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 © 2011 Elsevier B.V. All rights reserved.
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.

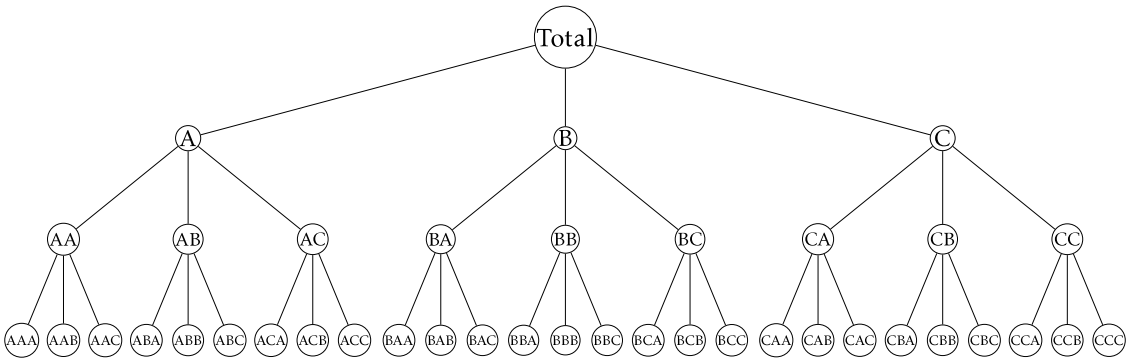

Fig. 1. A 3-level hierarchical tree diagram.

The performance of our optimal hierarchical forecasts is evaluated using a simulation exercise in Section 6, where we compare our method with the major existing methods. Then, in Section 7, we apply the various methods to some real data, Australian domestic tourism demand, disaggregated by geographical region and by purpose of travel. In both the simulations and the real data application, we find that our method, on average, produces significantly more accurate forecasts than existing approaches.

We conclude the paper by summarizing our findings and suggesting some possible extensions in Section 8.

2. Notation for hierarchical forecasting

Consider a multi-level hierarchy, where level 0 denotes the completely aggregated series, level 1 the first level of disaggregation, down to level K containing the most disaggregated time series. We use a sequence of letters to identify the individual series and the level of disaggregation. For example: A denotes series A at level 1; AF denotes series F at level 2 within series A at level 1; AFC denotes series C at level 3 within series AF at level 2; and so on.

To be specific, suppose we had three levels in the hierarchy with each group at each level consisting of three series. In this case, $K = 3$ and the hierarchy has the tree structure shown in Fig. 1.

It is assumed that observations are recorded at times $t = 1, 2, \ldots, n$, and that we are interested in forecasting each series at each level at times $t = n+1, n+2, \ldots, n+h$. It will sometimes be convenient to use the notation X to refer to a generic series within the hierarchy. Observations on series X are written as $Y_{\mathrm{X},t}$. Thus, $Y_{\mathrm{AF},t}$ is the value of series AF at time t. We use Y_t for the aggregate of all series at time t. Therefore

$$Y_t = \sum_i Y_{i,t}, \qquad Y_{i,t} = \sum_j Y_{ij,t}, \qquad Y_{ij,t} = \sum_k Y_{ijk,t}, \qquad Y_{ijk,t} = \sum_\ell Y_{ijk\ell,t},$$

and so on. Thus, observations at higher levels can be obtained by summing the series below.

Let m_i denote the total number of series at level i, $i = 0, 1, 2, \ldots, K$. So $m_i > m_{i-1}$ and the total number of series in the hierarchy is $m = m_0 + m_1 + m_2 + \cdots + m_K$. In the example above, $m_i = 3^i$ and $m = 40$.

