Computational Statistics and Data Analysis 55 (2011) 2579–2589

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Optimal combination forecasts for hierarchical time series

Rob J. Hyndman *, Roman A. Ahmed, George Athanasopoulos, Han Lin Shang

Department of Econometrics and Business Statistics, Monash University, VIC 3800, Australia

ARTICLE INFO

Article history:
Received 28 March 2008
Received in revised form 10 September 2010
Accepted 15 March 2011
Available online 23 March 2011

Keywords:
Bottom-up forecasting
Combining forecasts
GLS regression
Hierarchical forecasting
Reconciling forecasts
Top-down forecasting

ABSTRACT

In many applications, there are multiple time series that are hierarchically organized and can be aggregated at several different levels in groups based on products, geography or some other features. We call these "hierarchical time series". They are commonly forecast using either a "bottom-up" or a "top-down" method.

In this paper we propose a new approach to hierarchical forecasting which provides optimal forecasts that are better than forecasts produced by either a top-down or a bottom-up approach. Our method is based on independently forecasting all series at all levels of the hierarchy and then using a regression model to optimally combine and reconcile these forecasts. The resulting revised forecasts add up appropriately across the hierarchy, are unbiased and have minimum variance amongst all combination forecasts under some simple assumptions.

We show in a simulation study that our method performs well compared to the top-down approach and the bottom-up method. We demonstrate our proposed method by forecasting Australian tourism demand where the data are disaggregated by purpose of travel and geographical region.

1. Introduction

In business and economics, there are often applications requiring forecasts of many related time series organized in a hierarchical structure based on dimensions, such as product and geography. This has led to the need for reconciling forecasts across the hierarchy (that is, ensuring that the forecasts sum appropriately across the levels). We propose a new statistical method for forecasting hierarchical time series which (1) provides point forecasts that are reconciled across the levels of the hierarchy; (2) allows for the correlations and interactions between the series at each level of the hierarchy; (3) provides estimates of forecast uncertainty which are reconciled across the levels of the hierarchy; and (4) is sufficiently flexible that ad hoc adjustments can be incorporated, information about individual series can be allowed for, and important covariates can be included. Furthermore, our method provides optimal forecasts under some simple assumptions.

The problem of forecasting hierarchical time series arises in many different contexts. For example, forecasting manufacturing demand typically involves a hierarchy of time series. One of us has worked for a disposable tableware manufacturer who wanted forecasts of all paper plates, of each different type of paper plate, and of each type of plate at each distribution outlet. Another one of us has been involved in forecasting net labor turnover. Not only is it important to forecast the rate of job turnover in the economy as a whole and across major occupational groups, but it is also important to do so at the individual occupation level. The hierarchical structure according to the Australian Standard Classification of Occupations (ASCO), starting from the highest level, can be illustrated as follows:

- All employed persons
  - Professionals (major group)

* Corresponding author.
E-mail addresses: Rob.Hyndman@monash.edu (R.J. Hyndman), Roman.Ahmed@monash.edu (R.A. Ahmed), George.Athanasopoulos@monash.edu (G. Athanasopoulos), HanLin.Shang@monash.edu (H.L. Shang).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2011.03.006

- ∘ Educational professionals (sub-major group)
  - · School teachers (minor group)
    - * Pre-primary teachers (unit group)
    - * Primary teachers (unit group); etc.

There are 340 unit groups in ASCO. Further divisions of the unit groups can be made by gender and age variables. The series at the lowest level can be short in length with a high degree of volatility but aggregate behavior may be relatively smooth. Thus the problem here is that of forecasting a set of time series that are *hierarchical* in structure and clusters of which may be correlated. In Section 2, we introduce some notation to allow the problem of hierarchical forecasting to be defined more precisely.

The various components of the hierarchy can interact in varying and complex ways. A change in one series at one level, can have a consequential impact on other series at the same level, as well as series at higher and lower levels. By modeling the entire hierarchy of time series simultaneously, we will obtain better forecasts of the component series.

