Forecasting / Time Series Workshop (CERN)

Tuesday, Dec 3rd

Table of Content

Forecasting Procedures

- Time series
- Forward-looking methods
  - Capitalizing on information content embedded in derivatives (futures/options)
- Other methods:
  - Expert opinions
- Evaluation and combination of forecasting methods

- Density (probability function) forecasts (estimation and evaluation)
- Context / Domain: Applications for liquidity research

Goal of this meeting

- Sharing our research on forecasting with time series and experiences in commodity market risk management

- Very basic intro to a number of time series models

- Discussion about its usefulness in the context of liquidity measures

Models

- All models are wrong, but some are useful (George E. P. Box)

Time Series Models

- Why do we need and use time series models?
  - Financial data almost always includes a time element (i.e. prices, returns)
    - Thus time series and panel data dominate the empirical approach

- What do we do with models?
  - Description
  - Prediction (forecasting)
  - Establishing causality

Components of an observation

- At time t, we observe certain demand:

  D_t = Systematic Component + Random Component

- Systematic Component: Is the part that can be modeled and used for forecasting
- Random Component: Random processes cannot be predicted

$$F_{t+1} = E(D_t) + E(\varepsilon_t)$$

Component of a forecast / fitted value

$$F_{t+1} = E(D_t) + E(\varepsilon_t)$$

E	=	Expected value
F_t	=	Forecast of demand at period t
D_t	=	Demand at period t
$\varepsilon_t = F_t - D_t$	=	Forecast error at period t

What can be predicted?

The predictability of an event or a quantity depends on:

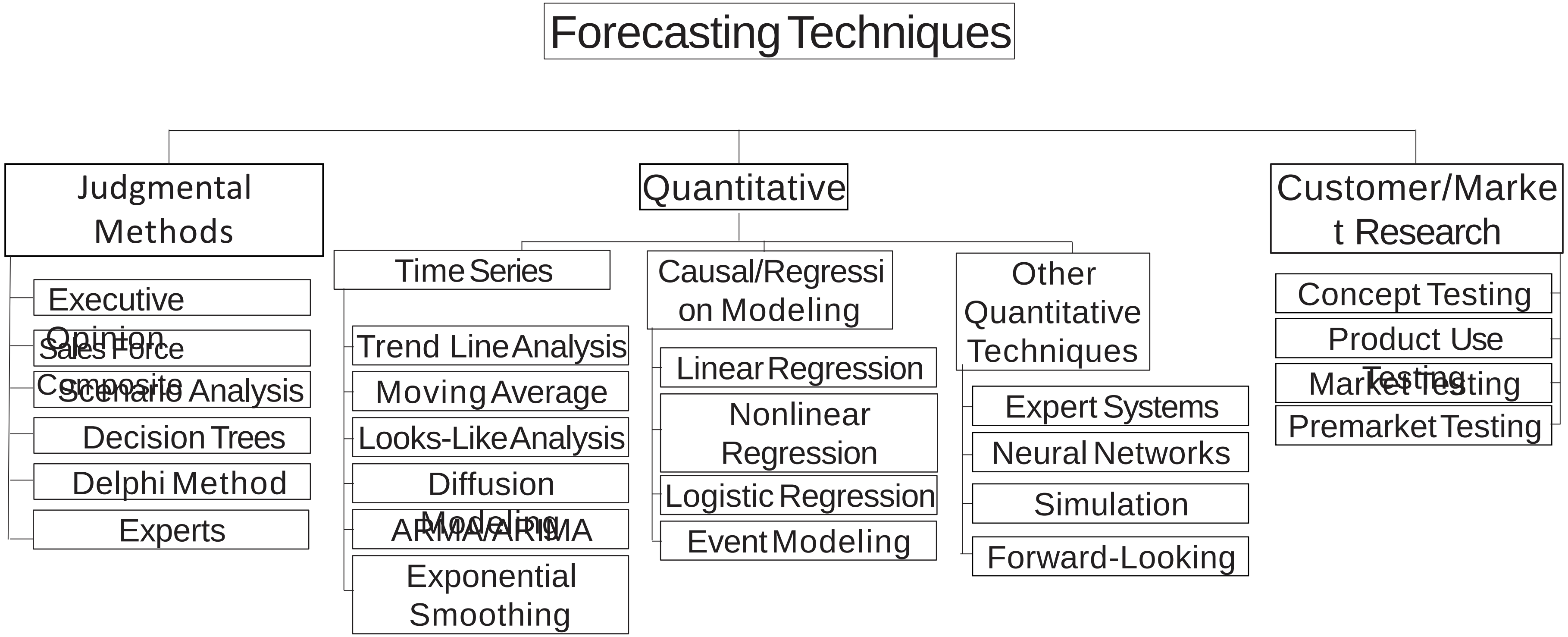
- What component dominates the process? Is it the systematic or the random?
- how well do we understand the factors that contribute to it?
- how much data are available?
- whether the forecasts can affect the thing we are trying to forecast.

What can be predicted

- What time will be the sunrise tomorrow?
- What will be the exchange rate euro/dollar in one month?
- What is the demand of electricity during the next year?
- Lotto number tonight?

- Systematic or random? In between?
- Do we understand the process?
- How much data do we have?
- Does the forecast influence the outcome?

Forecasting methods

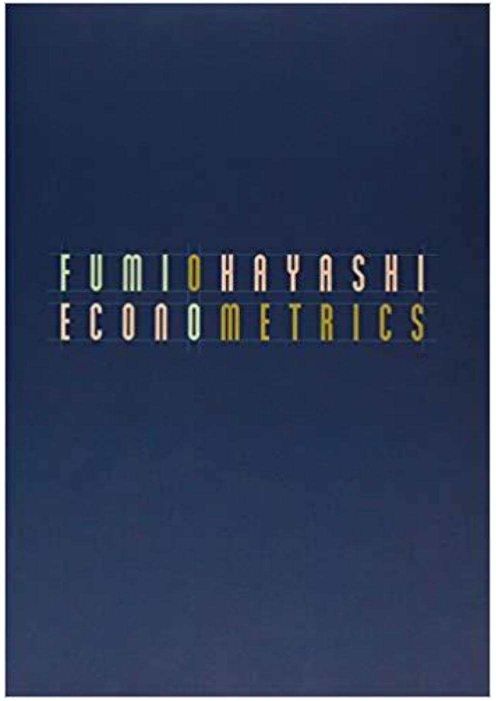


Forecasting methods - Quantitative

Involves mathematical and/or statistical techniques

- Depends on data availability

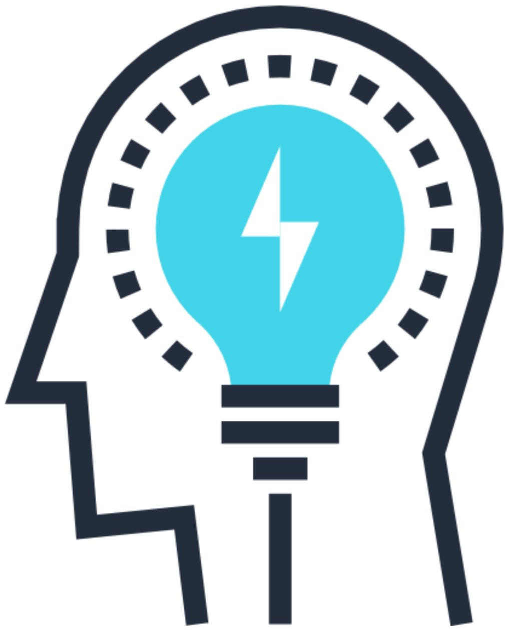
- Forward looking (aggregated market expectations)
- Causal models and time series (best when stable demand)
- Simulations (imitate consumer choices)



Forecasting methods - Qualitative

- Subjective methods, rely on human judgment, intuition, experience and opinion

- Used when situation is vague, little amount of data
  - New technology
  - New products



Experts

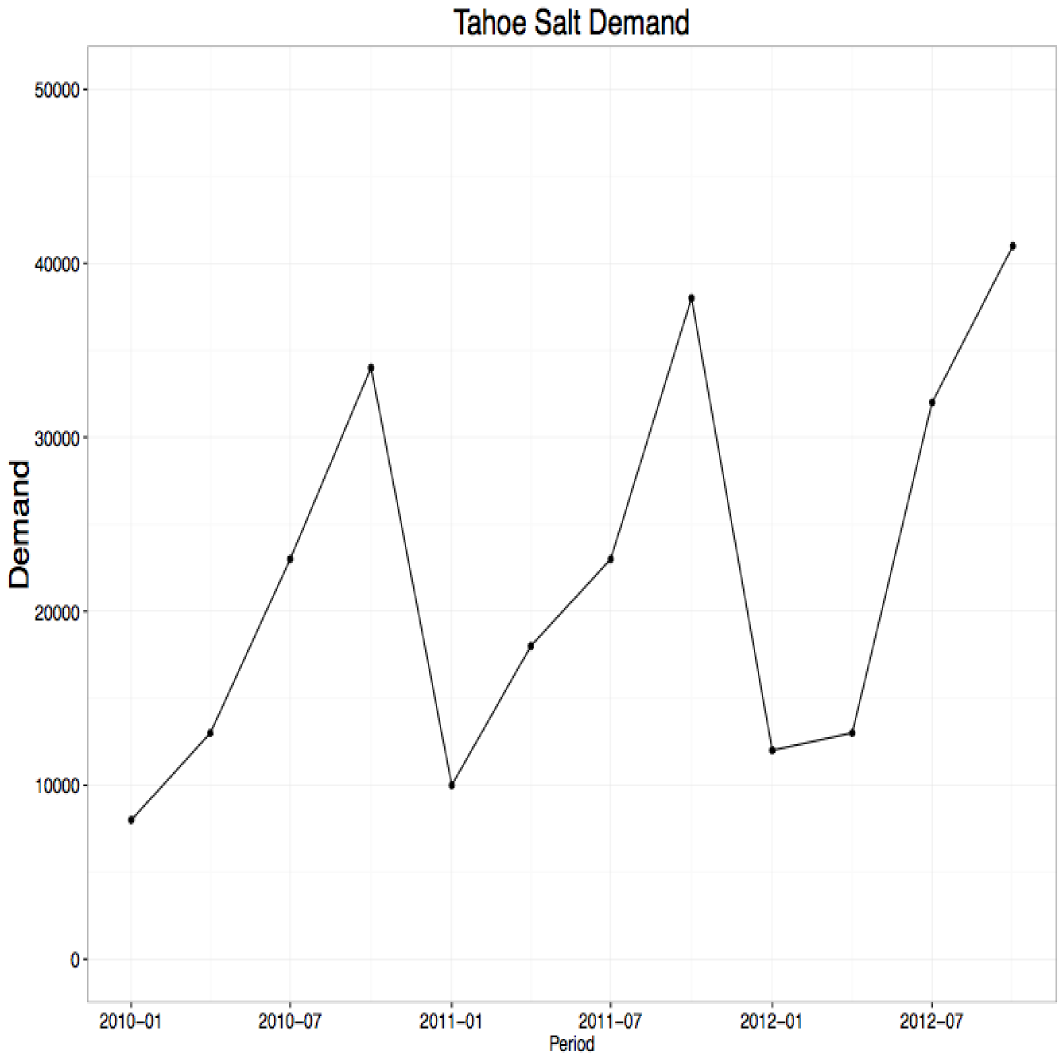
- USDA: Monthly reports on prices, yield, demand
  - Based on surveys to producers and field measurement data
  - Yield based on weather adjusted trends

