Advanced Forecasting Models

Romain Lafarguette, Ph.D.

Quant & IMF External Expert[1]

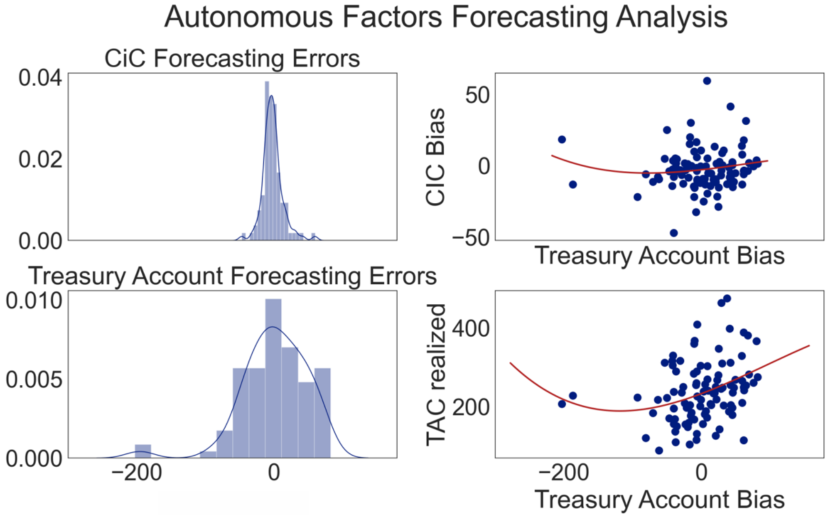
https://romainlafarguette.github.io/

Singapore Training Institute, 10 November 2022

[1] *This training material is the property of the International Monetary Fund (IMF) and is intended for use in IMF courses. Any reuse requires the permission of the IMF.*

Table of Contents

Bias Identification



Credit: *Author*

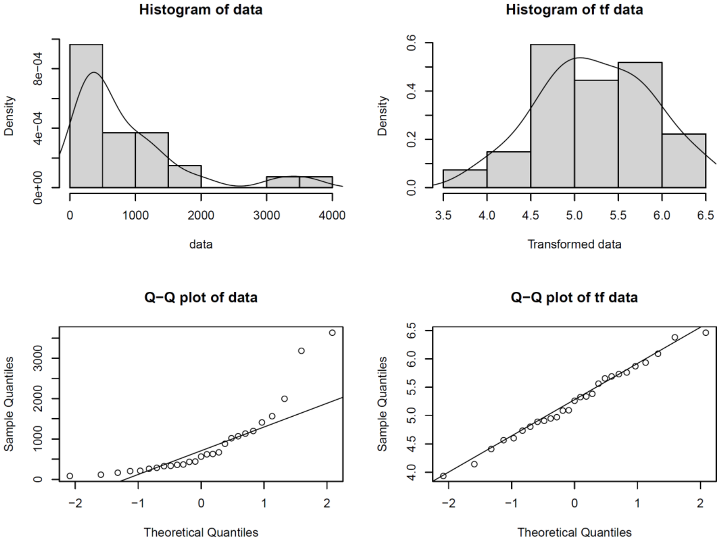
Advanced Models Typology

- **ETS with trigonometric** multiple seasonalities and indicators for events
- **ARIMA with trigonometric** multiple seasonalities and indicators for events
- **ARIMA with stochastic** multiple seasonalities and indicators for events
- **TBATS** – a state space model with exponential smoothing, stochastic trigonometric patterns, Cox-Box transformation and ARIMA errors. TBATS cannot incorporate events
- **Composite TBATS** – where the local level and events are modelled with ETS and the seasonality/trend with TBATS

Box Cox Transformation

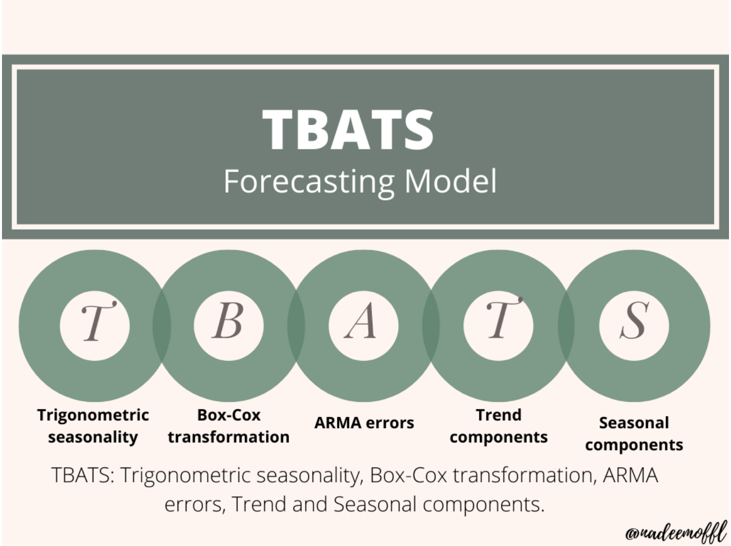
- The Box-Cox transformation is used to transform non-Gaussian data y_t into Gaussian data
- Working with Gaussian data makes it easier to apply the standard model in econometrics
- The Box Cox transformation works with a parameters λ that takes values between $[-5, 5]$
- For positive y_t data:
  - $y(\lambda) = \frac{y^{\lambda}-1}{\lambda}$ if $\lambda neq 0$
  - $\log(y)$ if $\lambda = 0$
- For negative y_t data, use an extra parameter λ_2
  - $y(\lambda) = \frac{(y+\lambda_2)^{\lambda}-1}{\lambda}$ if $\lambda neq 0$
  - $\log(y + \lambda_2)$ if $\lambda = 0$

Transforming Data with the Box Cox Transformation



Credit: *https://universeofdatascience.com/*

The TBATS Model: Overview



Credit: *https://medium.com/analytics-vidhya/*

The TBATS Model: Overview

- Combines in a **single state-space** model:
  - Trigonometric terms to account for seasonality
  - Box-Cox transformation to account for non-normality and heterogeneity
  - ARIMA errors to model the residuals short-term dynamic
  - Trend (potentially damped)
  - Seasonal (including multiple and non-integer period)

- Main specifications:
  - Handles non-integer seasonality, multiple seasonal periods
  - Entirely automated
  - Predictions interval often wide
  - Very slow on long series

The TBATS Model: Specification

Model:

$$y_t^{(\lambda)} = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^{T} s_{t-m_i}^{(i)} + d_t$$

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha d_t$$

$$b_t = \phi b_{t-1} + \beta d_t$$

$$d_t = \sum_{i=1}^{p} \varphi_i d_{t-i} + \sum_{i=1}^{q} \theta_i e_{t-i} + e_t$$

