilon_{T+2} - \phi_2 X_T = \phi_1 X_{T+1} + \varepsilon_{T+2}$

Forecast error variance :

- $Var(e_{T+1}) = E\left[(X_{T+1} - X_T(1))^2\right] = E\left[\varepsilon_{T+1}^2\right] = \sigma^2$
- $Var(e_{T+2}) = E\left[(X_{T+1} - X_T(1))^2\right] = E\left[(\phi_1 X_{T+1} + \varepsilon_{T+2})^2\right]$

ARMA forecasting

Proposition

The MSE increases with the forecast horizon up until $h = q$. For $h > q$ the forecast is the unconditional mean and the MSE is the unconditional variance of the series.

Show it for the AR(1) process.

ARIMA(p,d,q) forecasting

The class of integrated models is the most popular class of non-stationary time-series models.

- An integrated process $\{X_t\}$ is defined by the property that is is not stationary but d-th differences are. Notation : $(\Delta^d X_t)$. Only $d = 1$ and $d = 2$ are of empirical interest.
- Integrated processes model well random walk trends but not structural breaks, outliers, and other observed non-stationary features

ARIMA(p,d,q) forecasting

- *Box & Jenkins* called a process $\{X_t\}$, ARIMA(p,d,q) if $\Delta^d X_t$ is a stable and well-defined ARMA(p,q) process but $\Delta^{d-1} X_t$ is not.
- *Engle & Granger* called a process $\{X_t\}$ d-th order integrated, in symbols $I(d)$ if $\Delta^d X_t$ is stationary but $\Delta^{d-1} X_t$ is not stationary.
- $\Delta X_t = X_t - X_{t-1}$ and $\Delta^2 X_t = X_t - 2X_{t-1} + X_{t-2}$.

ARIMA(p,d,q) forecasting

1. take d–th order differences ;
2. fit ARMA models to the differenced data ;
3. forecast according to the identified ARMA structures ;
4. possibly integrate back (accumulate) to obtain forecasts for the original variable.

ARIMA(p,d,q) forecasting - Confidence interval

- Forecast errors are given by $X_{t+h} - x_T(h) = \sum_{j=0}^{h-1} \pi_j \varepsilon_{t+h-j}$
- How to determinate the confidence interval of $\hat{x}_T(h)$?
- Under normality assumptions, we show :

$$\frac{X_{t+h} - \hat{x}_T(h)}{Var(X_{t+h} - \hat{x}_T(k))^{1/2}} \xrightarrow{L} N(0,1)$$

- We know
$E\{[X_{t+h} - x_T(h)]^2\} = E\left[\left(\sum_{j=0}^{h-1} \pi_j \varepsilon_{t+h-j}\right)^2\right] = \sum_{j=0}^{h-1} \pi_j^2 \sigma_\varepsilon^2$
- And thus $IC = \left[\hat{x}_T(k) \pm t^{\alpha/2} \left(\sum_{j=0}^{k-1} \pi_j^2\right)^{1/2} \hat{\sigma}_\varepsilon^2\right]$

Reminders

The original ARCH model by Engle (1982) assumes, in its that X_t is white noise.

- $X_t = \mu + \varepsilon_t$
- $\varepsilon_t = \sqrt{h_t}\eta_t$ with $\eta_t \sim IID(0,1)$
- $h_t = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2$

The error term of $\{X_t\}$ is said to be an ARCH(q) process. The model is stable with $Var(X_t) < +\infty$:

- $\omega > 0$
- $\alpha_i \geq 0$
- $\sum_{i=1}^{q} \alpha_i < 1$

Reminders

The GARCH model, Bollerslev (1986) extends the ARCH model :

- $X_t = \mu + \varepsilon_t$
- $\varepsilon_t = \sqrt{h_t}\eta_t$ with $\eta_t \sim IID(0,1)$
- $h_t = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}$

The error term of $\{X_t\}$ is said to be an ARCH(q) process. The model is stable with $Var(X_t) < +\infty$:

- $\omega > 0$
- $\alpha_i \geq 0$, $\beta_i \geq 0$
- $\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i < 1$

Reminders

- $Var(X_t) = Var(\varepsilon_t) = E(\varepsilon_t^2) - E(\varepsilon_t)^2 = E(h_t \eta_t^2)$
- Thus, $Var(X_t) = \frac{\omega}{1 - \sum_{i=1}^{q} \alpha_i - \sum_{i=1}^{p} \beta_i}$.
- A GARCH(1,1) is like an ARMA model on squared residuals ε_t^2.

Reminders

To the time-series forecaster who models serially correlated variables, the most interesting extensions are ARMA-GARCH models :

- $X_t = \mu + \phi X_{t-1} + \varepsilon_t$
- $\varepsilon_t = \sqrt{h_t}\eta_t$ with $\eta_t \sim IID(0,1)$
- $h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$

Likelihood estimation in two stages.