- Financial analysts

- Advantages: Can account for structural breaks, shocks, rapid changes

Quantitative forecast: Time series

Period	Demand
2010-I	8000
2010-II	13000
2010-III	23000
2010-IV	34000
2011-I	10000
2011-II	18000
2011-III	23000
2011-IV	38000
2012-I	12000
2012-II	13000
2012-III	32000
2012-IV	41000



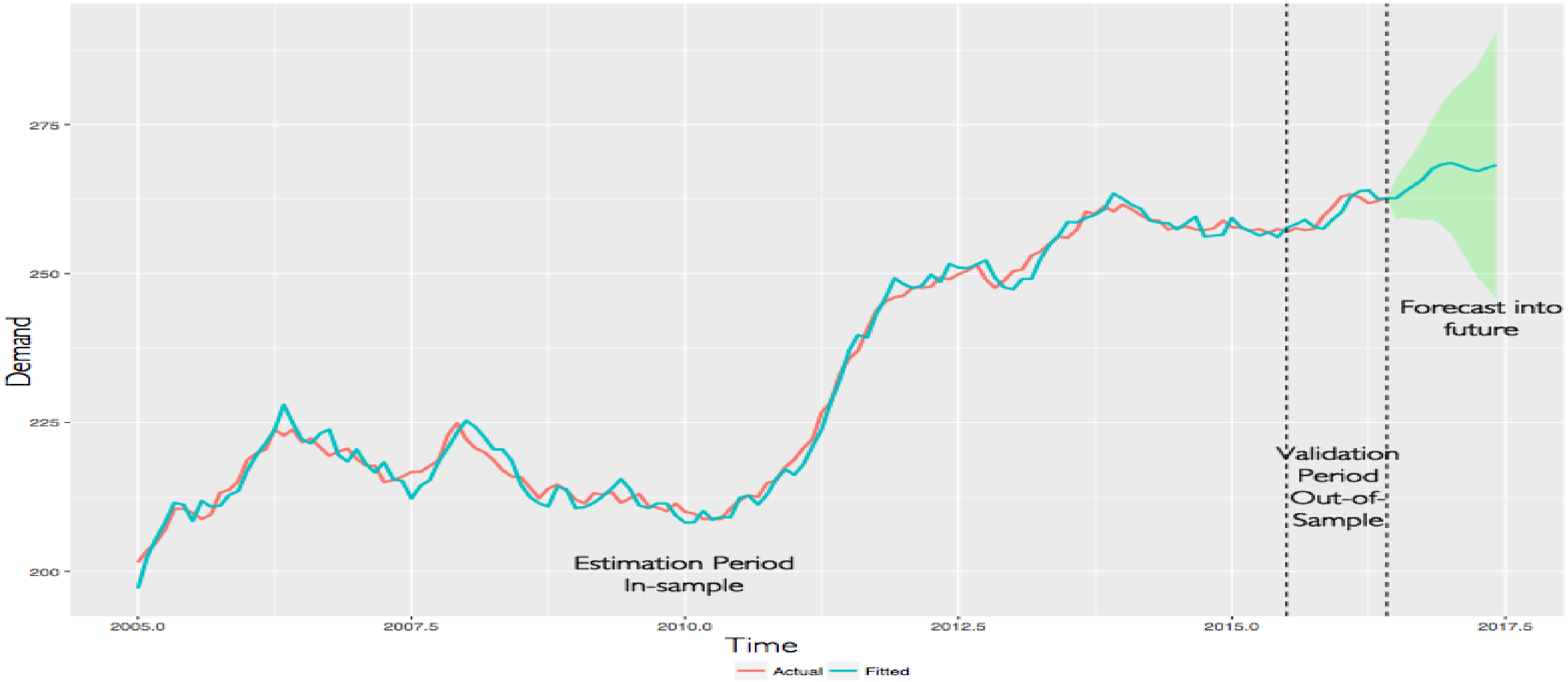
Time series

- A time series is a set of observations, each one recorded at a specific time t
- We use time series when we believe that past data can provide useful information about the future
- Time series models include Exponential Smoothing, Autoregression Moving Average (ARIMA), State-Space Models

Other quantitative forecast models

- Forward looking:
  - Does not assume the future will be like the past, instead tries to capture aggregated expectations from the market
  - Futures and options markets
  - Sentiment analysis from social media (Twitter, google trends)

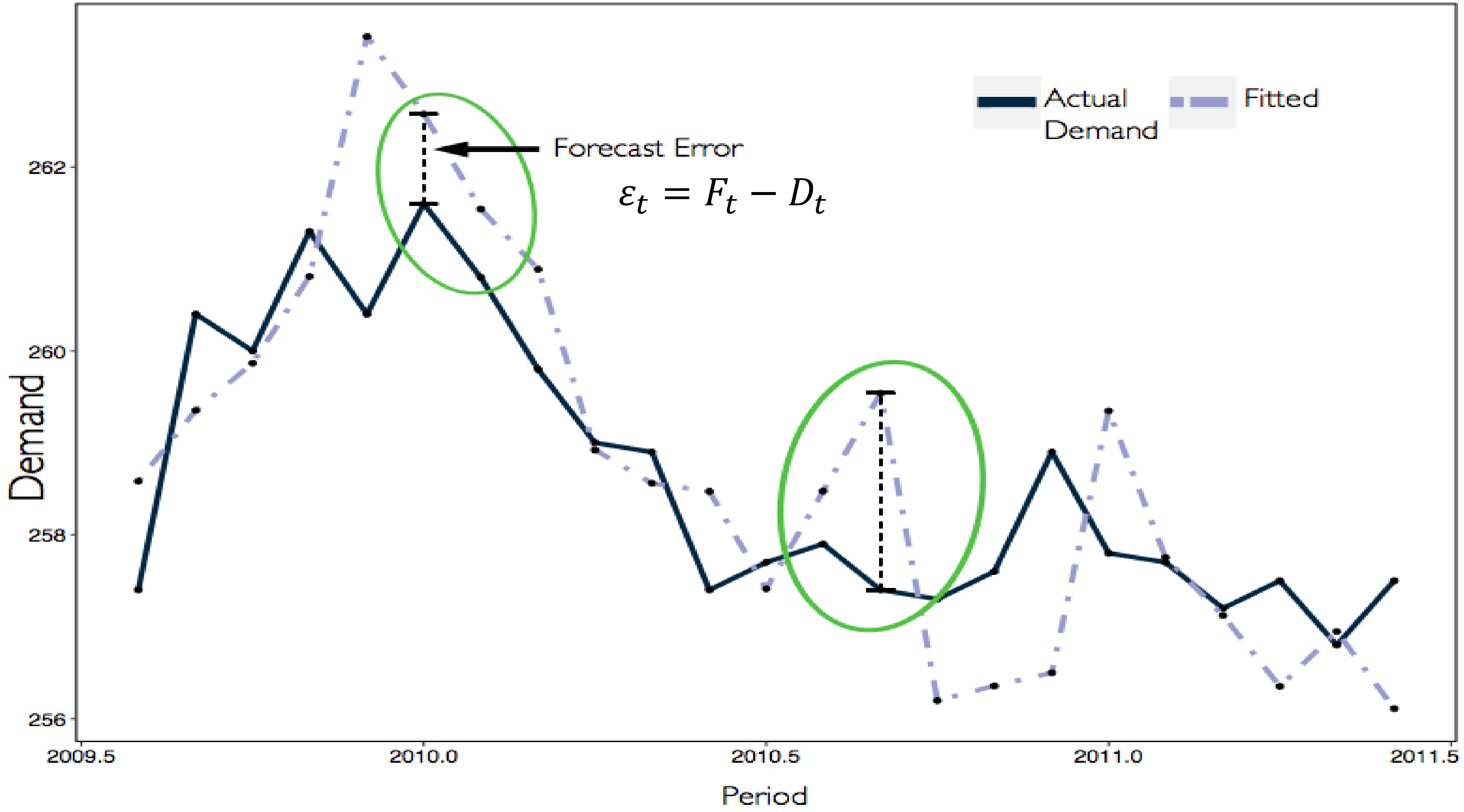
In-sample vs out-of-sample forecasts



- In-sample: fitted values, predictions done within the available data

- Out-of-sample: Data not used to estimate the level, trend, and seasonality (can be used to evaluate the performance of the forecast ex-post)

Forecast error



Evaluation: Forecast error (point forecast)

Forecast error:

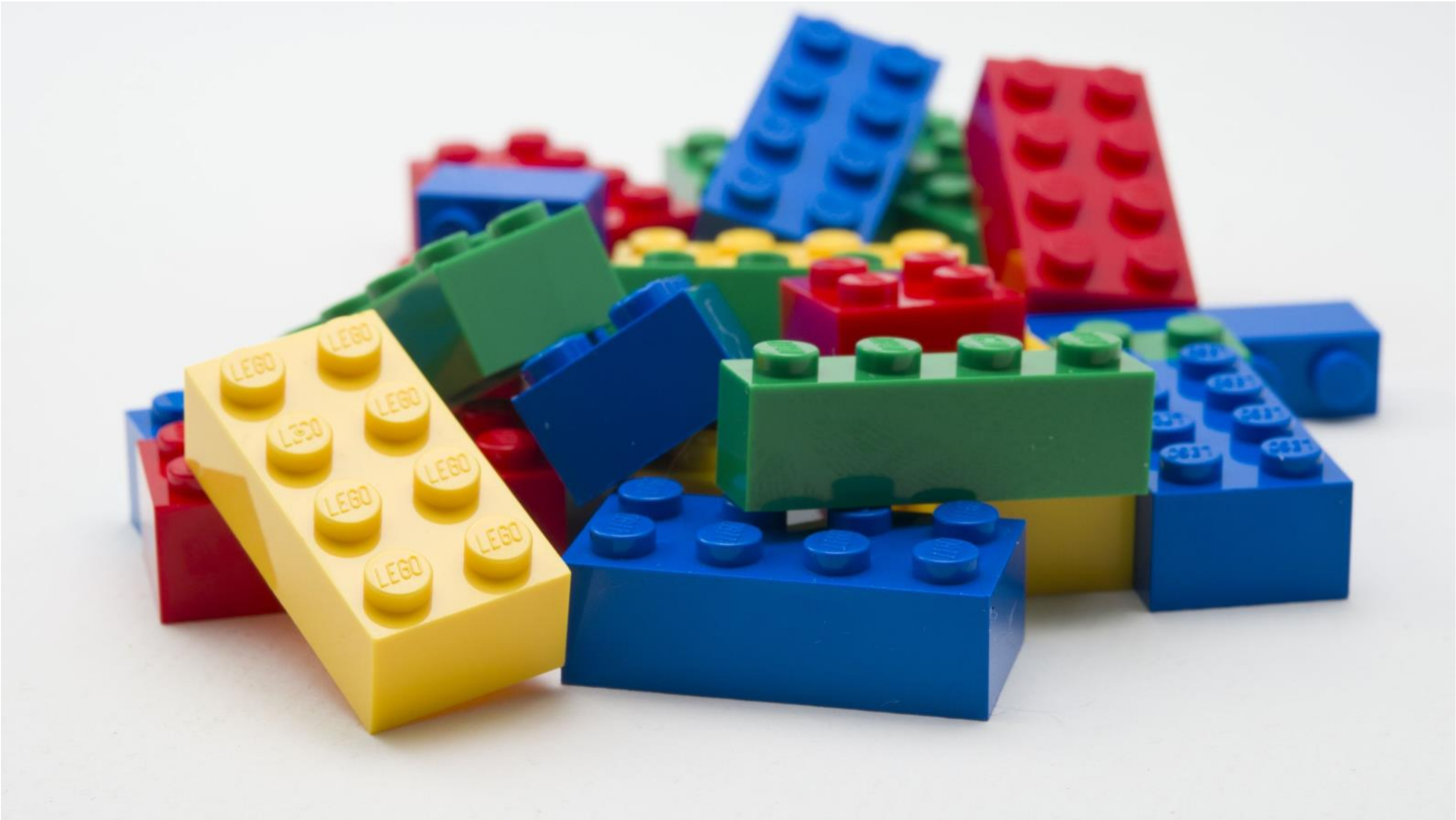
$$\varepsilon_t = F_t - D_t$$

$\varepsilon_t =$	Error in time t:
$F_t =$	Forecast in time t
$D_t =$	Demand in time t

Measures of Forecast Error	
Mean Absolute Deviation (MAD):	$\frac{1}{n}\sum_{t=1}^{n}\lvert\varepsilon_t\rvert$
Mean Square Error (MSE):	$\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^2$
Mean Absolute Percentage error (MAPE):	$\frac{\sum_{t=1}^{n}\left\lvert\frac{\varepsilon_t}{D_t}\right\rvert}{n}\cdot 100$

Forecasting with time series

Building Blocks:

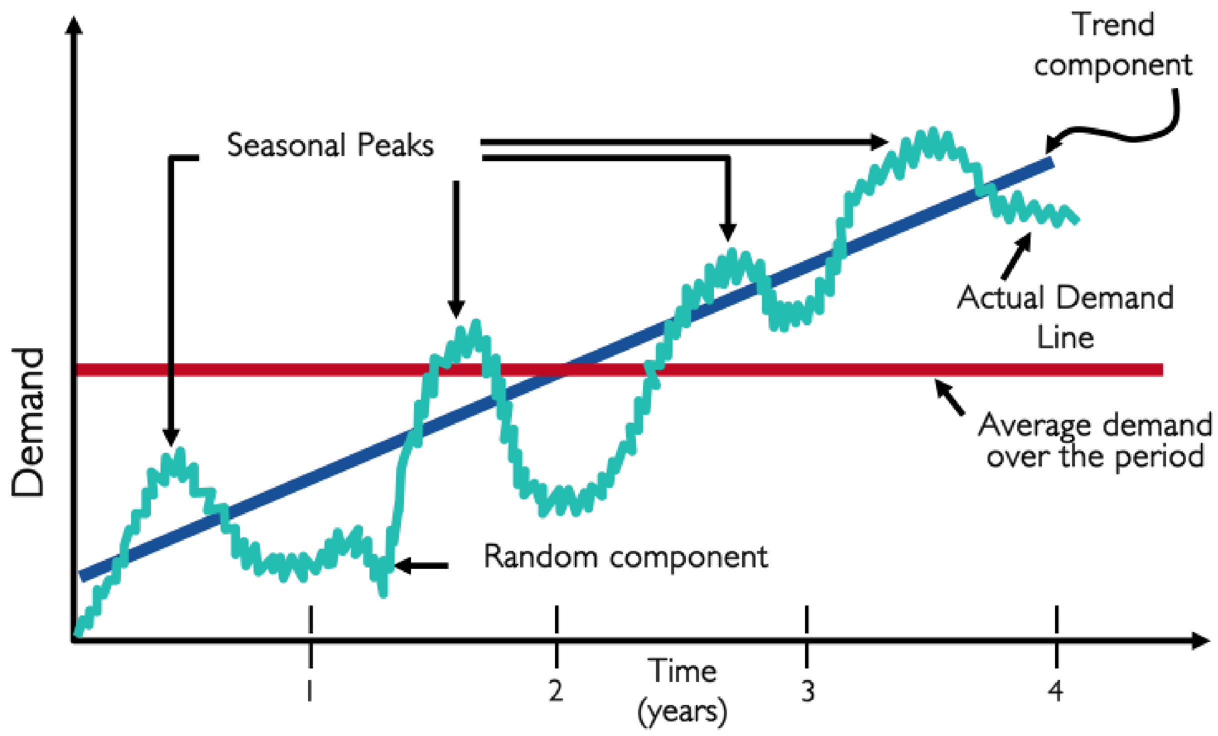


Time series models

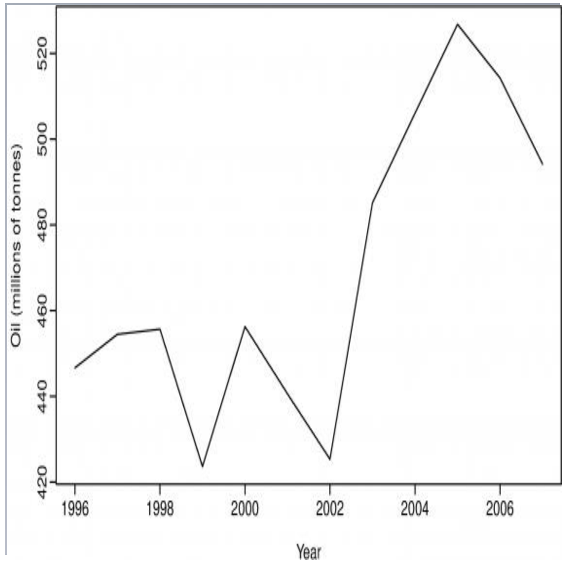
- Many time series models can be used in forecasting
- We will briefly go over basic elements of two models
  - Exponential smoothing
  - ARIMA
- These models and their extensions are among the most popular models in finance, marketing, and supply chain management

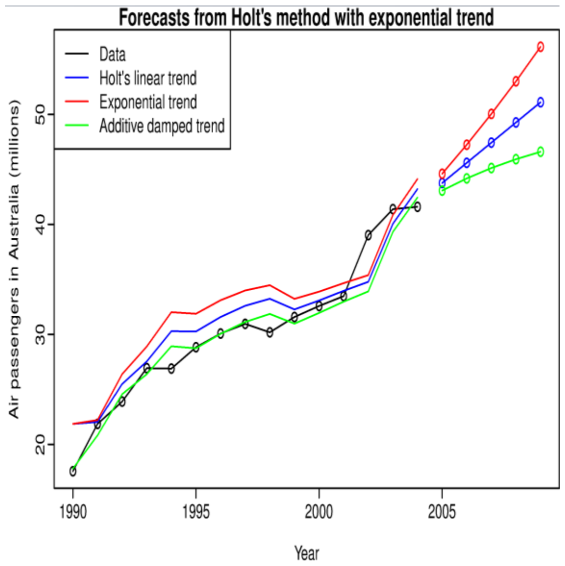
Time series decomposition

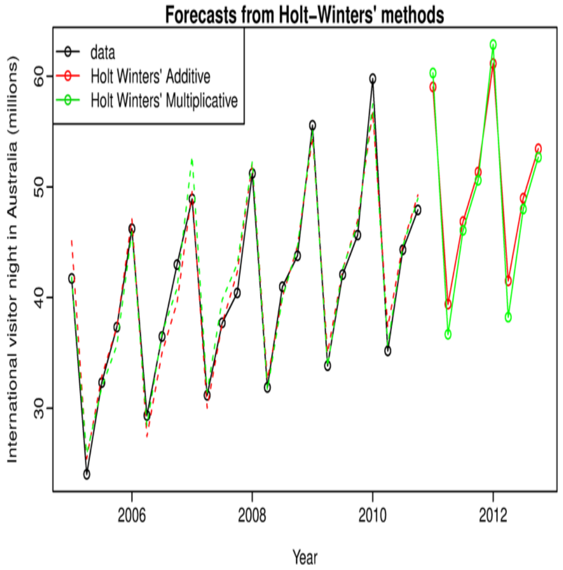
- D_t = *Systematic Component* (X_t) + *Random Component* (ε_t)
- *The systematic component is a function of: Level* (L_t), *Trend* (T_t), *and Seasonality* (S_t)



Exponential smoothing models







Simple Exponential Smoothing	Holt's model	Winter's model
No trend, no seasonality	Trend but not seasonality	Trend and seasonality

Exponential smoothing methods

- Exponential smoothing forecasts are weighted averages of past observations, with the weights decaying exponentially as the observations get older

- Estimates of level, trend, and seasonality are updated after each demand observation

- Estimates incorporate new data that are observed

- Updated observation:

Smoothing Constant * New Information + (1 - Smoothing Constant) * Past Info

Trend and seasonality corrected exponential smoothing

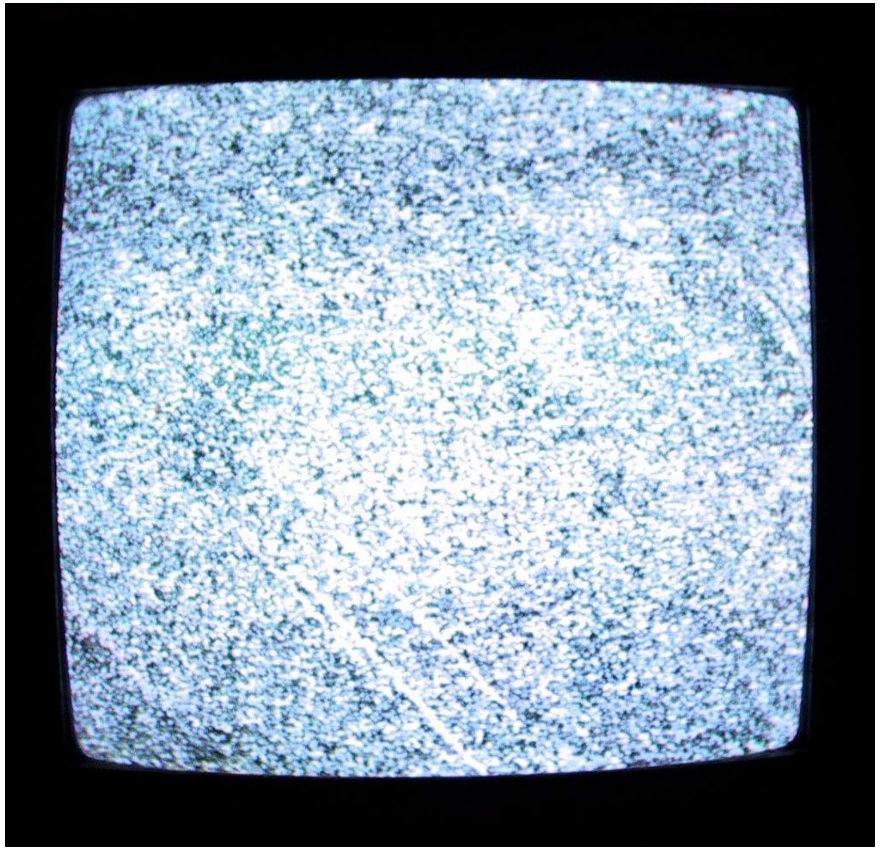
1. Initial estimate of Level (L_0), trend (T_0), and seasonal factors ($S_1, \dots, S_p$) Obtain using the same procedure of the static method	
2. Forecasts:	$F_{t+1} = (L_t + T_t)S_{t+1}$ $F_{t+n} = (L_t + nT_t)S_{t+1}$
3. Smoothing equations:	$L_{t+1} = \alpha(D_{t+1}/S_{t+1}) + (1-\alpha)(L_t + T_t)$ $T_{t+1} = \beta(L_{t+1} - L_t) + (1-\beta)T_t$ $S_{t+p+1} = \gamma(D_{t+1}/L_{t+1}) + (1-\gamma)S_{t+1}$
4. Estimate Error:	$\varepsilon_t = F_t - D_t$

Autoregression Integrated Moving Average (ARIMA)

- ARIMA models are very popular in finance
- Observes the behaviour of a random variable over time
- Based on autocorrelation pattern between observations of the time series

White noise

- Is the simplest model
  - $y_t = e_t$
- The current observation corresponds to a random component
- No information left for prediction
- We may assume its distribution with a mean of zero



Moving averages, MA (q)

- Current observations may be explained as a moving average of past errors

$$y_t = e_t + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q}$$

- A moving average of order q =2 would be:

$$y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2}$$

Autoregression models, AR(p)

- An autoregressive model forecasts the current value of variable (y_t) as a linear function of past observations (lags) of the variable $(y_{t-1}, y_{t-2}, \ldots, y_{t-n})$

- An autoregression model of order p can be written as:

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + e_t$$

- Then, an autoregression of order p =3 would be:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + e_t$$

Random walk

- Is a form of autoregression model AR(1) with $\phi_1 = 1$

$$y_t = y_{t-1} + e_t$$

- A random walk means that the best forecast of tomorrow is today's value

- Error term is like news, things that we cannot predict

- Some people say this is the best forecast for stock and commodity prices

ARMA (p,q) models

- We can combine AR(p) and MA(q) models

- Let's suppose our data can be explained by an ARMA (2,1) then the model would have 2 lags for the autoregression part and one for the moving average

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t + \theta_1 e_{t-1}$$

Let's rewind a bit: Measuring dispersion

- Variance: measures the average amount that data vary from the mean

  $Var\ (x) = S_x^2 = \frac{\sum(x_i - \bar{x})^2}{N-1} = \frac{\sum(x_i - \bar{x})^2}{N-1}$, where x_i are observations $\mathrm{i} = 1, \ldots, \mathrm{n}$ and $\bar{x} = \frac{\sum x_i}{N}$

- The standard deviation is the square root of the variance

  $S_x = \sqrt{Var}(x)$

Note: The following slides are not needed for this audience, but just to be self-contained

Let's rewind a bit: Measuring association

- The simplest way to see whether two variables are associated is to look at their covariance
- Covariance measures the common variation of two variables (x,y)
- $\mathrm{Cov}(\mathrm{x},\mathrm{y}) = S_{xy}\ \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{N-1}$
- We are interested on measuring whether changes in one variable are met with similar changes in the other variable
- When one variable deviates from its mean $(x_i - \bar{x})$, how does the other one deviates from its mean $(y_i - \bar{y})$?

Correlation

- Covariance is helpful measuring association between two variables but it depends on a scale, which may complicate its use
- To overcome this problem, covariance can be “standardized”
- The result is the correlation:

  - $r_{xy} = \frac{1}{N-1} \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{S_x \, S_y}$

- Measures the extent of the linear relationship between two variables

ARIMA models and correlation

- ARIMA models are based on finding the correlation between today's value and past values y_{t-n} or past forecast errors e_{t-n}

- We measure the association between

  y_t and y_{t-1}, y_t and y_{t-2}, y_t and y_{t-3}

  and / or

  y_t and e_{t-1}, y_t and e_{t-2}, y_t and e_{t-3}
- In that way we determine the amount of lags that "best" explain the data

Stationarity

- What does the I means in the ARIMA?
- Autocorrelations (covariance's, variances, means) have no meaning if they change through time (are non constant over the sample)
- A time series is stationary if the properties of its distribution do not change through time

Differencing time series

- A way to stabilize a series that is non-stationary, is to eliminate the trend by differencing the data
- Differencing: Compute the difference between two observations

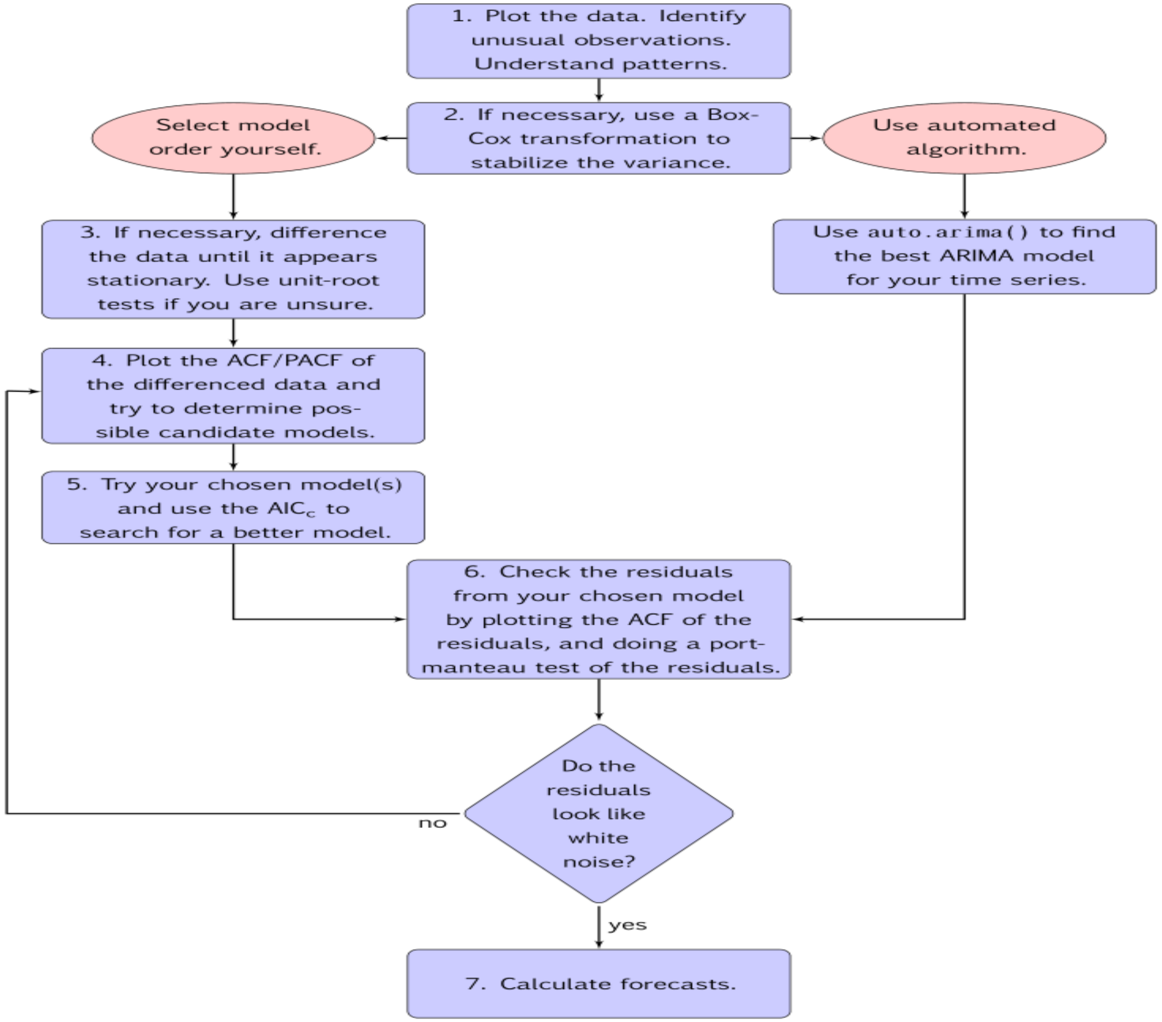
$$\Delta y_t = y_t - y_{t-1}$$

- If data is stationary after differencing it once it is called I(1), integrated of order one. Sometimes data needs to be differenced more than once to become stationary
- That is what the I stands in the ARIMA, the number of times the data needs to be differenced to become stationary