Seasonal part:

$$s_t^{(i)} = \sum_{j=1}^{(k_i)} s_{j,t}^{(i)}$$

$$s_{j,t}^{(i)} = s_{j,t-1}^{(i)} \cos(\omega_i) + s_{j,t-1}^{*(i)} \sin(\omega_i) + \gamma_1^{(i)} d_t$$

$$s_{j,t}^{*(i)} = -s_{j,t-1}^{(i)} \sin(\omega_i) + s_{j,t-1}^{*(i)} \cos(\omega_i) + \gamma_2^{(i)} d_t$$

$$\omega_i = 2\pi j / m_i$$

Where:

- $y_t^{(\lambda)}$ - time series at moment t (Box-Cox transformed)
- $s_t^{(i)}$ - ith seasonal component
- l_t - local level
- b_t - trend with damping
- d_t - ARMA(p,q) process for residuals
- e_t - Gaussian white noise

Model parameters:

- T - Amount of seasonalities
- m_i - Length of ith seasonal period
- k_i - Amount of harmonics for ith seasonal period
- λ - Box-Cox transformation
- α, β - Smoothing
- ϕ - Trend damping
- φ_i, θ_i - ARMA(p, q) coefficients
- $\gamma_1^{(i)}, \gamma_2^{(i)}$ - Seasonal smoothing (two for each period)

The TBATS Model: Decomposition

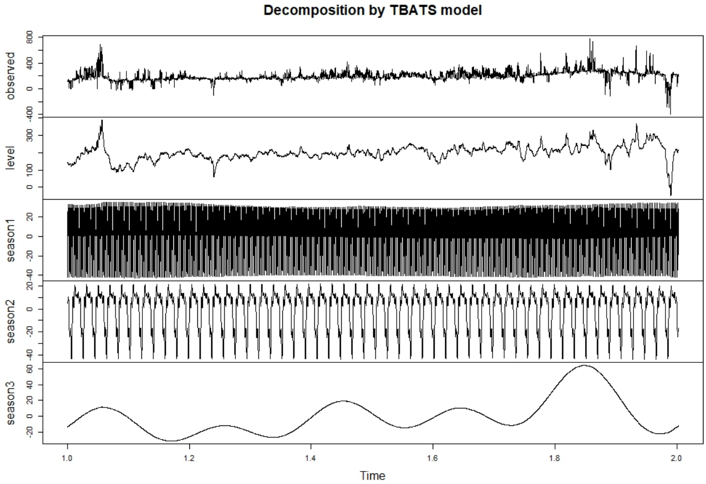
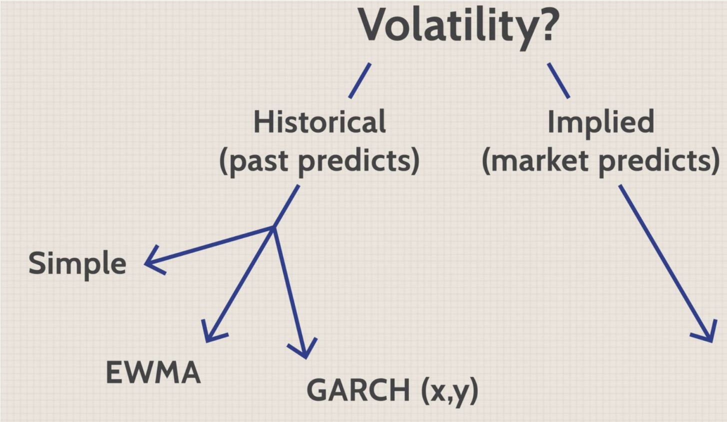


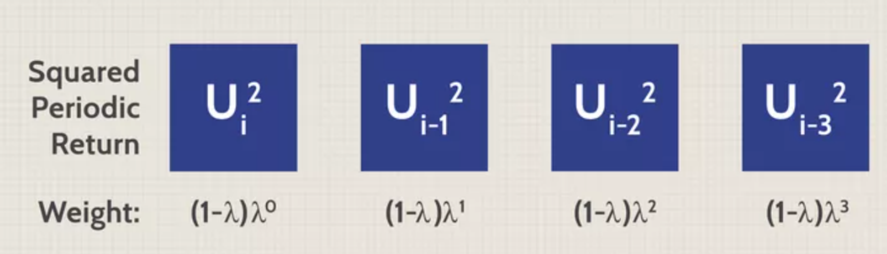
Table of Contents

Volatility Estimation



Credit: *Sabrina Jiang*

EWMA Model



Credit: *Sabrina Jiang*

$$\sigma_n^2(\text{EWMA}) = \lambda\sigma_{n-1}^2 + (1-\lambda)u_{n-1}^2$$

where:

EWMA = Exponentially weighted moving average

σ_n^2 = Variance today

λ = Degree of weighting

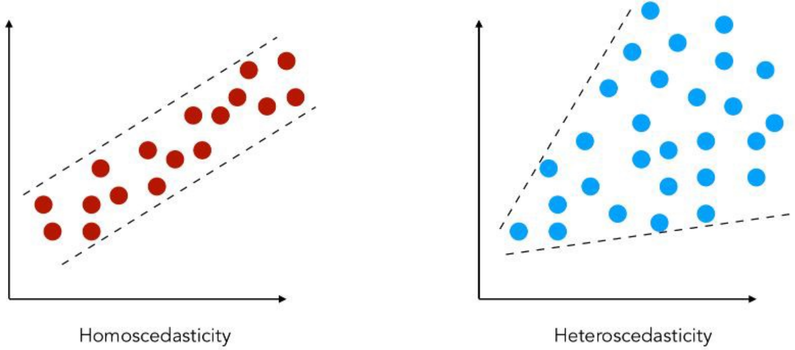
σ_{n-1}^2 = Variance yesterday

u_{n-1}^2 = Squared return yesterday

EWMA Weights

Simple	EWMA	
Equal Weight	**Weight (α)**	**αReturn^2 ($\alpha\mu^2$)**
0.196%	6.00%	0.000010%
0.196%	5.64%	0.000004%
0.196%	5.30%	0.000886%
0.196%	4.98%	0.000034%
0.196%	4.68%	0.001189%
0.196%	0.00%	0.000000%
100%	100%	

Homoscedasticity and Heteroscedasticity



Credit: *Medium.com*

GARCH Models

Meaning

GARCH: **G**eneralized **A**uto**R**egressive **C**onditional **H**eteroscedasticity

Reference: Bollerslev, T. (1986), Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics, 31, 307-327

Definition: GARCH Model

The stochastic process $\{\epsilon_t,\ t\ in \mathbb{Z}\}$ is said to be a **GARCH**(p,q) process if:

$$\epsilon_t = Z_t \sigma_t$$

where Z_t is a sequence of i.i.d variables with $\mathbb{E}(Z_t) = 0$ and $\mathbb{V}(Z_t) = 1$, and σ_t is a non-negative process such that:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2$$

with $\omega > 0$, $\forall\ i, (\alpha_i, \beta_i)\ \in\ \mathbb{R}^{+,2}$ and $\sum_{i=1}^{p} \alpha_i + \sum_{i=1}^{p} \beta_i < 1$

