Enterprise Business Metric Forecasting

Beatriz Etchegaray Garcia* Yasuo Amemiya†

Abstract

Regularly issued forecasts of various business metrics, such as revenue, cost, and event occurrences, play a key role in tracking and managing business performance. Some forecasts are for next several time periods, while others target specific term summaries such as quarterly totals. In an enterprise business, processes are often organized in a hierarchical and high-dimensional cube structure based on features such as product offerings or geography. Typically, forecasts need to be issued for every cell of the cube and must satisfy certain consistency relationships over the different dimensions. Forecasting for a hierarchical structure of time series is challenging because various components at different levels of the hierarchy can interact in a complex manner. A large number of time-dependent covariates are usually available, although the measurement periods may vary, and missing or unstable measurements are expected in some cells. Desirable forecast properties include reasonable accuracy at various levels or slices of a cube, and stability over time periods. In this talk we present actual enterprise forecasting problems and their challenges. We describe the development and implementation of methods for an operational solution effectively used by business.

Key Words: hierarchical forecasting, multivariate time series, enterprise structure

1. Enterprise Business Analytics

Business analytics is the use of quantitative analysis, modeling techniques and data to understanding business situations, improve business decisions and ultimately drive better planning and performance within an organization. Business analytics focuses on generating insightful information from data to create a measurable improvement in business objectives. Businesses use a variety of metrics, such as revenue, cost, sales, event occurrences, etc., to track the overall performance of the business as compared to previous months, years, or their competitors. Sequential tracking of business metrics is used to discover new knowledge and gain insight about the business.

Understanding what is happening and predicting what is likely to happen next helps drive strategic planning to manage and transform business processes. Forecasting is the process of making statements about future events. From a business perspective, forecasting is used to estimate future aspects of the business or operation. Forecasts can be used to aid in selecting a course of action to optimally achieve business objectives and performance targets. Internal and external data is gathered by businesses and used to look at past performance to determine future outcomes. For example, how much did we make last quarter or how much did we sell last year. Key to predicting future outcomes are understanding and learning past, current and new trends, things that are occurring now and in the future that will affect the future of the business.

The need to analyze and forecast the temporal evolution of business metrics motivates the use of times series and statistical forecasting methods. Time series models use past data to estimate future aspects of a business or operation based on a statistical model. In an enterprise business setting the time series of a business metric are typically organized in a hierarchical structure based on dimensions, such as client set and geography. Forecasting a business metric in a business enterprise setting, such as, for example, financial revenue,

*IBM Research, IBM T. J. Watson Research Center, Yorktown Heights, NY
†IBM Research, IBM T. J. Watson Research Center, Yorktown Heights, NY

would require modeling and forecasting many related time series. A quarterly revenue forecast that is issued every week during the quarter can be used as a metric to track business performance in terms of revenue and growth. Furthermore, deeper insights into business performance can be achieved using revenue forecasts and additional information, such as deviations from the budget or business target.

Tracking and forecasting business metrics requires a platform that includes the ability to access and organize data from multiple sources, can handle analytical reporting on a regular basis and integrates ad-hoc analysis capabilities to measure goals and performance. One important benefits of such a system is the ability to provide accurate and unbiased historical data as well as forecasts. In an enterprise business it is crucial to identify business opportunities and respond quickly to change in order to gain competitive advantage. Having a drill-down capability is also important as business enterprises are generally organized in a hierarchical structure spanning multiple brands, client sets, geographies, etc. By accurately identifying current performance levels though business metric analyses can help improve or modify future business strategy and goal setting.

In order to forecast business metrics it is essential to take into account current as well as historical business transactions that unbiasedly represent business transactions. Gathering and organizing historical data can be a challenge on its own in an enterprise business setting because of the inherent dynamical structure. Businesses are constantly changing through reorganizations, acquisitions and divestments. While gathering data may be a challenge, we will assume that organizations can restate the historical data that is necessary. In the following section we will present methodological issues specific to enterprise business analytics and business metric forecasting.

