Challenges in Forecasting High-Dimensional Time Series

Julie Novak

Shonan Meeting

02.12.2018

Outline

- Challenges in high-dimensional forecasting.
  - Example from IBM revenue forecasting.
- Overview of the forecast reconciliation problem.
- Future directions and open questions.

Motivation: IBM Revenue Forecasting

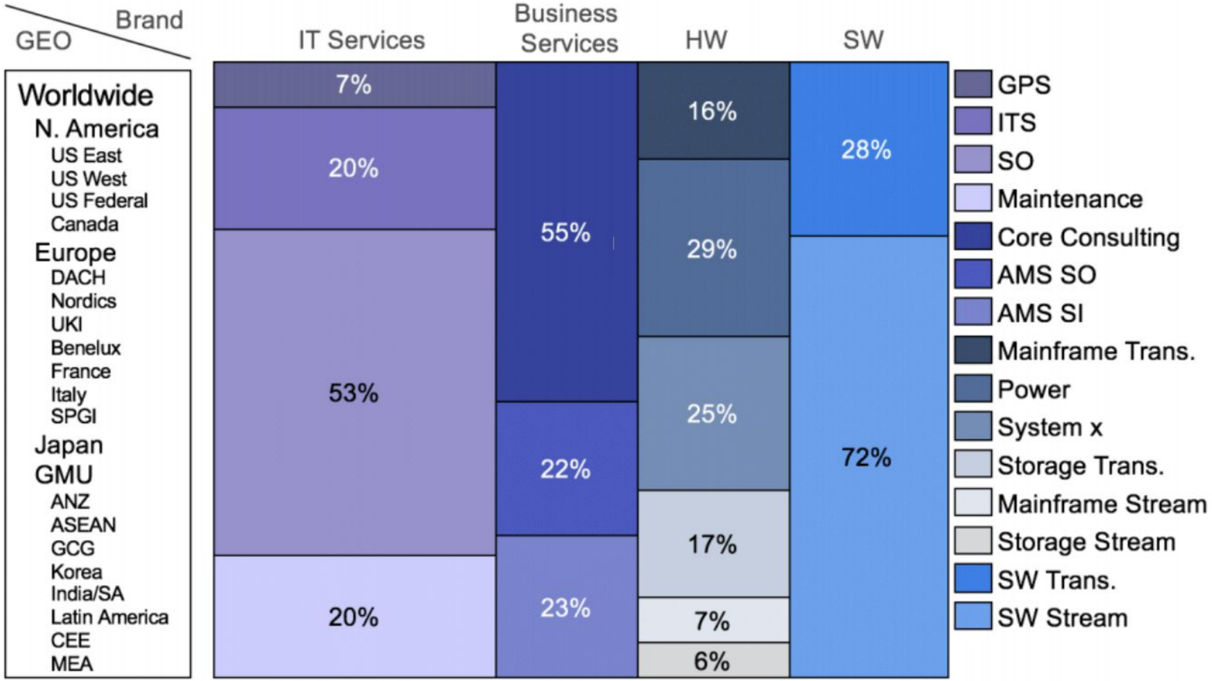
- Automated predictive analytics solution for IBM.
    - Revenue assessments for current and next quarter by market, division, and revenue type.
- Provide assessments for all levels of the hierarchy.
- Why? To aid both high level executives and lower level managers with decision making.
    - Objective system of revenue assessment.
    - Aggregate consistent.

Criteria for a "Successful" Solution

All these criteria are equally important for the solution.