GARCH: Intuition

- The conditional variance of a GARCH(p, q) depends on:
  - The first p lag of the ϵ_t^2 (e.g. the squared error terms)
  - The first q lag of the conditional variance σ^2

$$\sigma_t^2 = \omega + \underbrace{\sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2}_{\text{ARCH Components}} + \underbrace{\sum_{i=1}^{p} \beta_i \sigma_{t-i}^2}_{GARCH components}$$

- The parameters α_i are often called the **ARCH parameters**
- The parameters β_i are often called the **GARCH parameters**

GARCH(1, 1)

Tip

GARCH(1,1) specifications are generally sufficient to capture the dynamics of the conditional variance

Special Case: GARCH(1, 1)

The stochastic process $\{\epsilon_t,\ t\ in\mathbb{Z}\}$ is said to be a **GARCH**(1,2) process if:

$$\epsilon_t = Z_t \sigma_t$$

where Z_t is a sequence of i.i.d variables with $\mathbb{E}(Z_t) = 0$ and $\mathbb{V}(Z_t) = 1$, and σ_t is a non-negative process such that:

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

with $\omega > 0$, $\alpha \geqslant 0, \beta \geqslant 0$ and $\alpha + \beta < 1$

Conditional Variance Persistences: Intrinsic and Extrinsic

- The conditional variance $\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$ depends on two effects:
  - An **intrinsic persistence** effect through the first lag of the conditional variance
  - An **extrinsic persistence** effect
- Following a positive (or negative) shock at time *t-1*, the conditional variance at time t increases (impact effect) and thus it has an impact on $\epsilon_t = Z_t\sigma_t$

$$\text{shock } z_{t-1} \; > \; 0 \Rightarrow \; \epsilon_{t-1} \; \uparrow \Rightarrow \; \sigma_t \; \uparrow \; \ldots$$

- Starting from the next period (at time t), the effect of the shock at $t-1$ on the conditional variance at $t+1$ passes through the conditional variance at time t (intrisic persistence)

$$\text{shock } z_{t-1} \; \Rightarrow \; \sigma_t \; \uparrow \Rightarrow \; \sigma_{t+1}^2 \; \uparrow$$

- The overall effect of a shock can be decomposed into a

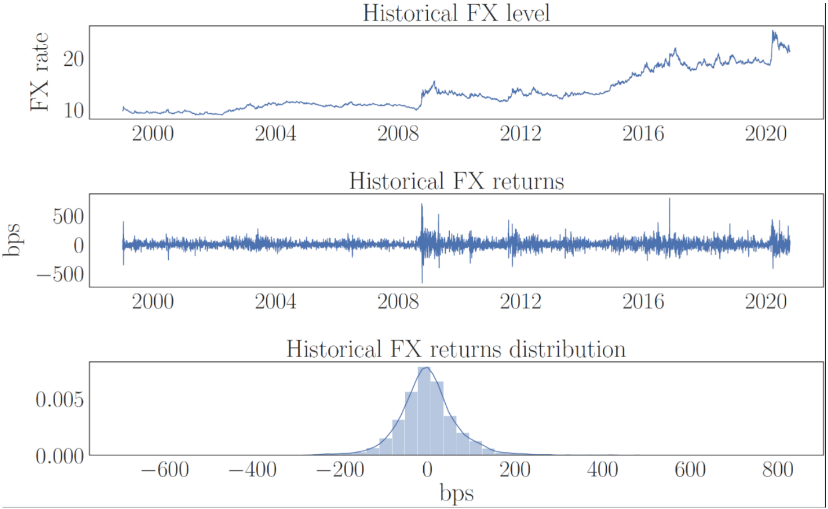
Remarks

- It is often the case that:
  1. The sum of the estimates of α and β are generally close (but below 1)
  2. The estimate of β is generally greater than the one of α
  3. The estimate of β is generally larger than 0.90 for daily returns and the estimate of α is below 0.1
- Be careful: it is not a general rule, just an observation.

GARCH Extensions

- I won't cover them during the course but many extensions have been proposed to extend the GARCH model, including model for dealing with **asymmetric responses**, **persistence** and **long-memory**
  - **IGARCH model**: Integrated GARCH (model persistence)
  - **GARCH-M model**: GARCH mean, when the mean depends on the volatility
  - **GJR-GARCH model**: Glosten, Jagannathan and Runkle (1993): asymmetric with non-linearities
  - **TGARCH model**: Threshold GARCH captures regimes change
  - **EGARCH model**: Exponential GARCH captures asymmetric responses and big shocks

Application: FX Modeling



Credit: *Lafarguette and Veyrune (2021)*

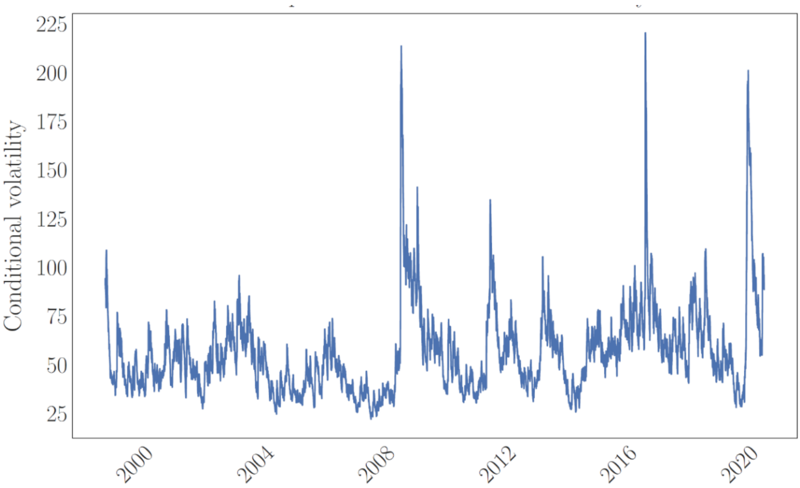
Application: GARCH Estimation

	Microstructure	CIP	Dollar move	Risk appetite	Baseline
Regressors					
Intercept	-2.33***	-2.29	-1.84	-2.55	-1.63
Lag FX log returns	-0.07***	-0.08	-0.08***	-0.08***	-0.08***
Bid-ask abs	5.71	24.39	-35.66	-2.42	3.23
Min-max abs	35.56***	34.63	34.32	34.55*	26.21
Forward points first difference	23.29***	17.79***	26.44***	19.8***	19.44***
Interbank rate vs Libor		33.61***	39.32***	34.75***	33.86***
EUR/USD log returns			-0.14***	-0.17***	-0.16***
VIX first diff				15.66***	15.37***
Oil prices log returns					-0.02***
FX intervention dummy lag					2.23
GARCH Parameters					
Omega	0.13***	0.13	0.12***	0.11***	0.12***
Alpha	0.17***	0.17*	0.16***	0.16***	0.15***
Gamma	0.07***	0.06***	0.06***	0.05***	0.05***
Beta	0.98***	0.99***	0.99***	0.99***	0.99***
Nu	8.33***	8.66***	8.92***	8.71***	8.54***
Lambda	0.08***	0.07	0.09	0.07*	0.08***
R2	0.058	0.067	0.104	0.273	0.276
R2 adjusted	0.058	0.066	0.104	0.272	0.275
Number of observations	5986	5986	5682	5682	5680
Significance thresholds					
**10%, **5%, ***1%*					