2. Methodological Issues Specific to Enterprise Business Analytics

Business analytics includes the processes, methods, and applications used to support data-driven business decision-making. Business metric tracking and forecasting uses both historical and real-time data to report, analyze and forecast quantities of interest to the business. Data needs to be extracted from a production system on a periodical basis and delivered to the appropriated end uses as well as to the applications that will process and analyze it for further use or reporting. In many ways, the practice of business analytics and business metric tracking and forecasting is closely tied to underlying information infrastructure of the business. There is no unique way of tracking, analyzing, and forecasting business metrics.

In order to build a tracking and forecasting tool for business metrics it is necessary to understand the dynamics in the data in order to use appropriate statistical methodology for analyzing and forecasting. There is a clear need and relevance to use statistical tools for forecasting. However, there are important methodological issues and considerations that are specific to enterprise business analytics. In the following sections we describe a few of the key issues that makes the business metric forecasting problem a nonstandard statistical time series problem.

2.1 Hierarchical High-Dimensional Structure

Business enterprises are traditionally organized in hierarchical structure. The structure is defined by the dimensions that best support internal operations and these may be based on the customer set, line of business, geography, etc. For example, IBM covers a large span of geographic regions around the world as well as different lines of businesses. Aligning the organizational hierarchy by line of business and geographical regions leads to better

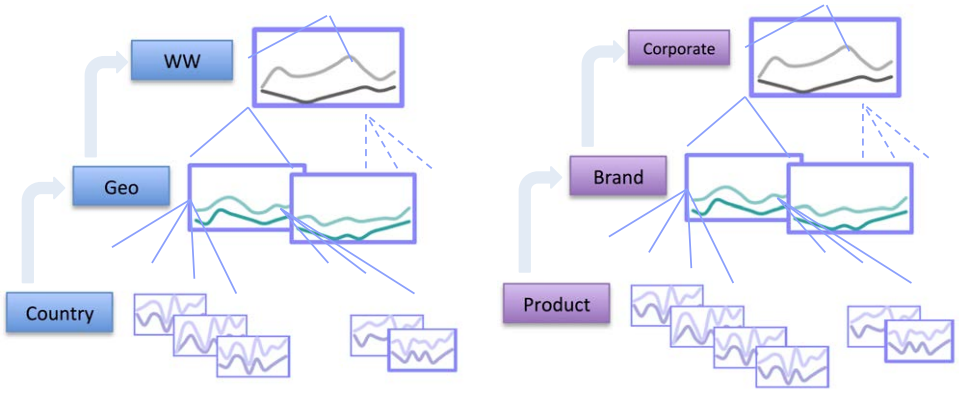

Figure 1: Business enterprises are traditionally organized in hierarchical structure with multiple dimensions.

management of the the overall business. On the other hand, having such a diverse and world-wide business makes the organizational hierarchy of an enterprise such as IBM a high-dimensional one (see Figure 1).

Time series and regression models provide a framework for forecasting. In a hierarchical setting, forecasting may be time consuming because of the dimensionally of the hierarchical structure. Furthermore, organizational hierarchies of enterprise businesses are comprised of multiple levels. Hence, there is the additional requirement that the forecasts in the hierarchy must be aggregate consistent (see Figure 2).

Typically three approaches are used: top-down, bottom-up, and middle-out. The top-down approach consists of forecasting the completely aggregated time series at the top of the hierarchy, and then disaggregating the forecasts. Different methods can be used to allocate the forecasts into the lower level such as the historical size or proportion of the metric. On the other hand, bottom-up consists of forecasting at the lowest level of the hierarchy. Aggregation is used to obtain the forecasts at higher levels. Middle-out is a combination of the two first approaches. With the middle-out approach an important additional issue is what intermediate level should be used to forecasts. None the methods described above capture information horizontally across the hierarchy.

In a hierarchical organizational structure it is important that the methodology is adapted for drill-down and drill-down capabilities. It is possible that the lowest levels of the hierarchy have little to no trend or seasonal dynamics. Nonetheless, forecasts at the lowest levels should align to realist expectations of the business.

Finally, the various cells in the hierarchy can interact with each other in complex ways. A change in the Asian economy will impact revenue of certain countries in the region, if not the whole world. Hence, series at one level may influence the dynamics of series at the same level, as well as series in higher and lower levels. As a result, it is important to consider the interdependencies in the hierarchy.