Existing approaches to hierarchical forecasting usually involve either a top-down or bottom-up method, or a combination of the two. The top-down method entails forecasting the completely aggregated series, and then disaggregating the forecasts based on historical proportions. Gross and Sohl (1990) discuss several possible ways of choosing these proportions. The bottom-up method involves forecasting each of the disaggregated series at the lowest level of the hierarchy, and then using simple aggregation to obtain forecasts at higher levels of the hierarchy. In practice, many businesses combine these methods (giving what is sometimes called the "middle-out" method) where forecasts are obtained for each series at an intermediate level of the hierarchy, and then aggregation is used to obtain forecasts at higher levels and disaggregation is used to obtain forecasts at lower levels. None of these methods take account of the inherent correlation structure of the hierarchy, and it is not easy to obtain prediction intervals for the forecasts from any of these methods.

Of course, it is also possible to forecast all series at all levels independently, but this has the undesirable consequence of the higher level forecasts not being equal to the sum of the lower level forecasts. Consequently, if this method is used, some adjustment is then carried out to ensure that the forecasts add up appropriately. These adjustments are usually done in an ad hoc manner.

In this paper, we present a framework for general hierarchical forecasting in Section 3, and show that existing methods are special cases of this framework. We also show how to compute prediction intervals for any of the methods that are special cases of our framework.

Most of the forecasting literature in this area has looked at the comparative performance of the top-down and bottom-up methods. An early contribution was Grunfeld and Griliches (1960) who argued that the disaggregated data are error prone and that top-down forecasts may therefore be more accurate. Similar conclusions were drawn by Fogarty et al. (1990) and Narasimhan et al. (1995). Fliedner (1999) also argued that aggregate forecast performance is better with aggregate level data. On the other hand, Orcutt et al. (1968) and Edwards and Orcutt (1969) argued that information loss is substantial in aggregation and therefore the bottom-up method gives more accurate forecasts. Shlifer and Wolff (1979) compared the forecasting performance of both methods and concluded that the bottom-up method is preferable under some conditions on the structure of the hierarchy and the forecast horizon. Schwarzkopf et al. (1988) looked at the bias and robustness of the two methods and concluded that the bottom-up method is better except when there are missing or unreliable data at the lowest levels.

Empirical studies have supported the efficacy of bottom-up forecasting over top-down forecasting. For example, Kinney (1971) found that disaggregated earnings' data by market segments resulted in more accurate forecasts than when firm level data were used. Collins (1976) compared segmented econometric models with aggregate models for a group of 96 firms, and found that the segmented models produced more accurate forecasts for both sales and profit. The study of telephone demand by Dunn et al. (1976) shows that forecasts aggregated from lower level modeling are more accurate than the top-down method. Zellner and Tobias (2000) used annual GDP growth rates from 18 countries and found that disaggregation provided better forecasts. Dangerfield and Morris (1992) constructed artificial 2-level hierarchies using the M-competition data with two series at the bottom level, and found that bottom-up forecasts were more accurate, especially when the two bottom level series were highly correlated.

In the econometric literature, there has also been some interest in the potential improvements in forecast accuracy that are possible by aggregating component forecasts rather than simply forecasting the aggregate itself (e.g., Fair and Shiller, 1990; Zellner and Tobias, 2000; Marcellino et al., 2003; Espasa et al., 2002; Hubrich, 2005).

Tiao and Guttman (1980) and Kohn (1982) used more theoretical arguments to show that the efficiency of aggregation depends on the covariance structure of the component series. Shing (1993) discussed some time series models and demonstrated that there is no uniform superiority of one method over the other. Fliedner and Lawrence (1995) concluded that current formal hierarchical forecasting techniques have no advantage over some informal strategies of hierarchical forecasting. Kahn (1998) suggested that it is time to combine the existing methodologies so that we can enjoy the good features of both methods, but no specific ideas were provided in that discussion. Another very good discussion paper is Fliedner (2001), which summarizes the uses and application guidelines for hierarchical forecasting. However, none of these papers provide any new methods.

In Section 4, we take up the call of Kahn (1998) by proposing a new methodology which takes the best features of existing methods, and provides a sound statistical basis for optimal hierarchical forecasting. We discuss computational issues associated with our method in Section 5.