Credit: *Lafarguette and Veyrune (2021)*

Application: MXN/USD Conditional Volatility

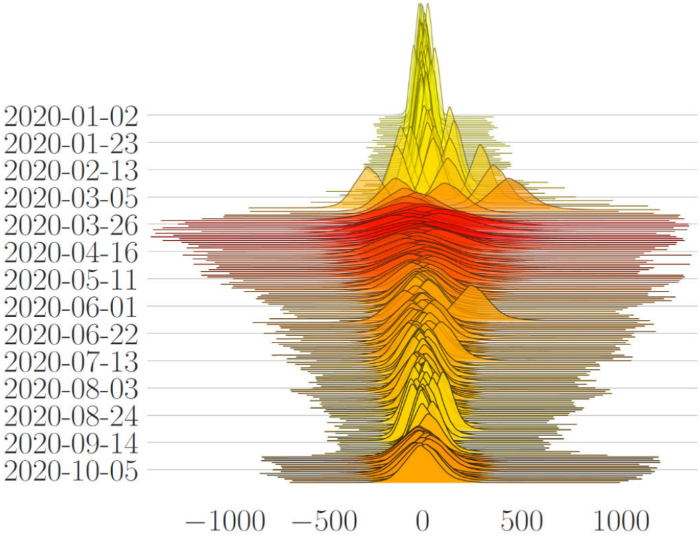
Figure 3. Conditional FX Volatility over Time



Credit: *Lafarguette and Veyrune (2021)*

Application: MXN/USD Conditional Density

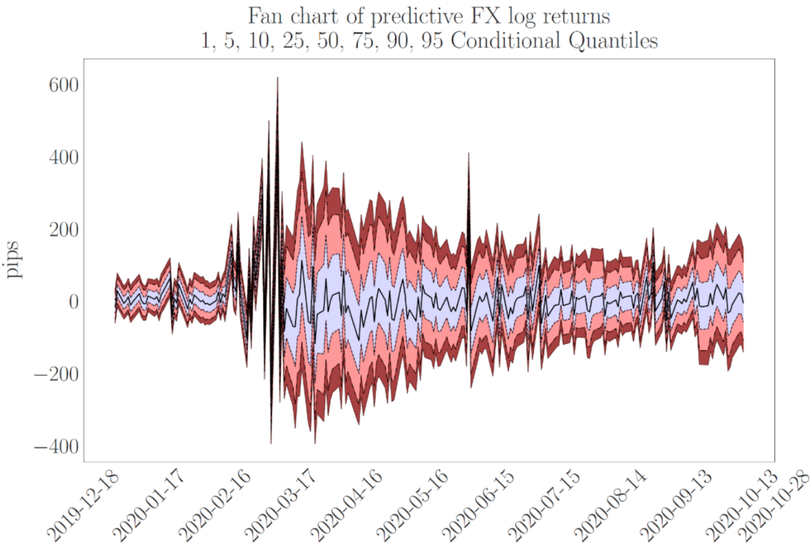
Figure 4. Out-of-Sample Conditional Density



Credit: *Lafarguette and Veyrune (2021)*

Application: MXN/USD Fan Chart

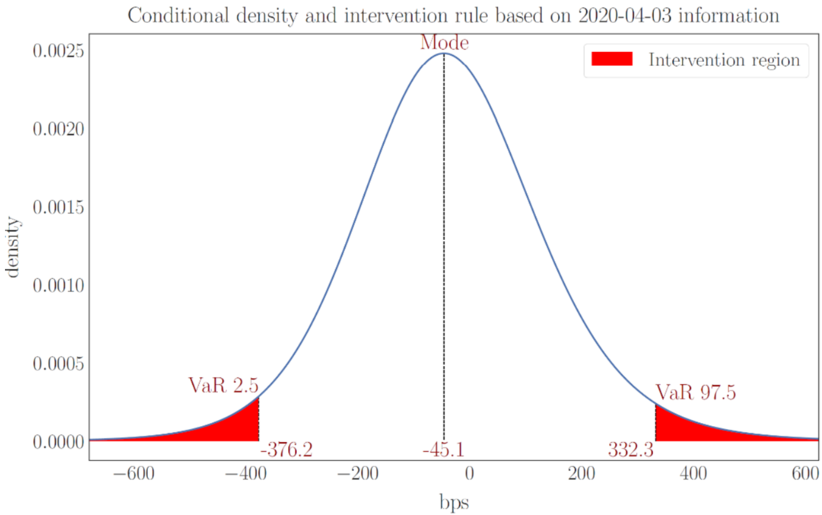
Figure 5. Out-of-Sample Fan Chart



Credit: *Lafarguette and Veyrune (2021)*

Application: MXN/USD Intervention Rule

Figure 7. VaR FX Intervention Rule Based on a Given Information Set



Credit: *Lafarguette and Veyrune (2021)*

Table of Contents

From Point Forecasts to Distributional Forecasts

- A forecast $\hat{y}_{T+h|T}$ is (usually) the mean of the conditional distribution: $Y_{T+h}|Y_1, \dots Y_T$
- Most models produces Gaussian distributed forecast
- Because models assumes Gaussian residuals: remember that the residuals are the stochastic components that are determining the distribution of both the estimators and the forecasts
- The forecast distribution describes the probability to forecast any future value

Distributional Forecasts

Assuming the **residuals are uncorrelated** with variance σ^2, and with an estimate $\hat{\sigma^2}$

Model Typology

- **Mean:** $Y_{T+h|T} = \mathcal{N}(\overline{Y}, (1 + \frac{1}{T})\hat{\sigma^2})$
- **Naive:** $Y_{T+h|T} = \mathcal{N}(Y_T, h\hat{\sigma^2})$
- **Seasonal Naive:** $Y_{T+h|T} = \mathcal{N}(Y_{T+h-m(k+1)}, (k+1)\hat{\sigma^2})$
- **Drift:** $Y_{T+h|T} = \mathcal{N}(Y_T + \frac{h}{T-1}(Y_T - Y_1), h\frac{T+h}{T}\hat{\sigma^2})$

- where k is the integer part of $\frac{h-1}{m}$
- Note that when $h = 1$ and T is large, these all give the same approximate variance $\hat{\sigma^2}$

Prediction Intervals

Definition

A prediction interval gives **a region** within which we expect Y_{T+h} to lie with **a specified probability**