It will be convenient to work with matrix and vector expressions. We let $\boldsymbol{Y}_{i,t}$ denote the vector of all observations at level i and time t and $\boldsymbol{Y}_t = \left[Y_t, \boldsymbol{Y}_{1,t}, \ldots, \boldsymbol{Y}_{K,t}\right]'$. Note that

$$\boldsymbol{Y}_t = \boldsymbol{S}\boldsymbol{Y}_{K,t} \tag{1}$$

where $\boldsymbol{S}$ is a "summing" matrix of order $m \times m_K$ used to aggregate the lowest level series. In the above example, $\boldsymbol{Y}_t = \left[Y_t, Y_{\mathrm{A},t}, Y_{\mathrm{B},t}, Y_{\mathrm{C},t}, Y_{\mathrm{AA},t}, Y_{\mathrm{AB},t}, \ldots, Y_{\mathrm{CC},t}, Y_{\mathrm{AAA},t}, Y_{\mathrm{AAB},t}, \ldots, Y_{\mathrm{CCC},t}\right]'$ and the summation matrix is of order 40×27 and is given by

$$\boldsymbol{S} = \begin{bmatrix} 111111111111111111111111111 \\ 111111111000000000000000000 \\ 000000000111111111000000000 \\ 000000000000000000111111111 \\ 111000000000000000000000000 \\ 000011100000000000000000000 \\ \vdots \\ 000000000000000000000000111 \\ 100000000000000000000000000 \\ 010000000000000000000000000 \\ \vdots \\ 000000000000000000000000001 \end{bmatrix}.$$

The rank of $\boldsymbol{S}$ is m_K. It is clear that the $\boldsymbol{S}$ matrix can be partitioned by the levels of the hierarchy. The top row is a unit vector of length m_K and the bottom section is an $m_K \times m_K$ identity matrix. The middle parts of $\boldsymbol{S}$ are vector diagonal rectangular matrices.

While aggregation is our main interest here, the results to follow are sufficiently general that $\boldsymbol{S}$ can be a general linear operator, and need not be restricted to aggregation.

3. General hierarchical forecasting

Suppose we first compute forecasts for each series at each level giving m base forecasts for each of the periods $n+1, \ldots, n+h$, based on the information available up to and including time n. We denote these base forecasts by $\hat{Y}_{\mathrm{X},n}(h)$, where X denotes the series being forecasted. Thus, $\hat{Y}_n(h)$ denotes the h-step-ahead base forecast of the total, $\hat{Y}_{\mathrm{A},n}(h)$ denotes the forecast of series A, $\hat{Y}_{\mathrm{AC},n}(h)$ denotes the forecast of series AC, and so on. We let $\hat{\boldsymbol{Y}}_n(h)$ be the vector consisting of these base forecasts, stacked in the same series order as for $\boldsymbol{Y}_t$.

All existing hierarchical forecasting methods can then be written as

$$\tilde{\boldsymbol{Y}}_n(h) = \boldsymbol{SP}\hat{\boldsymbol{Y}}_n(h) \tag{2}$$

for some appropriately chosen matrix $\boldsymbol{P}$ of order $m_K \times m$. That is, existing methods involve linear combinations of the base forecasts. These linear combinations are "reconciled" in the sense that lower level forecasts sum to give higher level forecasts. The effect of the $\boldsymbol{P}$ matrix is to extract and combine the relevant elements of the base forecasts $\hat{\boldsymbol{Y}}_n(h)$, which are then summed by $\boldsymbol{S}$ to give the final revised hierarchical forecasts, $\tilde{\boldsymbol{Y}}_n(h)$.

For example, bottom-up forecasts are obtained using

$$\boldsymbol{P} = \left[\boldsymbol{0}_{m_K \times (m-m_K)} \mid \boldsymbol{I}_{m_K}\right], \tag{3}$$

where $\boldsymbol{0}_{\ell\times k}$ is a null matrix of order $\ell \times k$ and $\boldsymbol{I}_k$ is an identity matrix of order $k \times k$. In this case, the $\boldsymbol{P}$ matrix extracts only bottom level forecasts from $\hat{\boldsymbol{Y}}_n(h)$, which are then summed by $\boldsymbol{S}$ to give the bottom-up forecasts.

Top-down forecasts are obtained using

$$\boldsymbol{P} = \left[\boldsymbol{p} \mid \boldsymbol{0}_{m_K \times (m-1)}\right] \tag{4}$$

where $\boldsymbol{p} = [p_1, p_2, \ldots, p_{m_K}]'$ is a vector of proportions that sum to one. The effect of the $\boldsymbol{P}$ matrix here is to distribute the forecast of the aggregate to the lowest level series. Different methods of top-down forecasting lead to different proportionality vectors $\boldsymbol{p}$.