Model selection and check assumptions

- Information criteria
  - Goodness of fit + penalty
- Check model assumptions
  - Test residuals for remaining autocorrelation, constant variance, normality, …

ARIMA algorithm



Extensions to the basic models

- If constant variance does not hold we can check for models of heteroscedasticity
    - ARCH, GARCH, Stochastic volatility

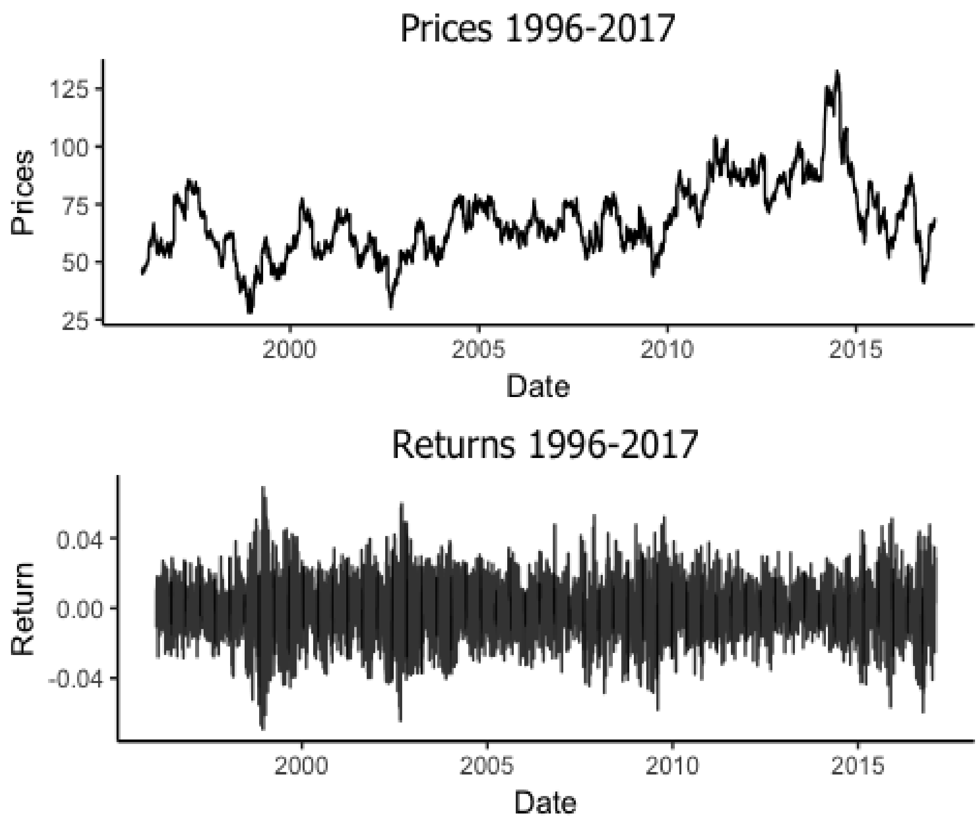
- If data exhibit structural breaks (differences in levels, in trends, or in seasonality between periods) we may use models like
    - Regime switch
    - Markov models

Forecasting Volatility

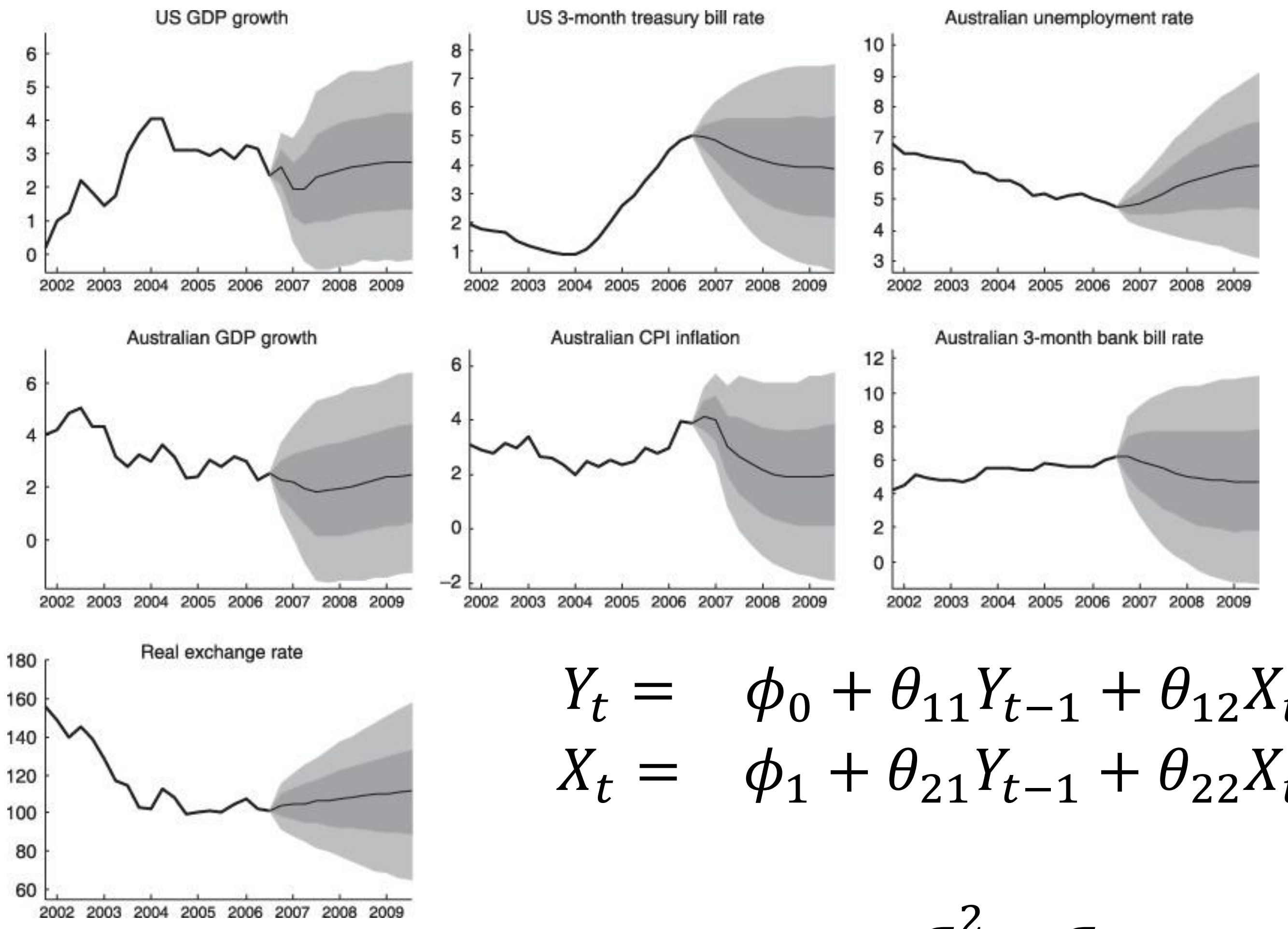
- GARCH (1,1)
- Generalized Autoregressive Conditional Heteroscedasticity

$$\sigma_{t+1}^2 = \omega + \alpha\varepsilon_t^2 + \beta\sigma_t^2$$

Where $\alpha + \beta < 1$



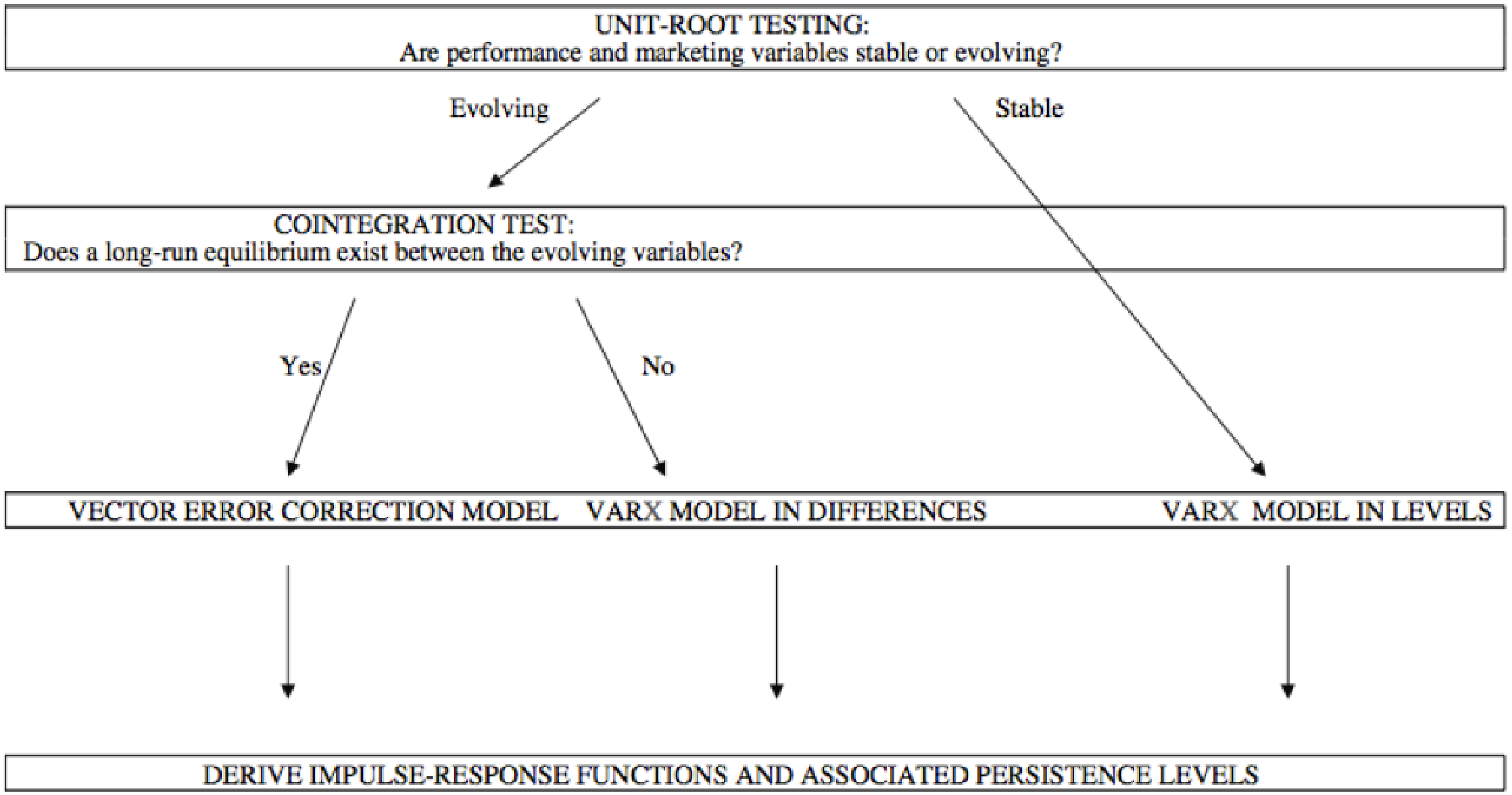
Multivariate models: Vector Autoregression



$$
\begin{aligned}
Y_t &= \quad \phi_0 + \theta_{11} Y_{t-1} + \theta_{12} X_{t-1} + \varepsilon_{1,t} \\
X_t &= \quad \phi_1 + \theta_{21} Y_{t-1} + \theta_{22} X_{t-1} + \varepsilon_{2,t}
\end{aligned}
$$

$$
\Sigma = \begin{matrix} \sigma_y^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_x^2 \end{matrix}
$$