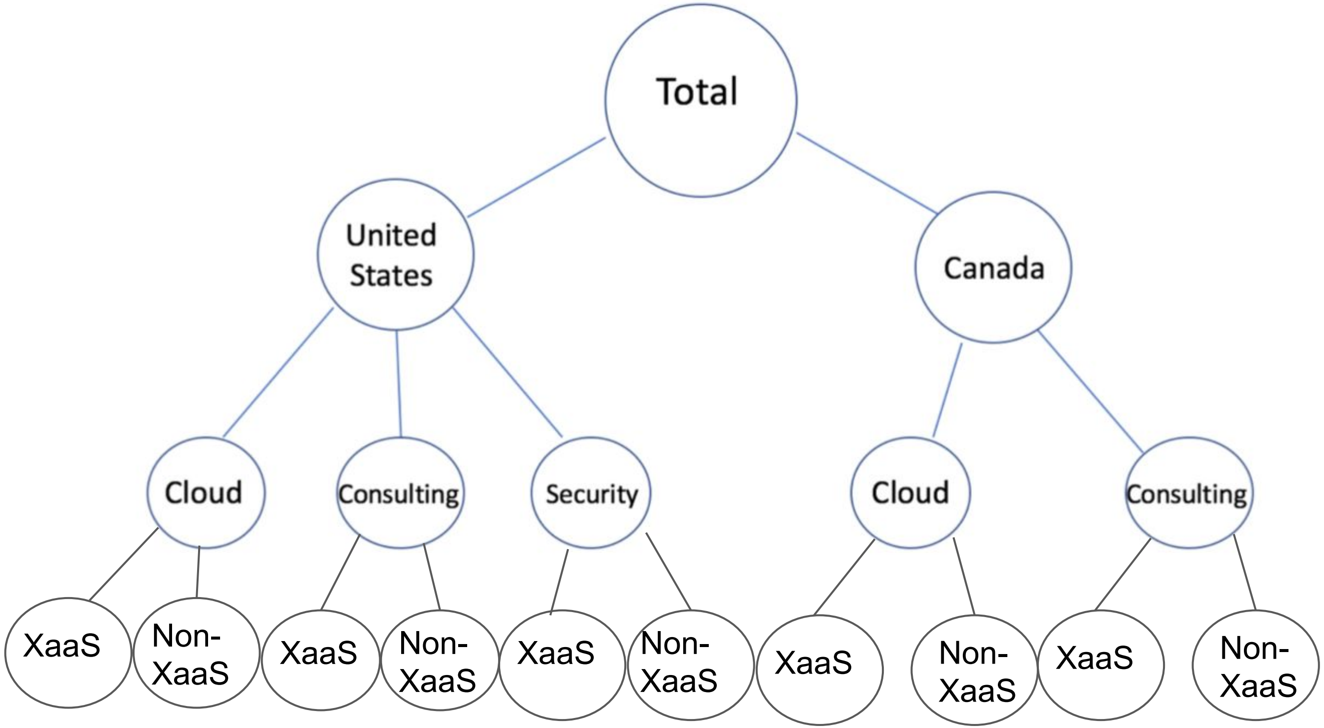
- Accuracy at all levels of the hierarchy.
- Robustness to outliers.
- Smoothness over weeks (for aiding data-driven decision making).
- Interpretability of week to week changes.

Hierarchical Structure of IBM



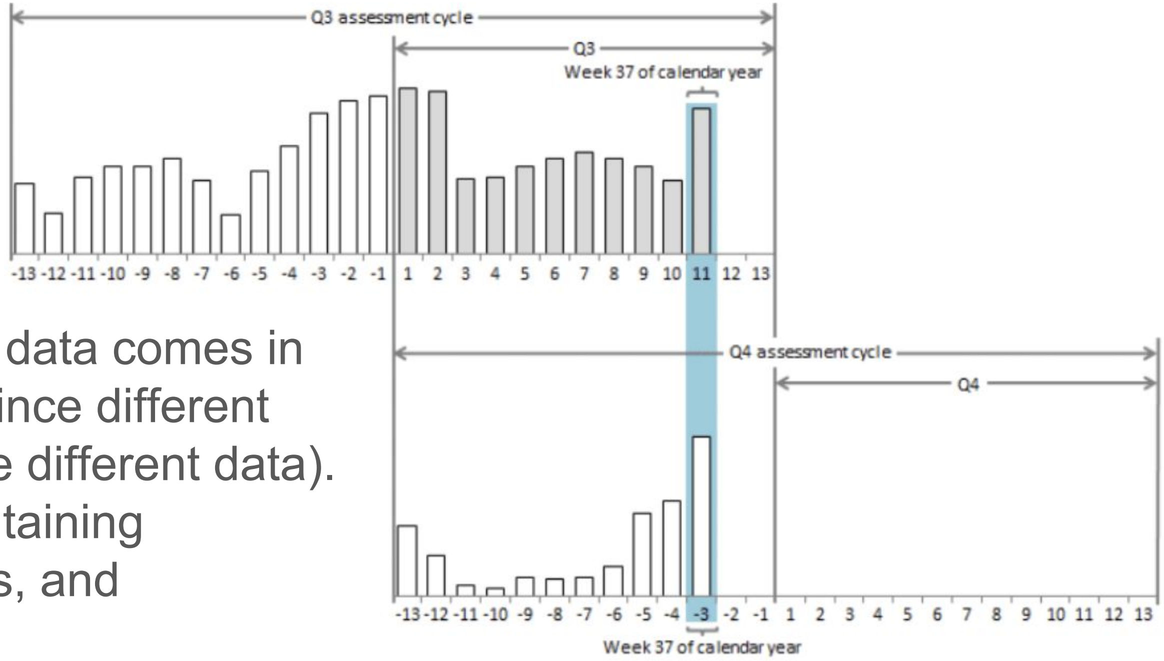
***Note this is an old org structure from years ago**