Variations such as middle-out forecasts are possible by defining the matrix $\boldsymbol{P}$ appropriately. This suggests that other (new) hierarchical forecasting methods may be defined by choosing a different matrix $\boldsymbol{P}$ provided we place some restrictions on $\boldsymbol{P}$ to give sensible forecasts.

If we assume that the base (independent) forecasts are unbiased (that is, $\mathrm{E}[\hat{\boldsymbol{Y}}_n(h)] = \mathrm{E}[\boldsymbol{Y}_n(h)]$), and that we want the revised hierarchical forecasts to also be unbiased, then we must require $\mathrm{E}[\tilde{\boldsymbol{Y}}_n(h)] = \mathrm{E}[\boldsymbol{Y}_n(h)] = \boldsymbol{S}\mathrm{E}[\boldsymbol{Y}_{K,n}(h)]$. Suppose $\boldsymbol{\beta}_n(h) = \mathrm{E}[\boldsymbol{Y}_{K,n+h} \mid \boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n]$ is the mean of the future values of the bottom level K. Then $\mathrm{E}[\tilde{\boldsymbol{Y}}_n(h)] = \boldsymbol{SP}\mathrm{E}[\hat{\boldsymbol{Y}}_n(h)] = \boldsymbol{SPS}\boldsymbol{\beta}_n(h)$. So, the unbiasedness of the revised forecast will hold provided

$$\boldsymbol{SPS} = \boldsymbol{S}. \tag{5}$$

This condition is true for the bottom-up method with $\boldsymbol{P}$ given by (3). However, using the top-down method with $\boldsymbol{P}$ given by (4), we find that $\boldsymbol{SPS} \neq \boldsymbol{S}$ for any choice of $\boldsymbol{p}$. So the top-down method can *never* give unbiased forecasts even if the base forecasts are unbiased.

Let the variance of the base forecasts, $\hat{\boldsymbol{Y}}_n(h)$, be given by $\boldsymbol{\Sigma}_h$. Then the variance of the revised forecasts is given by

$$\mathrm{Var}[\tilde{\boldsymbol{Y}}_n(h)] = \boldsymbol{SP}\boldsymbol{\Sigma}_h\boldsymbol{P}'\boldsymbol{S}'. \tag{6}$$

Thus, prediction intervals on the revised forecasts can be obtained provided $\boldsymbol{\Sigma}_h$ can be reliably estimated. Note that result (6) applies to all the existing methods that can be expressed as (2) including bottom-up, top-down and middle-out methods.

4. Optimal forecasts using regression

We can write the base forecasts as

$$\hat{\boldsymbol{Y}}_n(h) = \boldsymbol{S}\boldsymbol{\beta}_n(h) + \boldsymbol{\varepsilon}_h \tag{7}$$

where $\boldsymbol{\beta}_n(h) = \mathrm{E}[\boldsymbol{Y}_{K,n+h} \mid \boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n]$ is the unknown mean of the bottom level K, and $\boldsymbol{\varepsilon}_h$ has zero mean and covariance matrix $\mathrm{Var}(\boldsymbol{\varepsilon}_h) = \boldsymbol{\Sigma}_h$. This suggests that we can estimate $\boldsymbol{\beta}_n(h)$ by treating (7) as a regression equation, and thereby obtain

forecasts for all levels of the hierarchy. If $\boldsymbol{\Sigma}_h$ was known, we could use generalized least squares estimation to obtain the minimum variance unbiased estimate of $\boldsymbol{\beta}_n(h)$ as

$$\hat{\boldsymbol{\beta}}_n(h) = (\boldsymbol{S}'\boldsymbol{\Sigma}_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{\Sigma}_h^{\dagger}\hat{\boldsymbol{Y}}_n(h) \tag{8}$$

where $\boldsymbol{\Sigma}_h^{\dagger}$ is the Moore–Penrose generalized inverse of $\boldsymbol{\Sigma}_h$. We use a generalized inverse because $\boldsymbol{\Sigma}_h$ is often (near) singular due to the aggregation involved in $\boldsymbol{Y}_n$. This leads to the following revised forecasts

$$\tilde{\boldsymbol{Y}}_n(h) = \boldsymbol{S}\hat{\boldsymbol{\beta}}_n(h) = \boldsymbol{SP}\hat{\boldsymbol{Y}}_n(h)$$

where $\boldsymbol{P} = (\boldsymbol{S}'\boldsymbol{\Sigma}_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{\Sigma}_h^{\dagger}$. Clearly, this satisfies the unbiasedness property (5). The variance of these forecasts is given by

$$\mathrm{Var}[\tilde{\boldsymbol{Y}}_n(h)] = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{\Sigma}_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'.$$