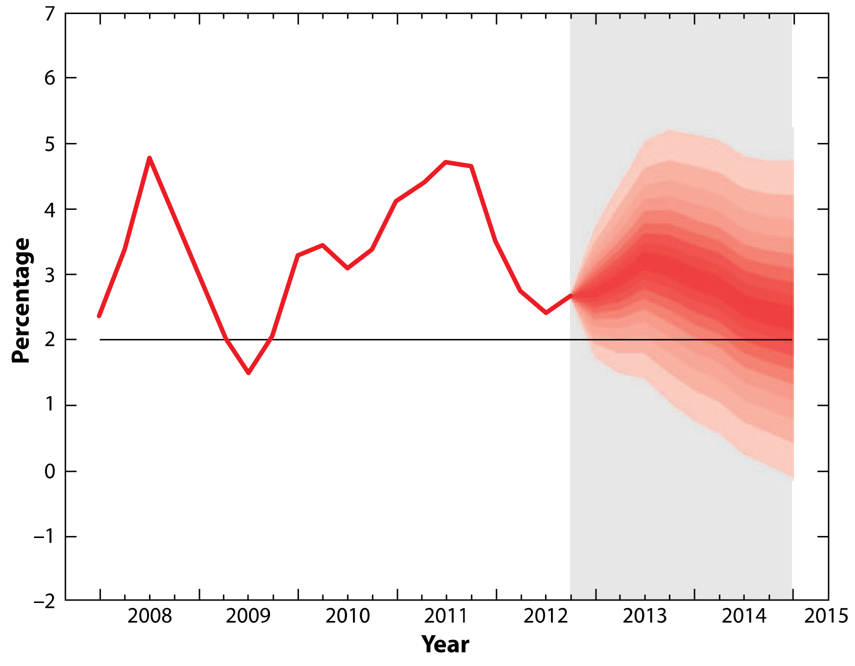
Multivariate models



Extension: Density forecasts

- Basic models look mainly at only two moments of the distribution, but higher moments may also play a role

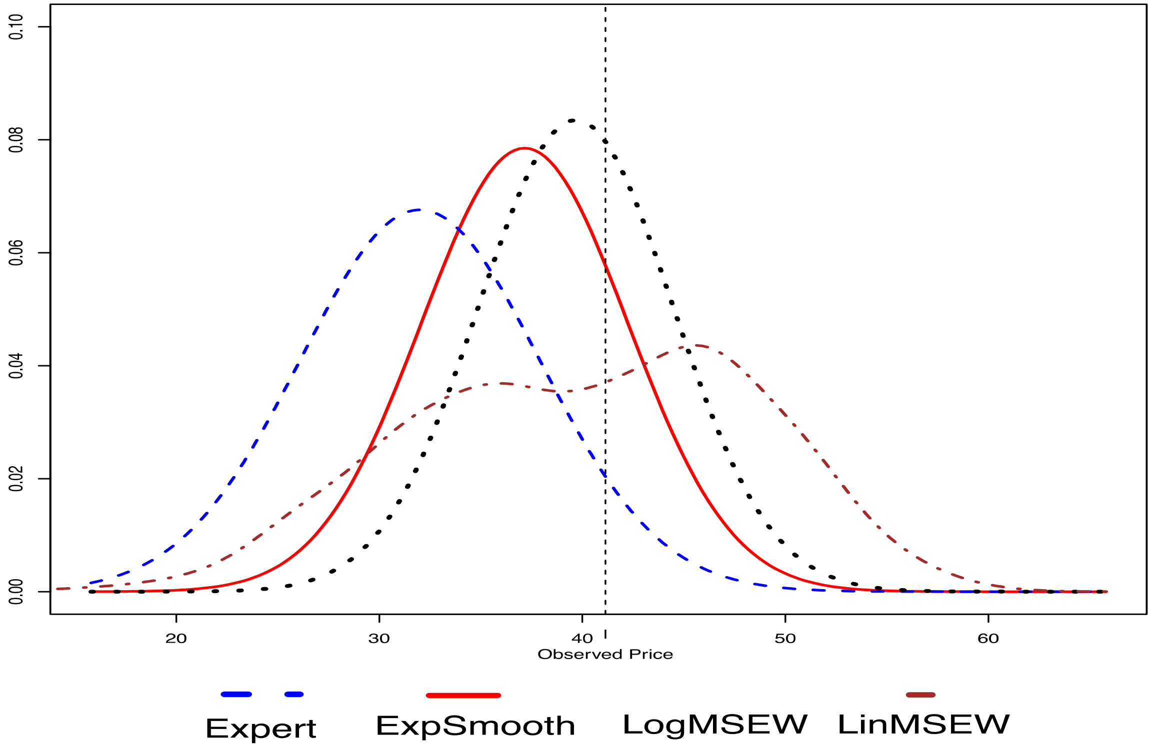
- Density forecast is a prediction in the form of a probability



Bank of England Fan chart

- Important in an environment of high variability

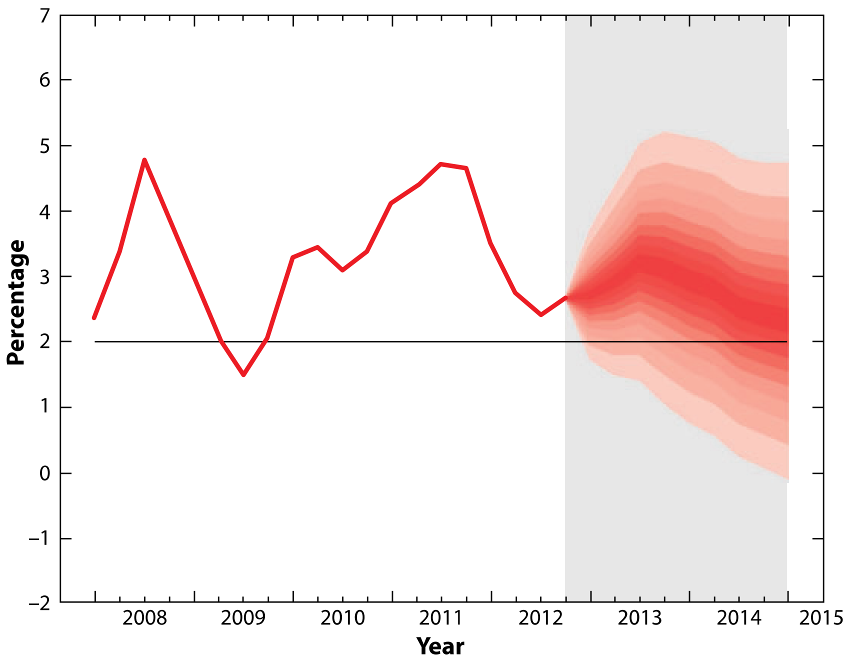
- Density forecasts provide an entire distribution over possible outcomes



Estimation

- Similar to the point-forecast case, density forecast can be generated by:

- Historical data
- Implied forward looking procedures
- Expert based



Bank of England Fan chart

Estimation from historical data

- Not hard to obtain predictive densities by using time series
- Assumption: The system will remain stable over time
- GARCH models and allowing the distributions of errors to be characterized by alternate forms (i.e. Normal, T, GED)
- Autoregressive Conditional Skewness and Kurtosis models also exist (i.e. Bali et al 2008)
- Quantile regression (Lima et al 2014)
- Non parametric approaches (i.e. Nearest-neighbor, kernel smoothing), Bayesian.

Estimation from forward-looking procedures

- Derivative markets are the main tools for risk management and price (and uncertainty) discovery. Therefore convey information about aggregated market expectations

- Predictive densities can be obtained from a set of option prices

Black-Scholes model

- Black-Scholes Price of a Call Option:

$$C = S + N(d_1) - PV(K) \times N(d_2)$$

Where:

S is the current price of the asset

$PV(K)$ is the present value of the strike (exercise) price

$N(d)$ is the cumulative normal distribution

Probability that an outcome from standard normal distribution is below certain value

- $d_1 = \frac{\ln(\frac{S}{PV(K)})}{\sigma\sqrt{T}} + \frac{\sigma\sqrt{T}}{2}$ and $d_1 = d_1 - \sigma\sqrt{T}$

- Where σ is the annual volatility, and T is time to expiration

- Only five inputs are needed to price the option
  - Asset Price, Strike Price, Exercise date, Risk-free rate
  - Volatility of the asset

Implied volatility

- Of the five requires inputs in the Black-Scholes formula, only σ is not observable directly:

- If we observe options prices, and have data of the other arguments of the function then we can obtain implied volatility
- The volatility of an asset's return that is consistent with the quote price of an option on the asset
- Forecast of variance

Estimation:

- Breeden and Litzenberger (1978) showed that a risk-neutral density (RND) can be inferred from European call prices c(X)

$$f(x) = e^{r_f T} \frac{\partial^2 y}{\partial x^2} = e^{r_f T} pdf(S_t)$$

- where x is the strike price, S_t asset's price, r_f risk free rate, and T time to expiration
    - Recall: the pdf is the derivative of the cdf