- Assuming that the forecasting errors are normally distributed, then a 95% prediction interval is:

$$Y_{T+h|T} \pm 1.96\hat{\sigma}_h$$

- where $\hat{\sigma}_h$ is the standard deviation of the h-step distribution
- when $h = 1$, $\hat{\sigma}_h$ can be estimated from the residuals

Prediction Interval Fit

```
brick_fc %>% hilo(level = 95)
```

```
## # A tsibble: 80 x 5 [1Q]
## # Key:       .model [4]
##    .model         Quarter       Bricks .mean         `95%`
##    <chr>            <qtr>       <dist> <dbl>        <hilo>
##  1 Seasonal_naive 2005 Q3 N(428, 2336)   428 [333, 523]95
##  2 Seasonal_naive 2005 Q4 N(397, 2336)   397 [302, 492]95
##  3 Seasonal_naive 2006 Q1 N(355, 2336)   355 [260, 450]95
##  4 Seasonal_naive 2006 Q2 N(435, 2336)   435 [340, 530]95
##  5 Seasonal_naive 2006 Q3 N(428, 4672)   428 [294, 562]95
##  6 Seasonal_naive 2006 Q4 N(397, 4672)   397 [263, 531]95
##  7 Seasonal_naive 2007 Q1 N(355, 4672)   355 [221, 489]95
##  8 Seasonal_naive 2007 Q2 N(435, 4672)   435 [301, 569]95
##  9 Seasonal_naive 2007 Q3 N(428, 7008)   428 [264, 592]95
```

Why Prediction Intervals Matter

- **Point forecasts are often useless without a measure of uncertainty** (such as prediction intervals)
- Prediction intervals require a **stochastic model** (with random errors, etc.)
- For most models, prediction intervals get wider as the forecast horizon increases
- The degree of confidence (the probability) impacts the width of the prediction interval
- Usually too narrow due to unaccounted uncertainty: pay attention

Difference between Confidence Interval and Prediction Interval

- A confidence interval informs about where the true parameter of a model can be
  - **Confidence interval quantifies the uncertainty about the model**, or the distance between the model and reality
  - Confidence interval are associated with a wide range of parameters, values, etc.
  - They inform about how the model represents well the reality
  - Wide confidence intervals are associated with a less accurate model, and/or a very volatile model

- The prediction interval predicts in what range a future individual observation will fall
  - **Prediction interval quantifies the uncertainty about the future**, or the distance between today and the future
  - Prediction interval are not about the parameters of the model, but about the dependent variable (y_t)
  - The problem is that prediction intervals tend to neglect the uncertainty about the parameters used to generate the forecasts...

Difference between Confidence Interval and Prediction Interval

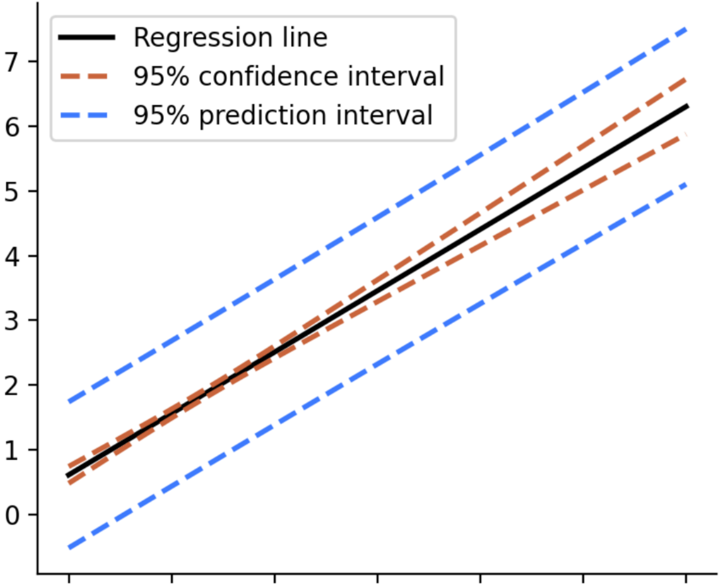


Table of Contents

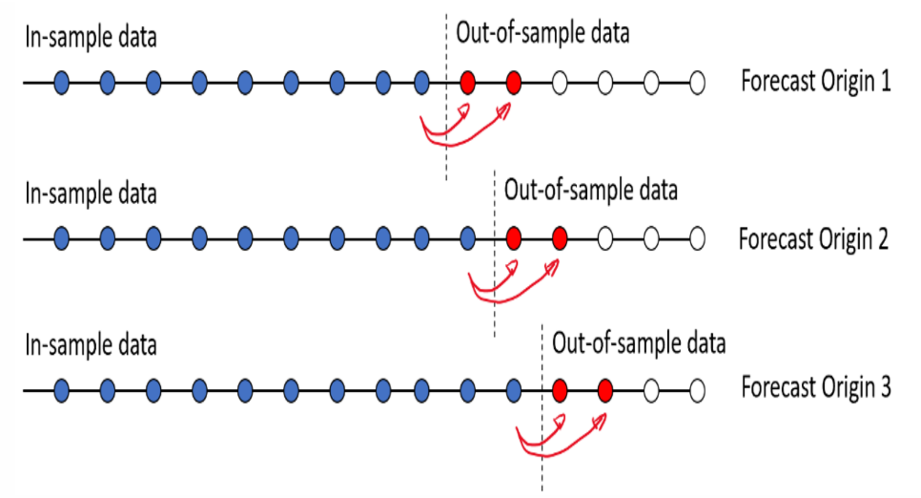
Fitting and Forecasting

Be careful

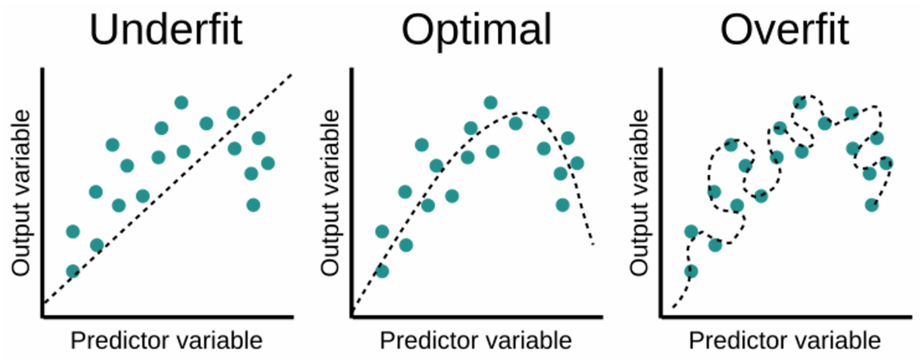
A model that fits the data well (in sample) might not necessarily forecast well

- A perfect in-sample fit can always be obtained by using a model with with enough parameters
- Over-fitting a model to data is just as bad as failing to identify a systematic pattern in the data
- Need to split the model between
- The test set must no be used to *any* aspect of model development or calculation of forecasts
- Forecast accuracy is only based on the test set