The difficulty with this method is that it requires knowledge of $\boldsymbol{\Sigma}_h$, or at least a good estimate of it. In a large hierarchy, with thousands of series, this may not be possible.

However, we can greatly simplify the computations by assuming that the error in (7) can be expressed as $\boldsymbol{\varepsilon}_h \approx \boldsymbol{S}\boldsymbol{\varepsilon}_{K,h}$, where $\boldsymbol{\varepsilon}_{K,h}$ is the forecast error in the bottom level. That is, we assume that the errors satisfy the same aggregation constraint as the original data (1). This assumption will be true provided the forecasts also approximately satisfy this aggregation constraint, which should occur for any reasonable set of forecasts. In some cases, it is possible for the errors to satisfy the expression exactly. For example, if a linear forecasting method (e.g., an ARIMA model) with fixed parameters was used for all series, then the errors would be exactly additive in this way. So the assumption is a reasonable approximation of what happens in practice.

The approximation leads to the result $\boldsymbol{\Sigma}_h \approx \boldsymbol{S}\boldsymbol{\Omega}_h\boldsymbol{S}'$, where $\boldsymbol{\Omega}_h = \mathrm{Var}(\boldsymbol{\varepsilon}_{K,h})$. We are now ready to state our main result.

Theorem 1. *Let $\boldsymbol{Y} = \boldsymbol{S}\boldsymbol{\beta}_h + \boldsymbol{\varepsilon}$ with $\mathrm{Var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma}_h = \boldsymbol{S}\boldsymbol{\Omega}_h\boldsymbol{S}'$ and $\boldsymbol{S}$ a "summing" matrix. Then the generalized least squares estimate of $\boldsymbol{\beta}$ obtained using the Moore–Penrose generalized inverse is independent of $\boldsymbol{\Omega}_h$:*

$$\hat{\boldsymbol{\beta}}_h = (\boldsymbol{S}'\boldsymbol{\Sigma}_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{\Sigma}_h^{\dagger}\boldsymbol{Y} = (\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{Y}$$

with variance matrix $\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\Omega}_h$. Moreover, this is the minimum variance linear unbiased estimate.

Proof. We write $\boldsymbol{\Sigma}_h = \boldsymbol{BC}$ where $\boldsymbol{B} = \boldsymbol{S}\boldsymbol{\Omega}_h$ and $\boldsymbol{C} = \boldsymbol{S}'$. Then, by Fact 6.4.8 of Bernstein (2005, p. 235), the Moore–Penrose generalized inverse of $\boldsymbol{\Sigma}_h$ is

$$\boldsymbol{\Sigma}_h^{\dagger} = \boldsymbol{C}'(\boldsymbol{CC}')^{-1}(\boldsymbol{B}'\boldsymbol{B})^{-1}\boldsymbol{B}' = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S})^{-1}(\boldsymbol{\Omega}_h'\boldsymbol{S}'\boldsymbol{S}\boldsymbol{\Omega}_h)^{-1}\boldsymbol{\Omega}_h'\boldsymbol{S}'. \tag{9}$$

Then $(\boldsymbol{S}'\boldsymbol{\Sigma}_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{\Sigma}_h^{\dagger} = (\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'$. The variance is obtained by substituting (9) into $(\boldsymbol{S}'\boldsymbol{\Sigma}_h^{\dagger}\boldsymbol{S})^{-1}$.