Hierarchical Structure of IBM.. from a Different Perspective



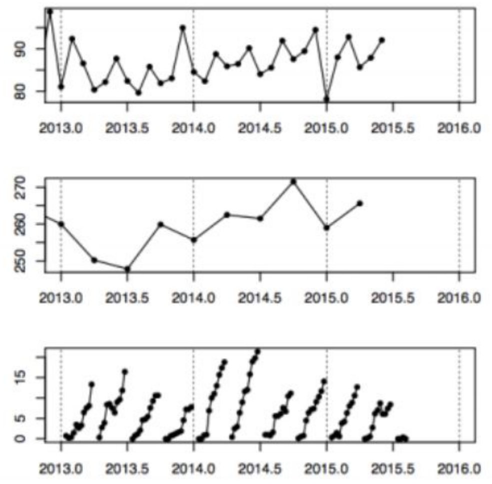
Challenge: Weekly Updates

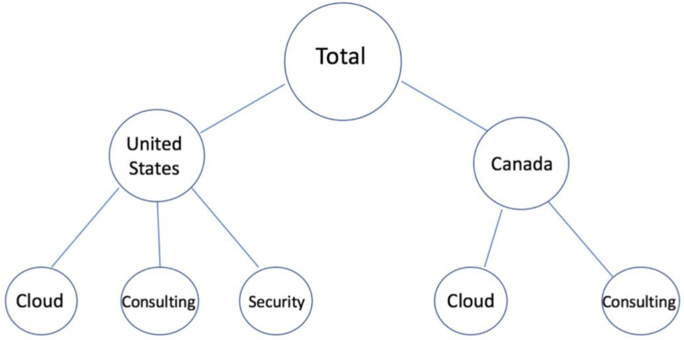
- For 26 weeks (13 ‘positive’, 13 ‘negative’), we target the same quarter assessment.
- As weeks go on, updated data comes in at different frequencies (since different parts of the business have different data).
- How to update while maintaining interpretability, robustness, and smoothness over time?



Challenge: Forecast Reconciliation

- Examples of multi-frequency data coming from different sources
- Data sources are updated at different times
- Data sources have different resolutions



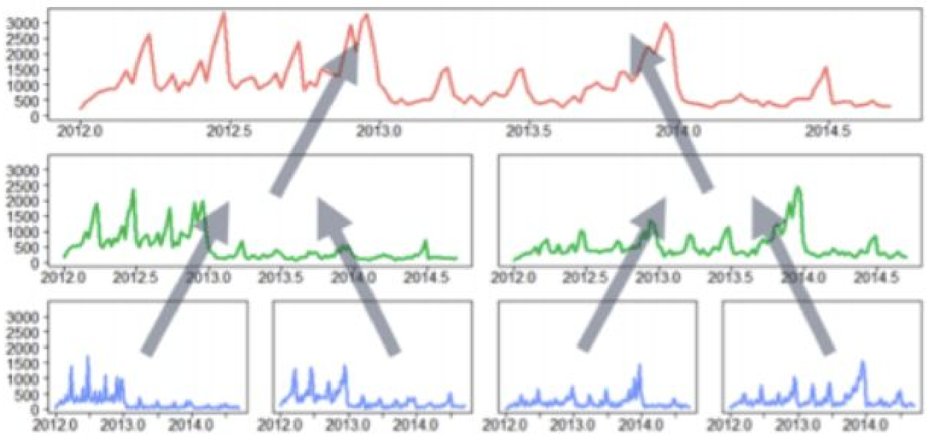


- It would make sense to forecast each node of the hierarchy separately… but how to we ensure aggregate consistency?