Train and Test Set



Underfit, Optimal, Overfit



Forecast Errors

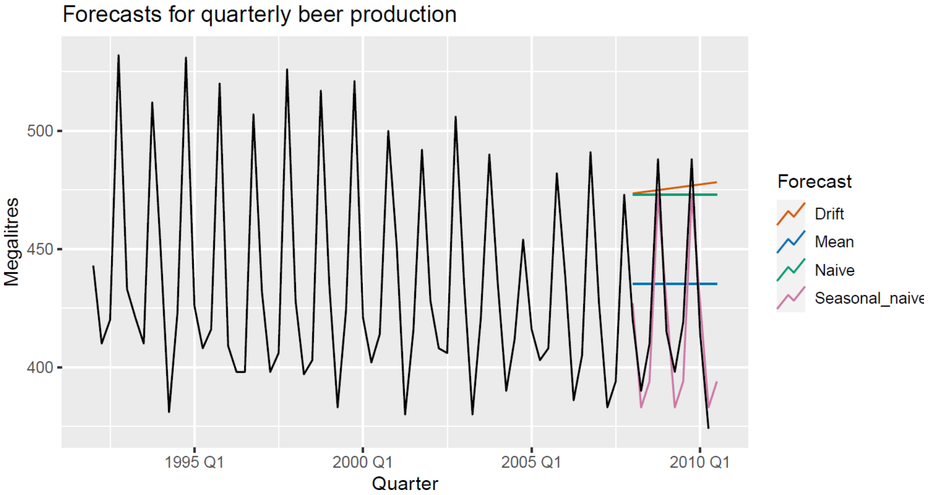
Definition: Forecast Errors

A forecast error is the difference between an observed value and its forecast

$$e_{T+h} = y_{T+h} - \hat{y}_{T+h}|Y_T, \ldots, Y_1$$

- The conditional set $Y_T, \ldots, Y_1$ should only be taken from the training dataset
- The true value y_{T+h} is taken from the test set
- Unlike residuals, forecast errors on the test involve multi-step forecasts
- These are the **true** forecast error, as the test data is not used to compute $\hat{y}_{T+h}$

Example: Forecasting Beer Production



Measures of Forecast Accuracy

Main Metrics

- **MAE**: mean absolute errors $\frac{1}{S}\sum_{s\in S}|e_{s,T+h}|$
- **MSE**: mean squared errors $\frac{1}{S}\sum_{s\in S}(e_{s,T+h})^2$
- **MAPE**: mean absolute percentage errors $\frac{1}{S}100 * \sum_{s\in S}\frac{|e_{s,T+h}|}{|y_{s,t+h}|}$
- **RMSE**: root mean squared errors: $\sqrt{\frac{1}{S}\sum_{s\in S}(e_{s,T+h})^2}$

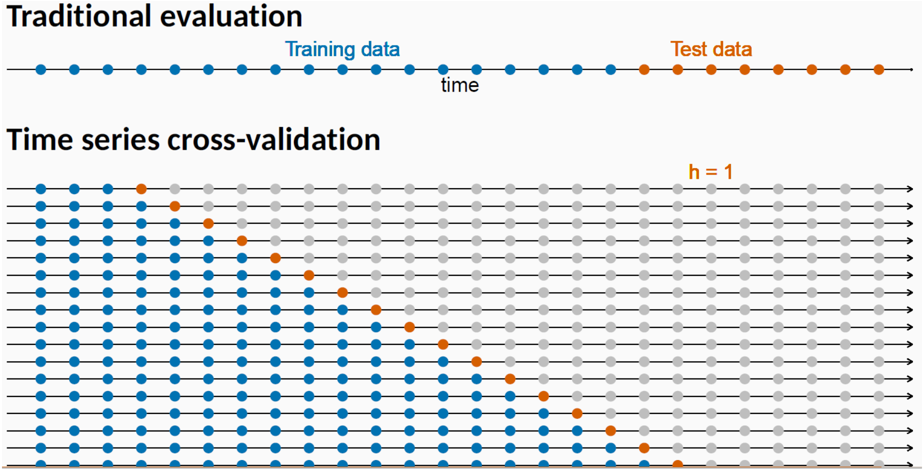
With:

- y_{T+h}: T+h observation, h being the horizon (h = 1, 2, ..., H)
- $\hat{y}_{T+h|T}$: the forecast based on data up to time T
- $e_{T+h} = y_{T+h} - \hat{y}_{T+h|T}$: The forecast errors
- S is the testing sample

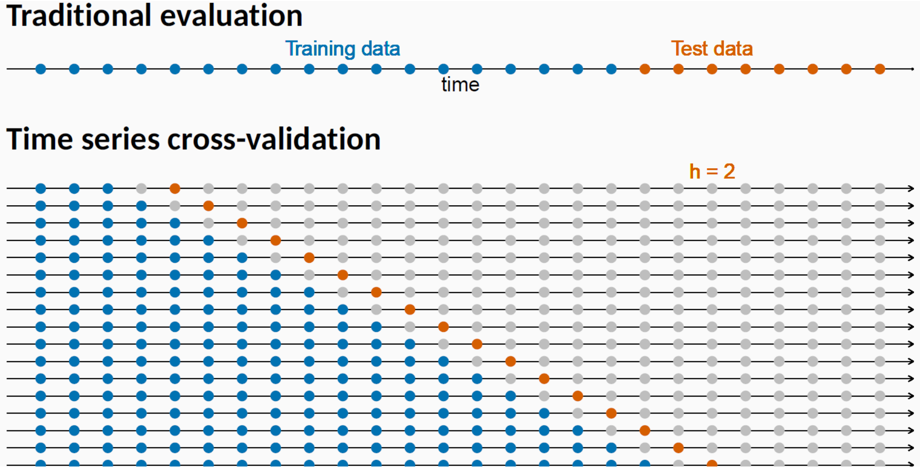
Scaling

- MAE, MSE and RMSE are all **scale dependent**
- MAPE is scale independent but is only sensible if $y_t >> 0 \qquad \forall\ t$
- **Most commonly used: Time Cross-Validation with the lowest RMSE**