Tian and Wiens (2006, Theorem 3) show that the GLS estimator will be the minimum variance unbiased estimator if and only if

$$\boldsymbol{S}\boldsymbol{S}^{\dagger}\boldsymbol{\Sigma}_h^{\dagger}\boldsymbol{\Sigma}_h(I - \boldsymbol{S}\boldsymbol{S}^{\dagger}) = \boldsymbol{0},$$

where $\boldsymbol{S}^{\dagger} = (\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'$. Using (9), it is easy to show that this condition holds. □

This remarkable result shows that we can use OLS rather than GLS when computing our revised forecasts, without the need for an estimate of the underlying covariance matrix. Thus it eases the computation especially for a large covariance matrix. That is, we use

$$\tilde{\boldsymbol{Y}}_n(h) = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'\hat{\boldsymbol{Y}}_n(h); \tag{10}$$

and $\boldsymbol{P} = (\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'$. The variance covariance matrix of the revised forecasts is $\mathrm{Var}[\tilde{\boldsymbol{Y}}_n(h)] = \boldsymbol{\Sigma}_h$. Therefore, prediction intervals still require estimation of $\boldsymbol{\Sigma}_h$. We leave the discussion of strategies for estimating this matrix to a later paper.

Eq. (10) shows that, under the assumption $\boldsymbol{\Sigma}_h = \boldsymbol{S}\boldsymbol{\Omega}_h\boldsymbol{S}'$, the optimal combination of base forecasts is independent of the data. For a simple hierarchy with only one level of disaggregation ($K = 1$) and with m_1 nodes at level 1, the weights are given by

$$\boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}' = (m_1+1)^{-1}\begin{pmatrix} m_1 & 1 & 1 & \dots & \dots & 1 \\ 1 & m_1 & -1 & -1 & \dots & -1 \\ 1 & -1 & m_1 & -1 & & -1 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 1 & -1 & & -1 & \ddots & -1 \\ 1 & -1 & -1 & \dots & -1 & m_1 \end{pmatrix}.$$

Similarly, sets of weights can be obtained for any particular hierarchy, independent of the observed data.

Our proposed forecasting method shares some similarities with the approach proposed by Stone et al. (1942), and further studied by Solomou and Weale (1991, 1993, 1996) and Weale (1985, 1988). These authors formulate the hierarchical problem using simultaneous linear equations in the context of balancing the national economic account. The national economic account is disaggregated into production, income and outlay, and capital transactions. Production is further classified into production in Britain and production in the rest of world; income and outlay and capital transactions are each further classified into persons, companies, public corporations, general government, and rest of world. The solution proposed in these papers is equivalent to the GLS estimate discussed here, but using notation that is more appropriate to account balancing than hierarchical forecasting, and without the simplifying assumptions allowing (10) to be used. They also do not consider how to implement the solution when the size of the $\boldsymbol{S}$ matrix is large, as is typical in hierarchical forecasting problems.

5. Computational pitfalls and remedies

The primary difficulty in implementing the forecasting method given by (10) is that the matrix $\boldsymbol{S}$ can be very large, and so the computational effort required to find the inverse of $\boldsymbol{S}'\boldsymbol{S}$ can be prohibitive. Even the construction of the $\boldsymbol{S}$ matrix can be difficult for large hierarchies. We discuss three solutions to this problem.

First, of the $m_K \times m$ elements in the $\boldsymbol{S}$ matrix, there are only $m_K \times K$ non-zero elements. Since K is usually much smaller than m, $\boldsymbol{S}$ is a sparse matrix. We can use sparse matrix storage and arithmetic (e.g., Duff et al., 2002) to save computer memory and computational time. Then we can use the algorithm of Ng and Peyton (1993) as implemented in Koenker and Ng (2010) to solve (10). This works well for moderately large hierarchies, but for some of the large hierarchies we encounter in practice, even this method is unsuitable.

For very large hierarchies, an iterative approach can be used based on the method of Lanczos (1950, 1952). This method was primarily developed for solving a system of linear equations but was subsequently extended by Golub and Kahan (1965) to work with least squares problems with lower bidiagonalization of the coefficient matrix. Based on these methods, Paige and Saunders (1982) developed an iterative algorithm for solving sparse linear least squares problems. Due to the iteration process, the results are an approximation to the direct solution, but in almost all cases in which we have used the method, the difference is negligible.