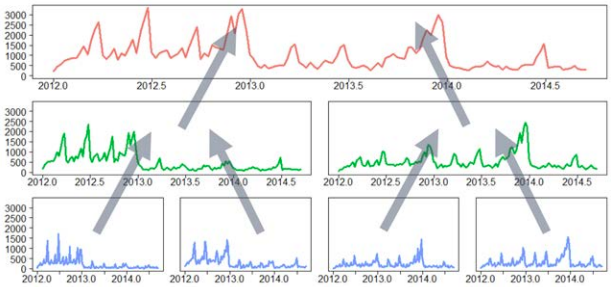

Figure 2: A basic requirement is that forecasts of a business metric must be aggregate consistent.

2.2 Multi-cycle Time Series

Time periods of the data need to forecast a business metric may vary. Data can be collected on a quarterly, monthly, weekly, daily, etc., basis. The business metric that is targeted will have a frequency that is dictated by what the business needs. For example, by quarter revenue is a natural frequency for enterprise businesses. However, the majority of the information used to predict end of quarter revenue may be available more frequently. The more granular data should be a valuable source of information. The January and February revenue actuals are good predictors of first quarter revenue. Using multi-frequency data is a methodological challenge when building a forecasting business metric tool (see Figure 3). Separate models can be used for data of different frequency. However, it may be the case that the use of multi-frequency data improves forecasting accuracy.

Another challenge is setting forecasting window. If the forecasting window is fixed, its value is usually set by what the business needs. For example, a business may need to forecasting volume for the next 13 weeks on a bi-weekly basis so as to optimally allocate resources and assign vacation in sales support center. On the other hand, it may be useful to the business to target one final value, multiple times. Such is the case when forecasting end-of-quarter revenue on a weekly basis. In this particular example, at the end of the quarter one will have 13 predictions of the same target: the end-of-the-quarter revenue. Each week, the forecast should become more accurate in light of newer, more updated information.

Another important issue to consider is the importance of model development and testing using a real-time data setting. Data timing issues are common in an enterprise business setting, there may be blackout dates and actual values of certain quantities may not be available immediately. Suppose we are looking to forecast fourth quarter revenue, and we will do so weekly. For the first couple of weeks, the most recent monthly information available is not the third quarter revenue. This may be due to delays in the system. What this may result in are worst forecasts, i.e., less accurate, for the first few weeks. A realistic evaluation of the model must take into account the nuances of the real-time data stream.

2.3 Multiple Userbase Targets

One important issue to consider are the multiple participants in building a system to forecast business metrics. Although technically this is not a methodological issue, it does con-

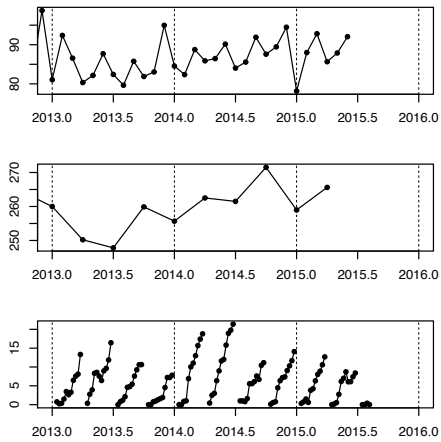

Figure 3: Using multi-frequency data is a methodological challenge when building a forecasting business metric tool.

strain methodology in the sense that information needs to be communicated, understood and transparent. In addition, the reporting systems need to be able to support the outputs. Enterprise statisticians and modelers are needed to build such a system with the aid and support of business administrators. Models and methodology need to be developed in line with the business goals and hence a discussion and inputs is needed from business decision makers and planners. Managers at different levels of an enterprise will be ultimately be the clients: the users for which the tracking and forecasting system is built. Information that is reported has to be in line with business expectations with respect to time and format.