Method for “Within Node Forecasting”

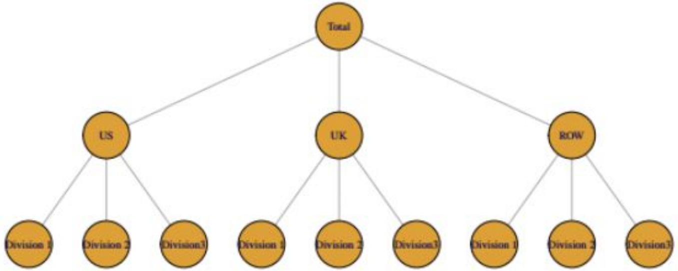
- For each node, the forecast is an ensemble of forecasts from different submodels
  - Time series models, regression models
- Weights for each forecast
  - Change every week (for 26 weeks)
  - Given weight 0 if they are considered outliers
  - Smoothed to prevent high variability in week to week changes

What Should We Do To Make the Forecasts Aggregate Consistent?



Methods for “Across Node Forecasting”/Forecast Reconciliation

1. Bottom-up: forecast series at the lowest level of the hierarchy and simply aggregate up.
2. Top-down: forecast the completely aggregated series (top level) and disaggregate based on historical proportions.
3. Hyndman et al. (2011, CSDA): standard least squares to ‘constrain’ forecasts to add up properly.
4. Hyndman et al. (2016, CSDA): weighted least squares to aggregate forecasts (e.g. weights based on one step ahead forecast accuracy).
5. Bayesian Hierarchical Forecasting: Bayesian regression perspective
6. And more…



Model Setup

Suppose we are provided with a vector of forecasts Y_t for each of the N nodes in a hierarchy graph:

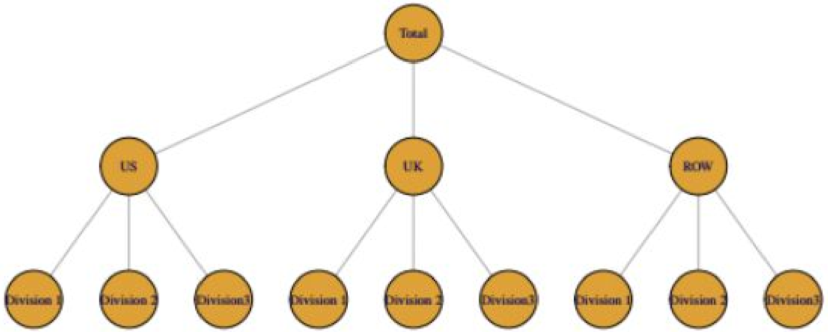
$$Y_t = \begin{pmatrix} Y_{t,1} \\ Y_{t,2} \\ \dots \\ Y_{t,N} \end{pmatrix}$$

The hierarchical structure imposes constraints on linear combinations of the $Y_{t,i}$'s:

$$Y_{t,1} + Y_{t,2} + Y_{t,3} = Y_{t,10}$$
$$Y_{t,4} + Y_{t,5} + Y_{t,6} = Y_{t,11}$$
$$\dots.$$
$$Y_{t,10} + Y_{t,11} + Y_{t,12} = Y_{t,13}$$

Model Setup

S, the aggregation matrix, combines the lowest level to give the higher level forecasts.



$$S = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Model Setup

We need to find a set of forecasts $\tilde{\beta}_t$ at the lowest level such that they add up properly and are as `similar' as possible to our given forecasts, Y_t.

$$Y_t \approx S\,\tilde{\beta}_t$$

The equation is written in this way to highlight the similarities to a standard linear regression model.

One Solution

Hyndman et al. (2011) proposes the least squares solution to the forecasting problem.

$$\tilde{\beta}_t = (S^T S)^{-1} S^T Y_t$$

$$\tilde{Y}_t = S\tilde{\beta}_t = S(S^T S)^{-1} S^T Y_t$$

This solution gives forecasts that are closest in the `least squares sense'.

Weighted Least Squares Solution

Hyndman et al. (2016) extended this work to a weighted least squares solution.

$$\tilde{\beta}_t = (S^T \Lambda S)^{-1} S^T \Lambda Y_t$$

$$\tilde{Y}_t = S\tilde{\beta}_t$$

where **Λ** is the diagonal matrix of weights (which take into account the variability of the base forecasts).

A Bayesian Approach

Another strategy to tackle this problem is as a Bayesian regression and infer the posterior for the forecasts.