The third approach relies on a reparameterization of the regression model (7). Rather than each parameter representing a bottom level series, we use a parameter for every series in the hierarchy, but impose some zero-sum constraints to avoid overparameterization. The new parameter vector is denoted by $\phi_n(h)$ and has elements in the same order as $\boldsymbol{Y}_t$. Each parameter measures the contribution of the associated level to the bottom level series below it. In the example shown in Fig. 1,

$$\phi_n(h) = \left[\mu_{\mathrm{T},h}, \mu_{\mathrm{A},h}, \mu_{\mathrm{B},h}, \mu_{\mathrm{C},h}, \mu_{\mathrm{AA},h}, \mu_{\mathrm{AB},h}, \ldots, \mu_{\mathrm{CC},h}, \mu_{\mathrm{AAA},h}, \mu_{\mathrm{AAB},h}, \ldots, \mu_{\mathrm{CCC},h}\right]',$$

and $\beta_{\mathrm{AAB},h} = \mu_{\mathrm{T},h} + \mu_{\mathrm{A},h} + \mu_{\mathrm{AA},h} + \mu_{\mathrm{AAB},h}$. Similarly, other β values are constructed by summing the aggregate nodes that contribute to that bottom level node. Then we can write $\boldsymbol{\beta}_n(h) = \boldsymbol{S}'\phi_n(h)$ so that

$$\hat{\boldsymbol{Y}}_n(h) = \boldsymbol{S}\boldsymbol{S}'\phi_n(h) + \boldsymbol{\varepsilon}_h. \tag{11}$$

Because of the overparameterization, we impose the constraints that, for every split, the sum of the estimated parameters is equal to 0; that is,

$$\sum_i \hat{\mu}_{i,h} = 0, \qquad \sum_j \hat{\mu}_{ij,h} = 0, \qquad \sum_k \hat{\mu}_{ijk,h} = 0, \ldots.$$

This formulation can be viewed as an ANOVA model and the parameters are estimated by least squares. The advantage of this formulation is that the estimators can be re-expressed in the usual ANOVA style, involving summations but no matrix inversion. The approach works for any hierarchy, but the equations are quite complicated when the hierarchy is unbalanced. Consequently, we only provide the estimation equations for balanced hierarchies (i.e., within a given level, each node has an equal number of sub-series).

In order to present these estimators, we need to introduce some new notation. Let X be a node in the hierarchical tree and let m_{X} be the number of series in the sub-tree headed by node X (including X itself). Also let j_{X} be the row of $\hat{\boldsymbol{Y}}_n(h)$ containing $\hat{Y}_{\mathrm{X},h}$ and let ℓ_{X} be the level where node X appears, i.e., $\ell_{\mathrm{X}} \in \{0, 1, \ldots, K\}$. Finally, let $\mathrm{X}^{(j)}$ be the node that is j levels above X, $\mathbf{1}_k$ be a row vector of 0s with 1 in the kth column, and $b_{\mathrm{X}} = \mathrm{diag}[0, \ldots, 0, 1, \ldots, 1]$ where the first $j_{\mathrm{X}} - 1$ diagonal elements are 0 and every element after that is 1. For example, in the hierarchy shown in Fig. 1, $m_{\mathrm{BC}} = 4, j_{\mathrm{BC}} = 10, \ell_{\mathrm{BC}} = 2$, and $\mathrm{BC}^{(1)} = \mathrm{B}$.

Then for a balanced hierarchy, the ANOVA estimators for the parameters are

$$\hat{\mu}_{\mathrm{T},h} = \frac{1}{(\mathbf{1}_1\boldsymbol{S}\boldsymbol{S})'(\mathbf{1}_1\boldsymbol{S}\boldsymbol{S}')'}\left\{\mathbf{1}_1\boldsymbol{S}\boldsymbol{S}'\hat{\boldsymbol{Y}}_n(h)\right\} \tag{12}$$