2.4 System Requirements and Usability

Another nonmethodological issue that does constrains methodology are the system requirements. A tracking and forecasting system must be flexible, configurable and robust to the continuous changes of the business. Implementation has to be effective and convenient to maintain over time. More importantly, the system needs a level of stability of operation over time. A certain level of robustness is needed against data entry errors and mistakes. Self-updating data is required for the automation of a solution. Finally, the system needs to have business intelligence capability so as to add value by gaining insights into the business, predicting accurately what is happing and prescribing the best action to take.

3. Enterprise Statistical Forecasting Framework

When building an enterprise statistical forecasting framework accuracy should be targeted, however we believe it should not be prioritized. Capturing the true dynamics and presenting the information in a way so that the users gain more insight into the business should also be targeted. This will result in a trade-off because many times optimality based on forecast accuracy will be in conflict with the needs and requirements of the business. As an example, in order to use a middle-out approach to hierarchical forecasting, the level of

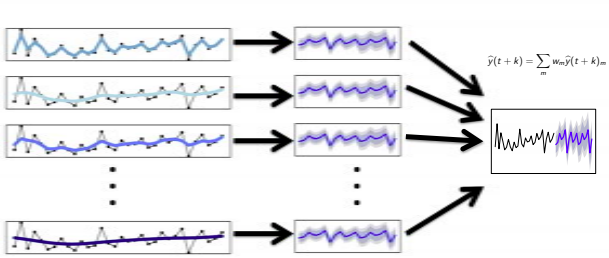

Figure 4: Ensemble forecasting consists of combining forecasts obtained from different models using a combination scheme.

the hierarchy must be selected a priori. Selecting an the optimum base level can be done by minimizing the forecast error. However, if forecasting is done on a weekly basis, the base level will likely change regularly. This in turn will may lead to significant week to week changes, not because of the nature of the business, but because we are using different models. Furthermore, choosing a hierarchical level based on accuracy may also contradict the nature of the business: accuracy is needed throughout the hierarchy because different managerial levels will make business decision based on the forecasts.

There is also the question of what time series models to include in order to create forecasts for all of the cells in the hierarchy. There are multiple time series models that can be used, such as the ARIMA and the exponential smoothing families. Models may or may not include covariates. Candidate models can and should be proposed by analyzing historical patterns in the cells of the hierarchy. Business logic should aid the selection of the set of proposed models. The set of candidate models should address commonly observed issues of seasonality, irregular cycles, volatility, break points, outliers and autocorrelation.

Automatic model selection can be used to determine the optimal model to use within each cell. However, in a business setting, this may not be ideal. If forecasts are produced on a weekly basis, then there is an issue with model selection. It may very well be the case that every week we pick a different "best" model. In this particular scenario, large variations will result in week to week comparison of the forecasts. Single model-based forecasting can be insufficient for producing both accurate and stable repeated forecasts.

Instead, an ensemble forecasting approach can be used (see Figure 4). Ensemble forecasting consists of combining forecasts obtained from different models using a combination scheme. The combination scheme consists of weights and hence the ensemble forecast is a weighted linear combination of the forecasts. Weights can be calculated in numerous ways. For example, weights can be based on historical performance of a model, i.e., forecasting accuracy. In addition, ensemble forecasting is appropriate for a hierarchical structure since candidate submodes that share information across the hierarchy can be included and considered. In addition, cross level models can be included. In an ensemble setting, forecasts and weights are specific to the cells of the hierarchy and on a specific time period. Smoothness and consistency on a week to week basis of the forecasts can be controlled by the weights.

REFERENCES

Clemen, R. T. (1989), "Combining forecasts: A review and annotated bibliography," *International journal of forecasting*, 5(4), 559-583.

Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., & Shang, H. L. (2011), "Optimal combination forecasts for hierarchical time series," *Computational Statistics & Data Analysis*, 55(9), 2579-2589.

Shan, J. Z., Tang, H. K., Wu, R., & Safai, F. (2005), "Dynamic modeling and forecasting on enterprise revenue with derived granularities," *2005 IEEE International Conference on Granular Computing*, (Vol. 1, pp. 243-248).

Zou, H., & Yang, Y. (2004), "Combining time series models for forecasting," *International Journal of Forecasting*, 20(1), 69-84.