- Assume the errors are independent (initially)

$$p(\{Y_t\}|\boldsymbol{\beta}_t) = p(Y_{t,1}|\boldsymbol{\beta}_t)p(Y_{t,2}|\boldsymbol{\beta}_t)\dots p(Y_{t,N}|\boldsymbol{\beta}_t)$$

- Assume a Gaussian distribution for the errors, ($\tilde{Y}_t = S\tilde{\beta}_t$ deal)

$$P(Y_{t,1}Y_{t,2}\cdots Y_{t,N}) = Ce^{\frac{-1}{2\sigma_{t,1}}(Y_{t,1}-\tilde{Y}_{t,1})^2}\cdots e^{\frac{-1}{2\sigma_{t,n}}(Y_{t,n}-\tilde{Y}_{t,n})^2}$$

- We now write

$$Y_t \sim N(S\beta_t, \boldsymbol{\Omega}_t)$$

Posterior Distributions

- Place noninformative priors on β_t and the variance

- Posterior for β_t:

$$\boldsymbol{\beta}_t | \sigma_t^2, \boldsymbol{Y}_t \sim N(\hat{\boldsymbol{\beta}}_t, \boldsymbol{V}_{\beta_t} \sigma_t^2)$$

- Posterior for σ_t^2:

$$\sigma_t^2 | \boldsymbol{Y}_t \sim Inv - \chi^2(m - m_K, s^2)$$

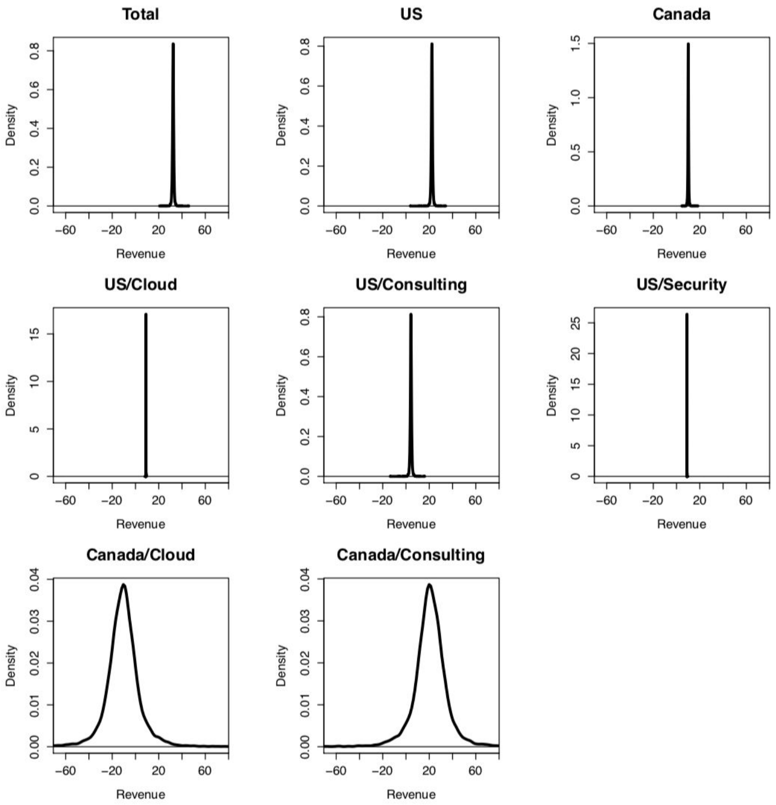
Loss Function

We are looking for an optimal $\tilde{\boldsymbol{Y}}_t^{\star}$ such that

$$\tilde{\boldsymbol{Y}}_t^{\star} = \underset{\boldsymbol{\beta}_t \in p_{\hat{\boldsymbol{\beta}}_t}}{\operatorname{argmin}} \ \mathbb{E}[\mathcal{L}(\boldsymbol{S}\boldsymbol{\beta}_t, p_{\tilde{\mathbf{Y}}_t})]$$

- The practitioner defines the loss given the context
- The uncertainty in the posterior of each of the nodes (both from prior knowledge and the data) is incorporated

Uncertainties in the Posterior Distributions



Future Directions

- Applications: hierarchies occur everywhere.
- Covariance matrix
  - Block diagonal
  - Best way to borrow information across the nodes?
- Missing data in certain nodes
- Priors from a Bayesian perspective


Questions?