$$\hat{\mu}_{\mathrm{X},h} = \frac{1}{(\mathbf{1}_{j_{\mathrm{X}}}\boldsymbol{S}\boldsymbol{S}')(\mathbf{1}_{j_{\mathrm{X}}}\boldsymbol{S}\boldsymbol{S}'b_{\mathrm{X}})'}\left\{\mathbf{1}_{j_{\mathrm{X}}}\boldsymbol{S}\boldsymbol{S}'\hat{\boldsymbol{Y}}_n(h) - \sum_{i=1}^{\ell_{\mathrm{X}}}(\mathbf{1}_{j_{\mathrm{X}}}\boldsymbol{S}\boldsymbol{S}')\left(\mathbf{1}_{j_{\mathrm{X}^{(i)}}}\boldsymbol{S}\boldsymbol{S}'b_{\mathrm{X}^{(i)}}\right)'\hat{\mu}_{\mathrm{X}^{(i)},h}\right\}, \tag{13}$$

where $\ell_{\mathrm{X}} \geq 1$.

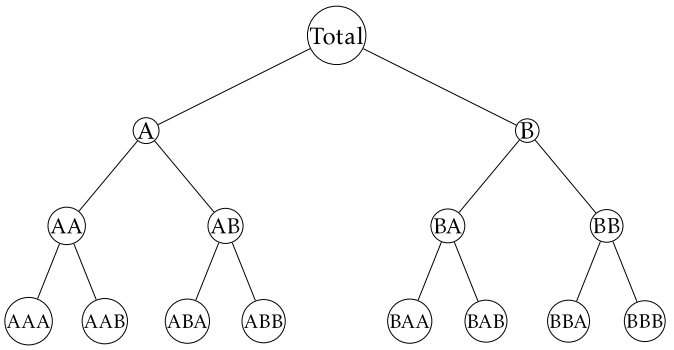

Fig. 2. A 3-level hierarchical tree diagram.

In practice, we would normally use the sparse matrix approach for small to moderate size hierarchies, the ANOVA approach for large balanced hierarchies, and the iterative approach for large unbalanced hierarchies. While it is possible to derive the ANOVA equations for unbalanced hierarchies, it is difficult to do so completely generally, and it is easier to use the iterative approach than derive the ANOVA equations for a specific case.

6. Numerical simulations

In order to evaluate the performance of our proposed methodology, we first perform a simulation study. We consider a hierarchy with $K = 3$ levels and $m = 15$ series in total. The completely aggregated series at the top level (level 0) is disaggregated into $m_1 = 2$ component series at level 1. Each of these series is further subdivided into two series at level 2; i.e., $m_2 = 4$. Finally, each one of the level 2 series is further disaggregated to two component series giving a total of $m_3 = 8$ series at the completely disaggregated bottom level. A diagram of the hierarchical structure is given in Fig. 2.

The data for each series were generated by an ARIMA (p, d, q) process with d taking values 1 or 2 with equal probability, and p and q each taking values 0, 1 and 2 with equal probability. For each generated series, the parameters of each process were chosen randomly based on a uniform distribution over the stationary and invertible regions.

We first generate the disaggregated data for all the series in the bottom level. These are then summed to obtain the aggregated data for the levels above. To allow for relationships between the series in the hierarchy, the simulated ARIMA processes were designed to have contemporaneously correlated errors. The error covariance structure from which the bottom level series are generated is,

$$\begin{bmatrix} 7 & 3 & 2 & 1 & 0 & 0 & 0 & 0 \\ 3 & 7 & 2 & 1 & 0 & 0 & 0 & 0 \\ 2 & 2 & 6 & 3 & 0 & 0 & 0 & 0 \\ 1 & 1 & 3 & 6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 7 & 3 & 2 & 1 \\ 0 & 0 & 0 & 0 & 3 & 7 & 2 & 1 \\ 0 & 0 & 0 & 0 & 2 & 2 & 6 & 3 \\ 0 & 0 & 0 & 0 & 1 & 1 & 3 & 6 \end{bmatrix}.$$

This structure allows for a positive error correlation between bottom level series with the same grandparent (series A or series B), and stronger correlation if they have the same parent (series AA, AB, BA or BB), but no correlation with other bottom level series.

We tried several other covariance structures, including negative correlations and a mixture of positive and negative correlations. These produced qualitatively similar results and we do not include them here in order to save space. Further details of the additional simulations can be obtained upon request from the authors.