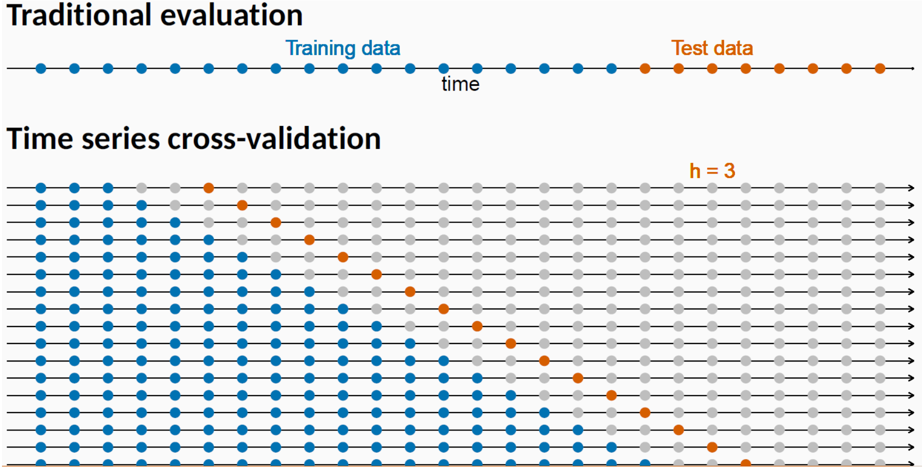
Time Series Cross-Validation



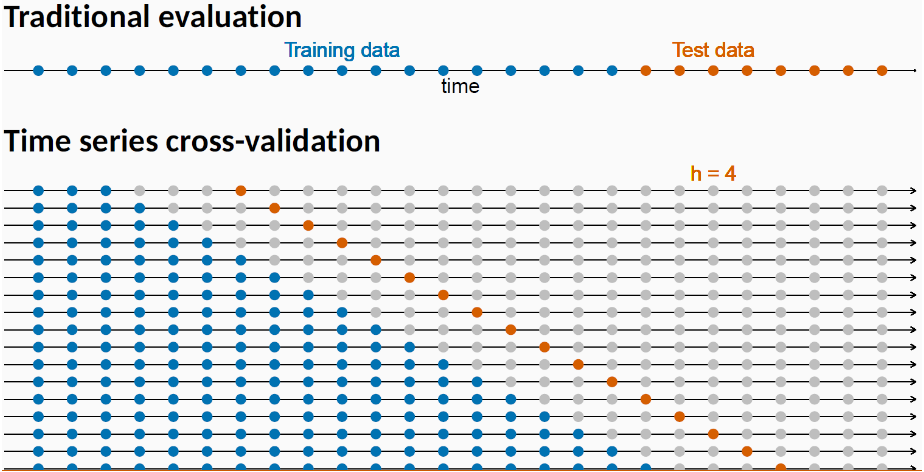
Time Series Cross-Validation



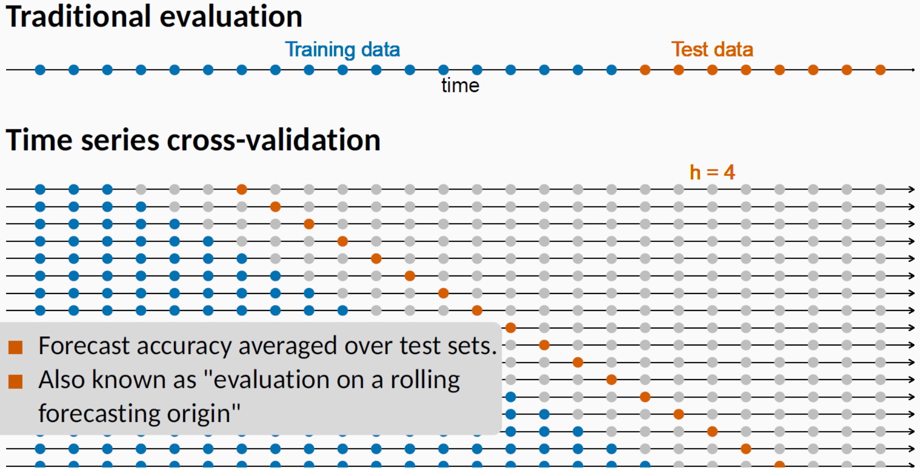
Time Series Cross-Validation



Time Series Cross-Validation



Time Series Cross-Validation



- Forecast accuracy averaged over test sets.
- Also known as "evaluation on a rolling forecasting origin"

Framework RMSE Output

RMSE

The RMSE is defined as

$$RMSE = \sqrt{\sum_{t \in T_{eval}} (y_t - \hat{y}_t(h))^2}$$

where y_t is the observed value and $\hat{y}_t(h)$ is the h-step ahead forecast of y_t. The RMSE for each model and HORIZON is shown below. Lower values indicate better forecasts, the best individual model is in bold, while combinations are colored.

Show 10 entries

Search:

	Method	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9	h=10	h=11	h=12	h=13	h=14	mean
1	Dynamic Model Selection	211.6	457.5	1,097.4	1,599.8	2,048.9	2,315.7	2,400.4	2,501.1	2,622.4	3,204.7	3,451.4	3,672.2	3,573.2	3,604.6	2,340.1
2	Model Average (best 2)	174.8	374.8	1,308.2	1,661.9	2,028.1	2,405.4	2,554.2	2,719.8	2,899.5	3,241.8	3,392.5	3,525.4	3,582.3	3,653.2	2,394.4
3	**ARIMA**	**210.0**	**491.7**	**1,435.5**	**1,785.8**	**2,146.9**	**2,468.8**	**2,608.2**	**2,775.3**	**2,965.3**	**3,245.6**	**3,371.7**	**3,472.3**	**3,484.7**	**3,549.0**	**2,429.4**
4	ETS	179.7	355.0	1,250.8	1,624.1	2,000.8	2,405.1	2,558.2	2,726.0	2,904.0	3,312.7	3,493.8	3,661.4	3,744.5	3,815.3	2,430.8
5	ETS (non-seasonal)	206.8	424.1	1,389.8	1,781.4	2,092.3	2,414.7	2,548.7	2,709.6	2,893.1	3,337.7	3,510.2	3,649.8	3,712.3	3,783.3	2,461.0
6	Model Average (all models)	175.3	371.9	1,342.3	1,721.6	2,085.8	2,471.4	2,622.7	2,795.4	2,985.0	3,401.1	3,576.6	3,729.1	3,819.1	3,896.2	2,499.5
7	Seasonal ARIMA	161.6	351.6	1,428.1	1,863.5	2,289.5	2,742.0	2,911.0	3,106.6	3,320.9	3,883.0	4,120.1	4,331.3	4,503.8	4,594.2	2,829.1

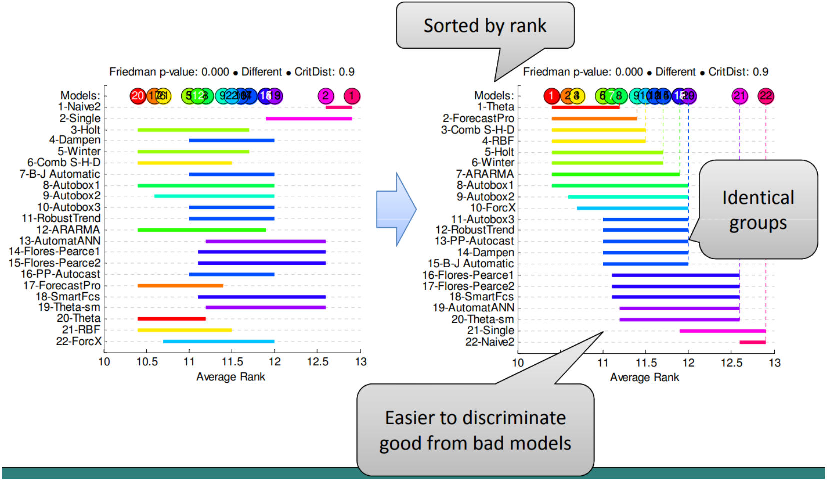
Credit: *IMF Framework*

Nemenyi Test

- We can rank the model by RMSE (or another metric), but are the RMSE significantly different?
- Maybe Model 1 can have a lower RMSE than Model 2, but the difference in RMSE is non-significant
- In which case, we could pool the two models together
- Use a non-parametric test to test the hypothesis of equal RMSE, with the test statistic:

$$r_{\alpha,K,N} \approx \frac{q_{\alpha,K}}{\sqrt{2}}\sqrt{\frac{K(K+1)}{6N}}$$

Nemenyi Test in Practice



Credit: *Nikolaos Kourentzes*

Table of Contents

Combination of Models

- Combining model provides a more reliable forecast, overcoming the risk of relying on a single model