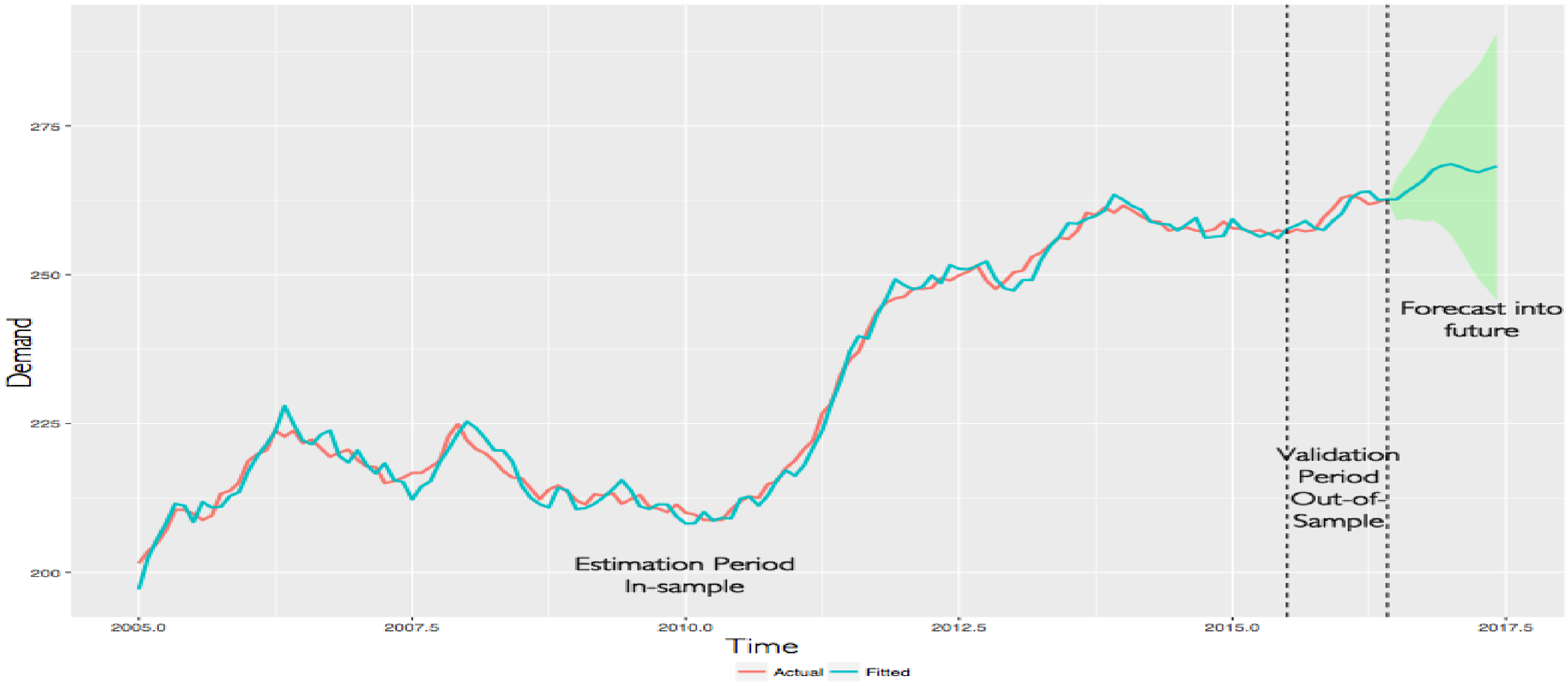
- The task is to find a method that captures RND and provides a reasonable approximation to observed market prices

Estimation:

- Many approaches exist, roughly fall into three categories:

- Fit a parametric density function:
  - Expansion methods: add corrections
  - Generalized distributions: use distributions with higher moments
  - Mixture models: create new distributions from combination
- Non Parametric Approximation:
  - Kernel methods (Ait-Sahalia and Lo, 1998)
  - Curve fitting: Interpolation of volatility smile (Shimko 1993)
- Model of return process: Implied trees, Black-Scholes

Training (Validation) period



Real-World Densities

- RND do not account for risk (risk-free)
- If investors are risk averse then RND would not provide correct distributions (deVincent-Humphreys and Noss 2012)
- We would need to calculate a distribution that reflects the dynamics of real prices. This is called real-world density (RWD) (Taylor 2007)
- From RND to RWD:
  - assume a form of risk aversion (utility function) (Bliss and Panigirtzoglou, 2004).
  - Use statistical methods (recalibration) Fackler and King (1990)

Experts based

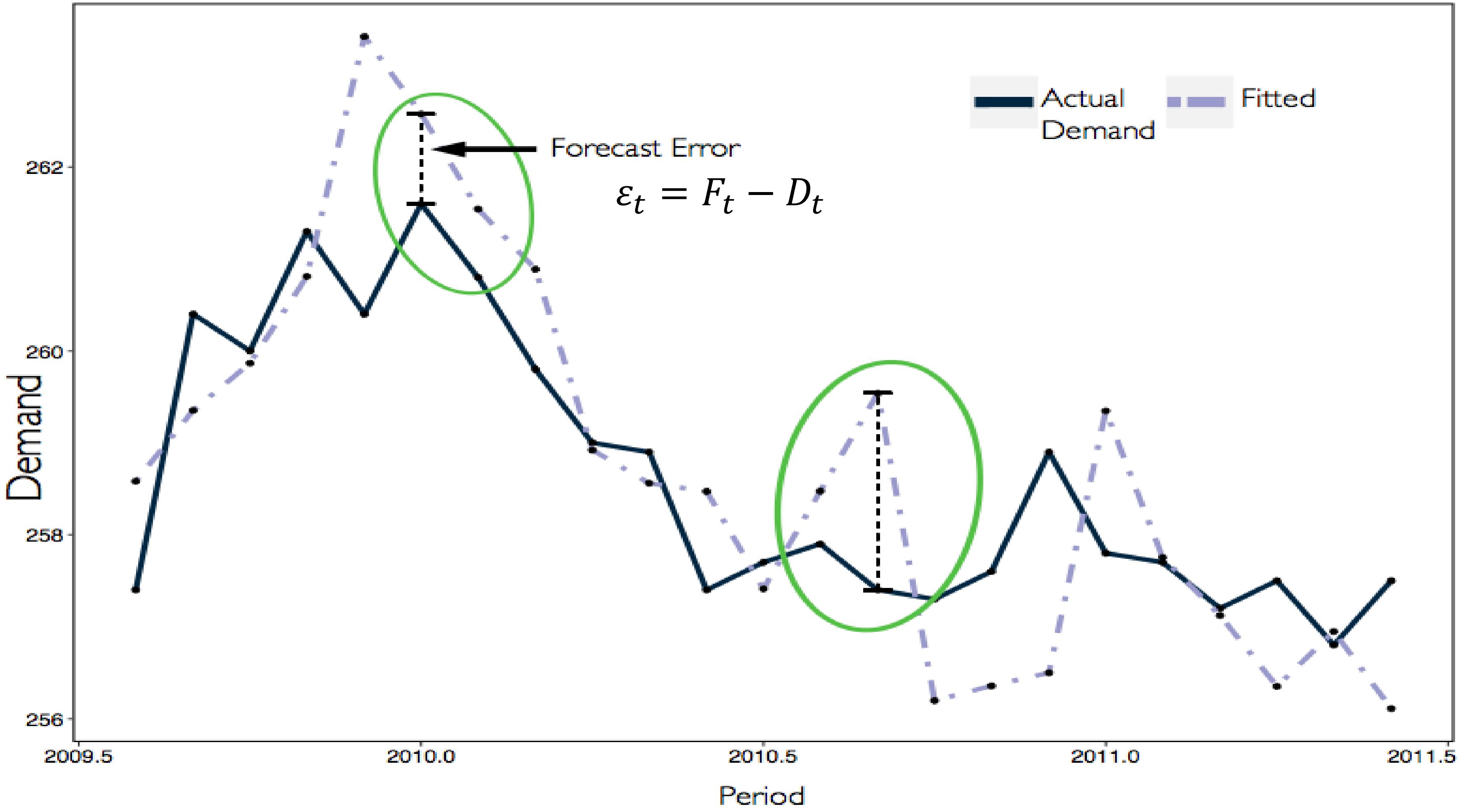
- Much work is needed to understand how this may work, how can it be implemented?
- Isengildina-Massa et al (2011) Empirical confidence intervals for USDA commodity price forecasts
- Density forecast are based on the distribution of historical forecast errors

Evaluation

- How do we evaluate out-of-sample performance of density forecasts (ex-post)?

- Hall and Mitchell (2007) propose:
  - Sharpness (Accuracy):
    - How accurate is the prediction?
  - Calibration (Goodness of fit):
    - Is the distribution correctly specified?
    - Statistical compatibility of probabilistic forecasts and observations; realizations should be indistinguishable from random draws from predictive distributions

Recall: Forecast error



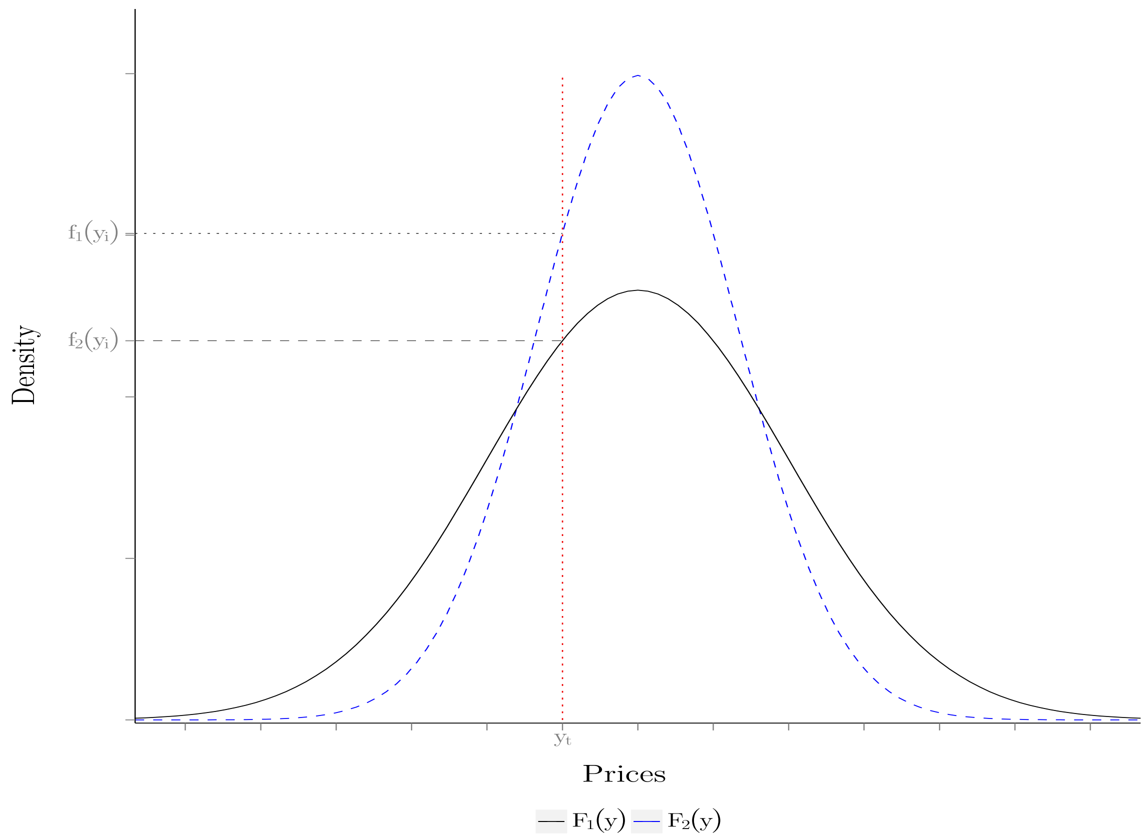
Sharpness (Predictive Accuracy)