Volatility forecasts

Why forecast volatility ?

- Risk management : to measure the potential future losses of a portfolio of assets ;
- Asset allocation : the Markowitz approach of minimizing risk for a given level of expected returns has become a standard approach ;
- Taking bets on future volatility.

Volatility forecasts

From the formula of the unconditional variance we can rewrite the GARCH equation :

- $h_t = (1 - \alpha_1 - \beta_1)E[\sigma^2] + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$;
- It is easy to see that next period's conditional variance is a weighted combination of the unconditional variance of returns, $E[\sigma^2]$, last period's squared residuals, ε_{t-1}^2 and last period's conditional variance h_{t-1}, with weights which sum to one.

Volatility forecasts

One-step forecast :

- $h_{T+1} = \omega + \alpha_1 E[\varepsilon_T^2 | I_T] + \beta_1 h_T = \omega + \alpha_1 h_T + \beta_1 h_T = \sigma^2 + (\alpha_1 + \beta_1)(h_T - \sigma^2)$;

Two-steps forecast :

- $h_{T+2} = \omega + \alpha_1 E[\varepsilon_{T+1}^2 | I_T] + \beta_1 h_{T+1} = \omega + \alpha_1(\sigma^2 + (\alpha_1 + \beta_1)(h_T - \sigma^2)) + \beta_1 h_{T+1} = \sigma^2 + (\alpha_1 + \beta_1)^2 (h_T - \sigma^2)$;

Code R

- A simple forecast ;
- a rolling forecast ;
- a rolling forecast with reestimation of model.

Reference manual, source code and the vignette are available at : `http://cran.r-project.org/web/packages/rugarch/index.html`
The author of the rugarch package Alexios Ghalanos has a blog : `http://www.unstarched.net/blog/`

Code R

To fit a GARCH-model the first step is to create an object from *uGARCHspec* It serves the purpose of specifying the model to be estimated.

```
model=ugarchspec(
variance.model = list(model = "sGARCH", garchOrder = c(1, 1)),
mean.model = list(armaOrder = c(1, 1), include.mean = TRUE),
distribution.model = "norm")
```

Code R

```
getSymbols('^GSPC', from='2012-01-01', to='2018-01-01')
stock_prices = GSPC[,4]
stock_returns =  diff(log(stock_prices),lag=1)*100
stock_returns = stock_returns[!is.na(stock_returns)]
fit=ugarchfit(spec=model,data=stock_returns)
modelfor = ugarchforecast(fit, data = NULL, n.ahead = 10, n.roll = 0,out.sample = 0)
fitted(modelfor)
sigma(modelfor)
```

Introduction

The Value-at-Risk is a measure of risk that tries to answer the following question :

- How bad can things get ?
- Developed by JP-Morgan in the 90s.
- Expected Shortfall (or C-VaR) is the expected loss given that the loss is greater than the VaR level

Definition

In a more probabilist way : With probability α, I will not lose more than x euros in time T with :

- α the confidence level. In theory, it may be freely chosen by the risk manager. In reality, it is often determined by the law or other prescriptions (e.g. Basel II-III).
- x the VaR ;
- T the time horizon.

$$VaR(\alpha) = F^{-1}(\alpha)$$

Example

If the profit of a portfolio during 3 month assuming returns are normally distributed with mean 2 Millions and standard error 5 Millions. What is VaR for $\alpha = 99\%$?

Definition

VaR can be computed using two different distributions :

- gain distribution, where a loss is a negative gain.
- losses distribution.

$$VaR(\alpha) = inf\{P(L \geq l) \leq 1 - \alpha\}$$

Definition

Underlying loss distribution is not known :

- directly estimate the associated quantile of historical data
- estimate model for underlying loss distribution, and evaluate inverse cdf at required quantile

Derivation of VaR from a model for the loss distribution can be further decomposed :

- analytical solution for quantile
- Monte Carlo Simulation when analytic formulas are not available

Modeling the loss distribution inevitably entails model risk, which is concerned with possibly misleading results due to model misspecification

Definition

The VaR presented is based on the unconditional distribution :

$$VaR(\alpha) = F^{-1}(\alpha)$$

R is the return of a financial asset, we suppose that this return is a random variable with a density $f_R(r)$. We can define the conditional density of this variable : $f_R(r|\Omega)$.

$$VaR(\alpha) = F_R^{-1}(\alpha|\Omega)$$

Value-at-Risk with GARCH forecast

- $X_t = \mu + \varepsilon_t$
- $\varepsilon_t = \sqrt{h_t}\eta_t$ with $\eta_t \sim N(0,1)$
- $h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$

You can forecast h_{T+1}. Thus

$$P\left[r_{t+1} < VaR_{T+1|T}(\alpha)|\Omega_T\right] = \alpha$$

$$\Rightarrow P\left[\eta_{t+1} < \frac{VaR_{T+1|T}(\alpha) - \mu}{\sqrt{h_{T+1}}}|\Omega_T\right] = \alpha$$

Part 1 : Model free methods

- Extrapolation
- Surveys

Part 1 : Extrapolation

Extrapolation methods are reliable, objective, inexpensive, quick, and easily automated :

- it is widely used, especially for inventory and production forecasts, for long-term forecasts in some situations.
- The basic assumption is that the variable will continue in the future as it has behaved in the past.
- can be used for cross-sectional data. The assumption is that the behavior of some actors at a given time can be used to extrapolate the behavior of others.