- Estimate all the models, determine the cross-validated errors

- Define a pool of "good" models and reject the other models, based on their out of sample performance. Define a **rejection threshold** to determine the "bad" models, based on the distribution of the RMSE

$$\text{Threshold} = Q(0.75) + 1.5 * \text{IQR}$$

  - Where $Q(0.75)$ is the 75th quantile of the RMSE distribution and IQR the interquantile range $(Q(0.75) - Q(0.25))$

- Then, weight each forecast of the eligible model by the relative performance of their model, the best models having the highest weight

Dynamic Model Selection: Principles

- Some models are better for modelling long-term dynamics, while others are better are short-term dynamics
- Time-varying volatility matters. As volatility increases, models focusing on the short term have an advantage
- Dynamically account for changes in the behaviour of the time series:
    - Combining high performing forecasts
    - Selecting a well performing forecast based on local performance

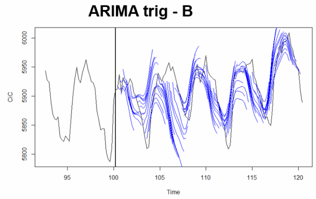
Dynamic Model Selection: Implementation

- **Combining high-performance forecasts:**
  - Compute unweighted average of forecasts. Although it seems better to estimate "optimal" weights, weights estimation introduces additional uncertainty

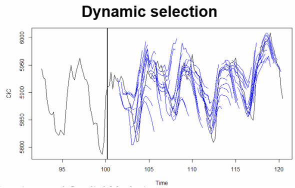
- Selecting a well performing forecast based on **recent performance**
  - **Rolling performance of the previous week** to select the forecast for the next origin.

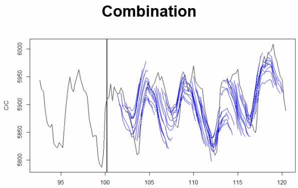
- Consider ETStrig, ARIMAtrig, ARIMA stoch, and TBATS comp. All these models capture information differently
  - **Model diversity is advantageous** both for selection and combination between forecasts

Out-of-Sample Performances via Rolling Origins



- Both selection and combination have fewer extreme forecasts.
  - Combination provides smoother forecasts (combining models acts as a shrinkage estimator)
  - while the selection is able to pick up details





Credit: *Author*

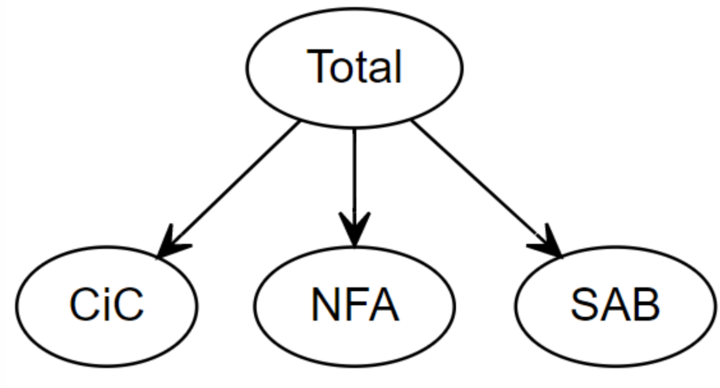
Table of Contents

Forecasts Reconciliation of the Autonomous Factors

Forecast the autonomous factors using a hierarchical modeling approach, with different models for each autonomous factor:

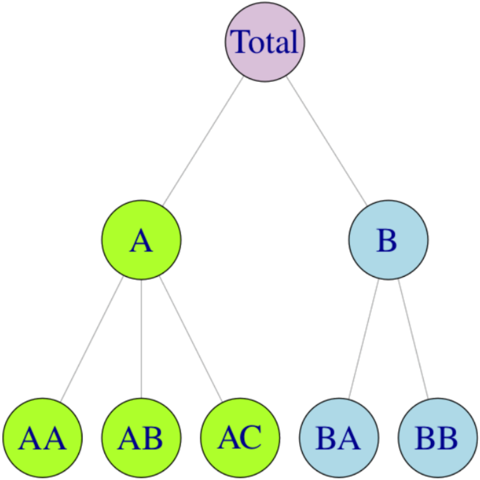
- Forecasting the sum of the autonomous factors separately
- Reconciling the forecasts using aggregation technics as described for instance in Hyndman, Rob J., et al. "Optimal combination forecasts for hierarchical time series." (2011)
- Use either OLS or the “minT” approach: minimize the trace of the VarCov forecast matrix. Exploit the covariance among regressors to maximize the accuracy of the hierarchical forecasts

Reconciliation via Hierarchical Forecasts



Credit: *Author*

General Approach



Credit: *https://otexts.com/fpp3/reconciliation.html*

Formalization

- We can express the total (top level, y_t) via a summing matrix S of the low level elements y_t^L

$$y_t = S y_t^L$$

- Then, we can express the whole system via a grouping matrix summing all the individual elements (also called base forecast y_t^b), so have a consistent approach irrespective of the level structure

$$\tilde{y}_t = S * G * y_h^b$$

$$\begin{bmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \end{bmatrix} \qquad \boldsymbol{G} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Credit: *otexts.com/fpp3/reconciliation.html*

The Min Trace Approach

Forecasts Errors in a Hierarchical System

$$V_h = \mathbb{V}[y_{T+h} - \tilde{y_h}] = SGW_hG'S'$$

where $W_h = \mathbb{V}[y_{T+h} - \tilde{y_h^b}]$ the variance-covariance matrix of the base vector

- The objective is to find a matrix G to minimize the variance-covariance matrix of the corresponding base forecast errors
- Because the error variances are on the diagonal of V_h, minimizing the forecasting errors is equivalent to minimizing the trace of V_h
- Wickramasuriya et al. (2019) show that the minimized trace matrix is given by:

$$G = (S'W_h^{-1}S)^{-1}S'W_h^{-1}$$

- This approach is called the **MinT (or Minimum Trace) optimal reconciliation approach**