- To measure accuracy we use scoring rules:

- Log of the pdf at the realized value

- Average log score:

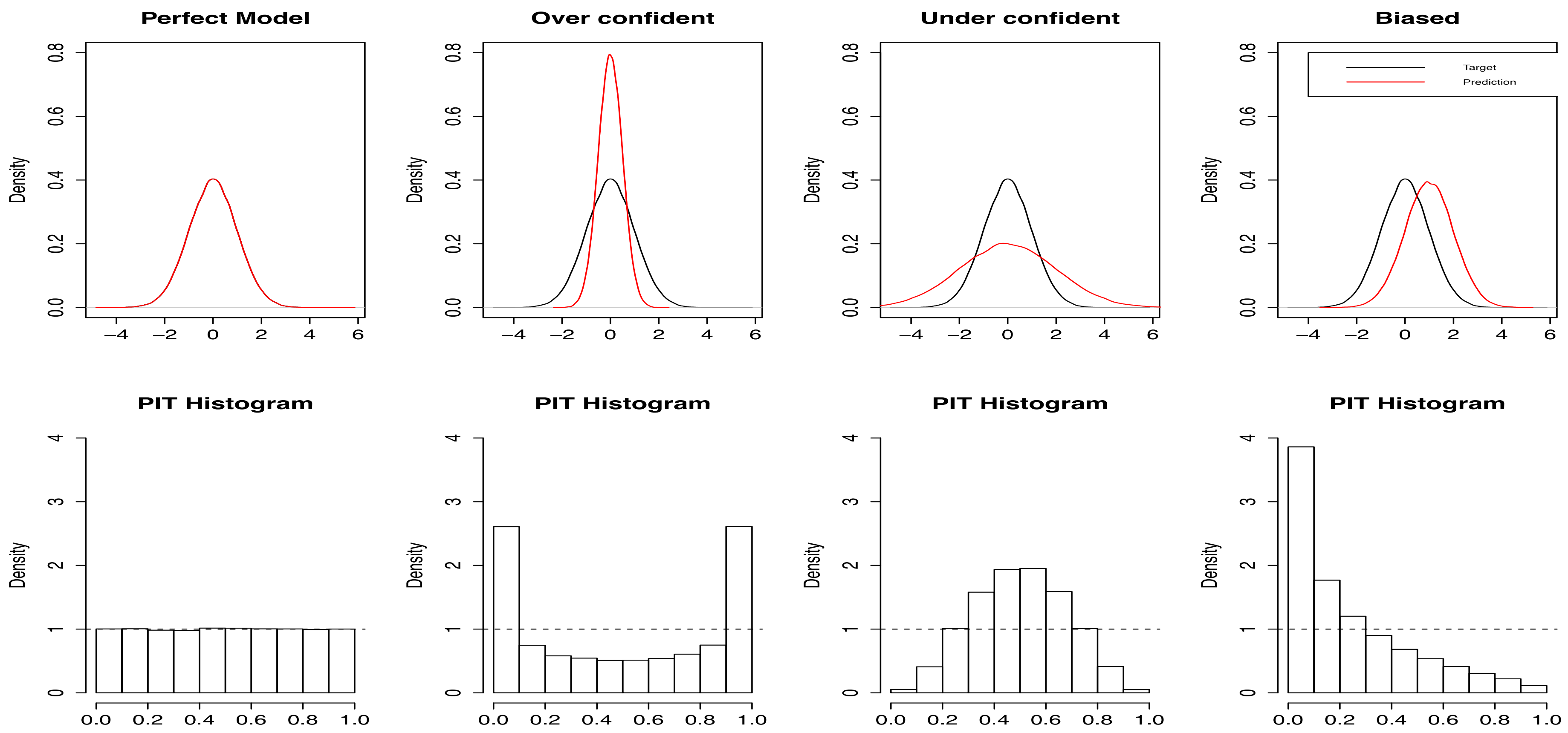
$$\frac{1}{n}\sum_{t=0}^{n-1}\log(f_i(y_t))$$



Calibration

- Density forecast are optimal if the model for the density is correctly specified (Diebold, Gunther, and Tsay 1998)
- Calibration can be measured with Probability Integral Transform (PIT)
- PIT is the CDF of the forecast at the realized observation:

$$PIT = CDF(Y_t)$$



If distribution coincides then PIT are iid U(0,1)

Combination

- The point-forecast literature has found that forecast combinations usually outperform any individual forecasts
    - Decreases the risk of choosing the wrong model (Diversification)
    - Increases the amount of information from different sources
- The question is, would forecast combination also work for density forecasts?

- In Trujillo-Barrera, Garcia, and Mallory (2016) we develop and evaluate quarterly out-of-sample individual and composite density forecast for U.S. hog prices using data from 1975.I to 2014.IV
- Estimation methods:
  - Time series
  - Expert-Based (USDA, Iowa State University)
- How do we combine individual forecasts?

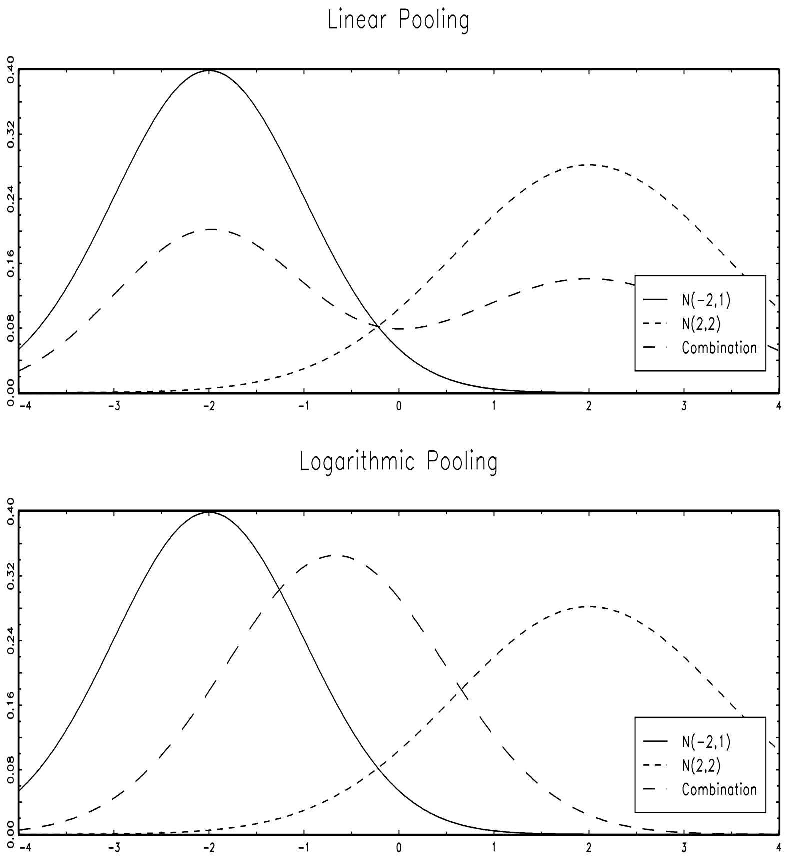
How do we combine individual forecasts?

- To aggregate the densities we use linear and logarithmic combinations

- Linear: $F_c = \sum_{i=1}^{N} \omega_t \, F_{t,i}$

- Logarithm:

$$F_c = \frac{\prod_{i=1}^{N} \omega_{t,i} F_{t,i}}{\int \prod_{i=1}^{N} \omega_{t,i} F_{t,i}}$$



Weights

- How do we obtain the weights of the weighted average of individual models?
- Weighting schemes include:
- Equal weights (EW) (simple average)
- Mean squared error weights (MSEW): based on point forecast metrics
- Recursive log scores weights (RLSW): based on density forecast metrics

Findings

- In the context of U.S. Hog prices we found:
- Equal weighted logarithmic combinations of density forecasts always dominate individual forecast and linear combinations
- Therefore room for improvement over experts' forecasts
- Equal weights outperform more complex weighting schemes (Forecast combination puzzle) No need for complicated weighting?
- Point forecast techniques outperform density forecasts on times of lower uncertainty, while density forecasts dominate in times of higher volatility

Other developments: Machine learning

- Use of big data
- Precision data
- CERN’s domain

- Google: http://www.unofficialgoogledatascience.com
- Facebook: https://research.fb.com/prophet-forecasting-at-scale/

- Algorithms that look for patterns in data:
  - Classification: i.e. Cluster analysis, neural networks, …
  - Dimension Reduction: i.e. Factor analysis
  - Supervised learning: Regressions

Relationship with liquidity

- A lot of financial tools deal with prices
- Prices are hard to understand and predict
  - Market efficiency theory vs behavioral finance

- The liquidity process and its determination is still not well understood, particularly in the high frequency domain

- Time series have been used in the finance literature to model the limit order book, but complications arise from imposing regular time intervals
  - Aggregation
  - Snap-shots

Recent work

- Hasbrouck (2019) proposes the use of VHAR (Vector Heterogenous Autoregressive Model)
- VHAR consist of a subsample of lags in a VAR that aims to capture, long, medium and short run dynamics a process
    - Instead of using all lags up to certain number
- De Boer, Gardebroek, Trujillo-Barrera, and Pennings (2019) use this model to explore liquidity spillovers in the soybean crush process

Mixed frequency models

- We may want to include in a model variables linked to different time frequencies
  - Take for instance in a macroeconomic example the use of GDP indicators (quarterly), inflation (monthly), and prices (daily).

- The “longer” term variable would have missing values that would be treated as latent
  - Solutions:
    - Impose a functional form to the structure of the missing values obtained from the observed values (i.e. Mixed frequency models (MIDAS))
    - Use of State-Space models (this may be a promising venue for our research)

Discussion on how does this models apply to liquidity

- Advantages
- Problems
- Implementations
- Further work