Extrapolation : drawbacks

Clean up the data :

- Obtain data that represent the forecast situation ;
- Use all relevant data, especially for long-term forecasts ;
- Structure the problem to use the forecaster's domain knowledge ;
- Clean the data to reduce measurement error ;
- Adjust data for historical events.

Extrapolation : first stage

- Intuition : the value of tomorrow is the average of past observations
- $\hat{y}_{T+1|T} = \frac{1}{T}\sum_{t=1}^{T} y_t$
- All the observations have the same weight.

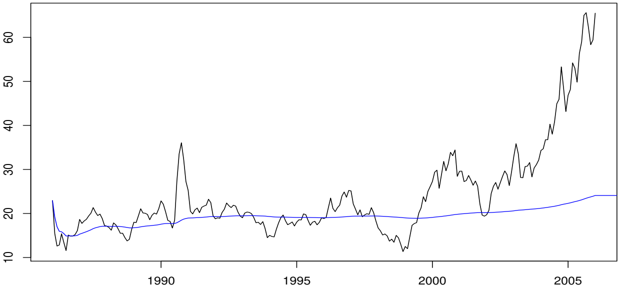
Extrapolation :R code

```r
library("TSA")
data("oil.price")
plot(oil.price)
hfor = 10
T = length(oil.price)
m_oilprice = cbind(rep(0, length(oil.price)+hfor))

for (t in 1:length(oil.price)){
m_oilprice[t] = mean(oil.price[1:t])}
for (t in 1:hfor){
m_oilprice[t+length(oil.price)] = mean(oil.price)}

m_oilprice = ts(m_oilprice, start = c(1986,1), frequency = 12)
plot(oil.price, xlab = "", ylab = "")
lines(m_oilprice, col = "blue")
```

Extrapolation : naïve methods



Extrapolation

- Intuition : the value of tomorrow is the average of last observations
- We can, for example, make the mean of the last five observations.
- How to choose the number of observations to take account ?

Extrapolation : naïve methods

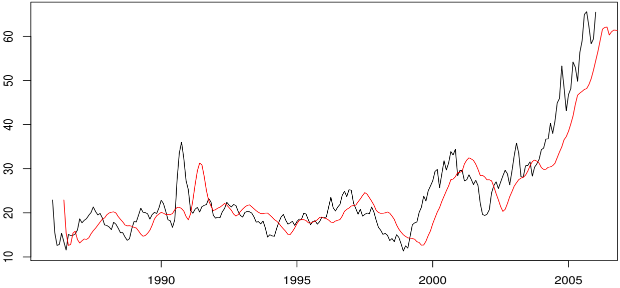
```
m_oilprice_last = cbind(rep(0, length(oil.price)+hfor))
m_oilprice_last[1:5] = oil.price[1:5]

for (t in 5:length(oil.price)){
i = t-4
m_oilprice_last[t+1] = mean(oil.price[i:t])}

for (t in 1:hfor){
i = T-5+t
j = T+t-1
m_oilprice_last[t+length(oil.price)] = mean(m_oilprice_last[i:j])}

m_oilprice_last = ts(m_oilprice_last, start = c(1986,6), frequency = 12)
plot(oil.price, xlab = "", ylab = "")
lines(m_oilprice_last, col = "red")
```

Extrapolation : naïve methods



Extrapolation : Exponential smoothing

- a causal filter determines the filtered value of a data point $\hat{x}_T$ from observations x_T, $t < T$;
- a smoother determines the smoothed value of a data point $\hat{x}_T$ from observation x_t, $t < T$ and $t \geq T$;

Extrapolation calculates a smoother in order to apply a causal filter afterwards.

