Computational Statistics and Data Analysis 97 (2016) 16–32

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Fast computation of reconciled forecasts for hierarchical and grouped time series

Rob J. Hyndman [a,*], Alan J. Lee [b], Earo Wang [a]

[a] *Department of Econometrics and Business Statistics, Monash University, VIC 3800, Australia*

[b] *Department of Statistics, University of Auckland, New Zealand*

ARTICLE INFO

Article history:
Received 17 November 2014
Received in revised form 17 November 2015
Accepted 17 November 2015
Available online 30 November 2015

Keywords:
Combining forecasts
Grouped time series
Hierarchical time series
Reconciling forecasts
Weighted least squares

ABSTRACT

It is shown that the least squares approach to reconciling hierarchical time series forecasts can be extended to much more general collections of time series with aggregation constraints. The constraints arise due to the need for forecasts of collections of time series to add up in the same way as the observed time series. It is also shown that the computations involved can be handled efficiently by exploiting the structure of the associated design matrix, or by using sparse matrix routines. The proposed algorithms make forecast reconciliation feasible in business applications involving very large numbers of time series.

1. Introduction

Time series can often be naturally disaggregated in a hierarchical or grouped structure. For example, a manufacturing company can disaggregate total demand for their products by country of sale, retail outlet, product type, package size, and so on. As a result, there can be tens of thousands of individual time series to forecast at the most disaggregated level, plus additional series to forecast at higher levels of aggregation.

The most disaggregated series often have a high degree of volatility (and are therefore hard to forecast), while the most aggregated time series is usually smooth and less noisy (and is therefore easier to forecast). Consequently, forecasting only the most disaggregated series and summing the results tends to give poor results at the higher levels of aggregation. On the other hand, if all the series at all levels of aggregation are forecast independently, the forecasts will not add up consistently between levels of aggregation. Therefore, it is necessary to reconcile the forecasts to ensure that the forecasts of the disaggregated series add up to the forecasts of the aggregated series (Fliedner, 2001).

In some cases, the disaggregation can be expressed in a hierarchical structure. For example, time series of sales may be disaggregated geographically by country, state and region, where each level of disaggregation sits within one node at the higher level (a region is within a state which is within a country).

In other cases, the disaggregation may not be uniquely hierarchical. For example, time series of sales may involve a geographical grouping (by country) and a product grouping (by type of product sold). Each of these grouping variables leads to a unique hierarchy, but if we wish to use both grouping variables, the order in which the disaggregation occurs is not unique. We may disaggregate by geography first, then by product type, or we may disaggregate by product type, then by

* Corresponding author.
E-mail addresses: rob.hyndman@monash.edu (R.J. Hyndman), lee@stat.auckland.ac.nz (A.J. Lee), yiru.wang@monash.edu (E. Wang).

http://dx.doi.org/10.1016/j.csda.2015.11.007

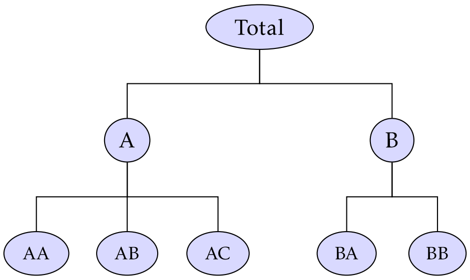

Fig. 1. A simple hierarchy of time series containing five bottom-level series.

geography. Both types of disaggregation may be useful, so choosing only one of these hierarchies is not appropriate. We call this a grouped time series. In general, there can be many different grouping variables, some of which may be hierarchical.

Hyndman et al. (2011, CSDA) proposed a method for optimally reconciling forecasts of all series in a hierarchy to ensure they add up. Our first contribution (Section 2) is to show that this result is easily extended to cover non-hierarchical groups of time series, and groups with a partial hierarchical structure.

The optimal reconciliation method involves fitting a linear regression model where the design matrix has one column for each of the series at the most disaggregated level. Consequently, for problems involving tens of thousands (or even millions) of series, the model is impossible to estimate using standard regression algorithms such as the QR decomposition.

The second contribution of this paper is to propose a solution to this problem, exploiting the unique structure of the linear model to efficiently estimate the coefficients, even when there are millions of time series at the most disaggregated level. Our solution involves two important cases: (1) where there can be any number of groups that are strictly hierarchical (Section 3); and (2) where there are two groups that may not be hierarchical (Section 4). All other cases can be handled using sparse matrix algorithms as discussed in Section 5.

For smaller hierarchies, where the computation is tractable without our new algorithms, good results have been obtained in forecasting tourism demand in Australia (Athanasopoulos et al., 2009) and inflation in Mexico (Capistrán et al., 2010).

Our algorithms make forecast reconciliation on very large collections of grouped and hierarchical time series feasible in practice. The algorithm has applications in any situation where large numbers of related time series need to be forecast, particularly in forecasting demand in product hierarchies, or geographical hierarchies. In Section 6 we demonstrate the different algorithms on a collection of quarterly time series giving the number of people holding different occupations in Australia.

2. Optimal reconciliation of hierarchical and grouped time series

2.1. Hierarchical time series

To introduce the notation and key ideas, we will use a very small and simple hierarchy, shown in Fig. 1, with five series at the most disaggregated level. The "Total" node is the most aggregate level of the data, with the tth observation denoted by y_t. It is disaggregated into series A and B, which are each further disaggregated into the five bottom-level series. The tth observation at node X is denoted by $y_{X,t}$.

For any time t, the observations of the bottom level series will aggregate to the observations of the series above. To represent this in matrix notation, we introduce the $n \times n_K$ "summing" matrix $\boldsymbol{S}$. For the hierarchy in Fig. 1 we can write

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

or in more compact notation $\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{b}_t$, where $\boldsymbol{y}_t$ is a vector of all the observations in the hierarchy at time t, $\boldsymbol{S}$ is the summing matrix as defined above, and $\boldsymbol{b}_t$ is a vector of all the observations in the bottom level of the hierarchy at time t.

We are interested in generating forecasts for each series in the hierarchy. Let $\hat{y}_{X,h}$ be an h-step-ahead forecast for the series at node X and let $\hat{y}_h$ be an h-step-ahead forecast generated for the "Total" series, computed using the data to time T. These forecasts are computed independently for each series, using only the history of the series and not taking account of any relationships with other series in the hierarchy. We refer to these as "base forecasts". The sum of the base forecasts at lower levels are unlikely to be equal to the base forecast of their parent nodes.