Extrapolation : Exponential smoothing

The most simple exponential smoothing is the Single Exponential Smoothing (SES). It determines the filtered value $\hat{x}_t$ from a weighted average over a past observation and a past filtered value :

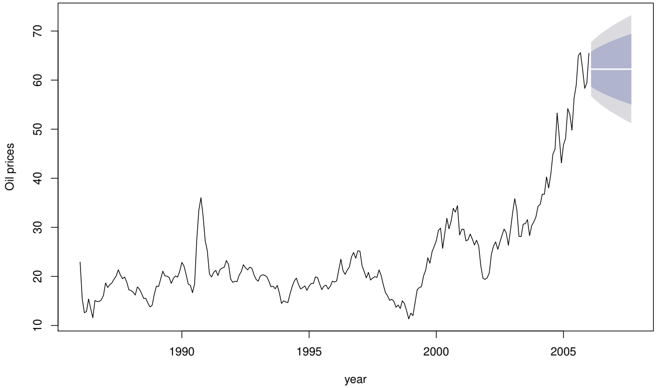
$$\hat{x}_t = \alpha x_t + (1-\alpha)\hat{x}_{t-1}$$

with $\alpha \in (0, 1)$.

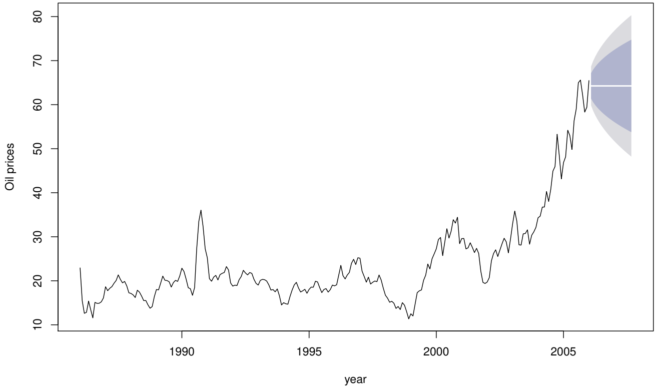
Extrapolation : Exponential smoothing

```
library("forecast")
fit1 <- ses(oil.price, alpha=0.4, initial="simple", h=20)
fit2 <- ses(oil.price, alpha=0.8), initial="simple", h= 20)
plot(fit1, ylab="Oil prices", xlab = "year", main="", fcol="white", type="o")
plot(fit2, ylab="Oil prices", xlab = "year", main="", fcol="white", type="o")
accuracy(fit1)
```

Extrapolation : Exponential smoothing



Extrapolation : Exponential smoothing



Extrapolation : Exponential smoothing

- Largely dependent of the choice α ;
- Can be chosen by minimizing the Mean Square Errors (MSE) : $MSE = \frac{1}{T}\sum_{t=2}^{T}(x_{t+1} - \hat{x}_t)^2$.
- Same for x_0.

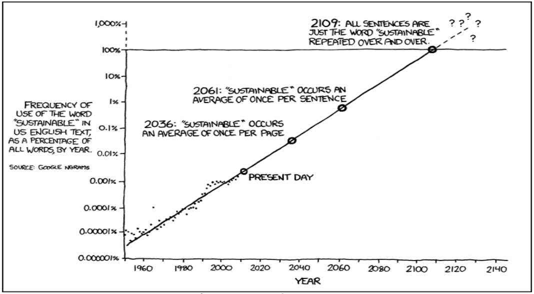
Extrapolation : is it a good method ?

- SES is a quick and simple procedure that can perform well ;
- SES can be shown to be equivalent to a time-series forecast based on ARIMA(0,1,1) model ;
- Many other methods exist to take into account trend and seasonal pattern of a time series(Holt-Winters).

So, it seems to be a good way to forecast... but

Extrapolation : is it a good method ?

- You should always keep in mind that forecasts generally depend on the future being like the past.

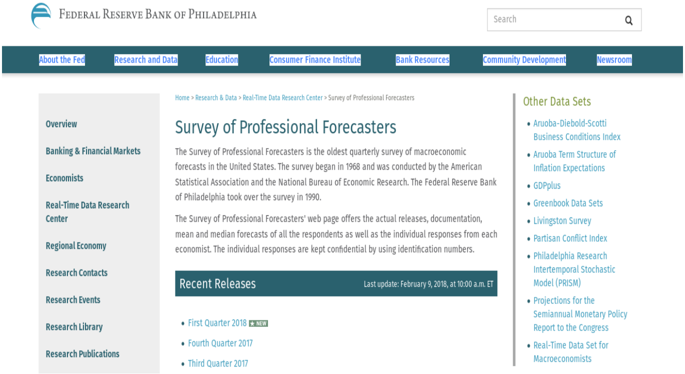


THE WORD "SUSTAINABLE" IS UNSUSTAINABLE.

Survey : a brief summary

- Survey method is one of the most common and direct methods of forecasting in the short term.
- External forecasters answer to a survey about their own forecast.
- An organization conducts surveys to determine their expectations.
- It helps to to forecast the macro economy in the short run.

Survey : a brief summary



Survey : a brief summary

- This information can be used in regression analysis to explain macroeconomic variables.
- It can be used in Vector Autoregressive Model (VAR), to extract shocks of expectations on key variables.
- ...
- add articles