For each series we generated 100 observations. We divided the data into two parts: the first 90 observations of each series was used as a training set for estimating parameters and the last 10 observations as a test set. Applying the statistical framework for exponential smoothing methods developed by Hyndman et al. (2002), we fit a state space model to the training data and obtain independent base forecasts for each of the 15 series for 1–10 steps ahead.

Final forecasts are produced using three different methods: (i) optimally combining the base forecasts as advocated in Section 4; (ii) using the conventional bottom-up method (i.e., aggregating the bottom level independent forecasts all the way to the top level); (iii) applying the conventional top-down approach where the historical proportions of the data are used to disaggregate the top level forecasts all the way down to the bottom level. This process was repeated 1000 times. For comparison purposes, we have also included the independent forecasts obtained by treating each time series separately, although these are not aggregate consistent.

In Tables 1 and 2, we present the mean absolute error (MAE) and root mean squared error (RMSE) for each series in the hierarchy. Apart from the bottom level, the optimal combination method consistently achieves the best forecast accuracy,

Table 1
MAE for each level series.

	MAE			
	Independent	Top-down	Bottom-up	Optimal
Total	62.24	62.24	59.66	57.60
A	39.41	144.78	38.91	37.53
B	33.36	144.76	33.16	31.91
AA	24.32	116.17	24.91	23.95
AB	20.56	108.68	21.26	20.25
BA	20.36	106.56	20.84	19.84
BB	18.24	100.88	18.67	17.76
AAA	13.31	66.66	13.31	14.20
AAB	13.35	66.71	13.35	14.27
ABA	11.11	63.39	11.11	12.02
ABB	11.56	61.58	11.56	12.46
BAA	11.29	61.80	11.29	12.13
BAB	10.86	59.20	10.86	11.64
BBA	9.91	57.41	9.91	10.65
BBB	10.08	59.40	10.08	10.88
Average	20.66	85.35	20.59	20.47

Table 2
RMSE for each level series.

	RMSE			
	Independent	Top-down	Bottom-up	Optimal
Total	75.64	75.64	72.88	71.10
A	47.77	149.89	47.33	46.03
B	40.32	149.10	40.20	39.07
AA	29.51	118.93	30.23	29.25
AB	24.93	111.06	25.76	24.72
BA	24.64	108.91	25.21	24.19
BB	22.08	103.01	22.58	21.66
AAA	16.21	68.19	16.21	17.20
AAB	16.24	68.25	16.24	17.26
ABA	13.48	64.61	13.48	14.51
ABB	14.08	62.99	14.08	15.07
BAA	13.70	63.08	13.70	14.64
BAB	13.19	60.45	13.19	14.06
BBA	12.02	58.60	12.02	12.85
BBB	12.25	60.58	12.25	13.14
Average	25.07	88.22	25.02	24.98

followed by the bottom-up method. The bottom-up method performs significantly better than the conventional top-down method. The optimal method is marginally worse than the independent forecasts.

Pairwise t-tests were used for each pair of methods to examine whether the differences in the forecasting results are statistically significant. The p-values for each pair of the four methods are smaller than 0.001 as measured by MAE and RMSE. These results indicate that all four methods differ significantly from each other, even allowing for multiple comparisons.

7. Australian domestic tourism forecasts

We now apply our optimal combination approach to forecasting Australian domestic tourism data. A more detailed analysis of these data was provided in Athanasopoulos et al. (2009). For each domestic tourism demand time series we have quarterly observations on the number of visitor nights which we use as an indicator of tourism activity. The available data covers 1998:Q1–2009:Q4 and were obtained from the National Visitor Survey which is managed by Tourism Research Australia. The data are collected by computer-assisted telephone interviews with approximately 120,000 Australians aged 15 years and over on an annual basis (Tourism Reseach Australia, 2005).

We compare the forecasting performance of our method to the bottom-up and conventional top-down approaches. The structure of the hierarchy is shown in Table 3. At the top level we have aggregate domestic tourism demand for the whole of Australia. At level 1, we split this demand by purpose of travel: Holiday, Visiting friends and relatives, Business, and Other. In the next level the data are disaggregated by the states and territories of Australia: New South Wales, Queensland, Victoria, South Australia, Western Australia, Tasmania, and the Northern Territory. At the bottom level we further disaggregate the data into tourism within the capital city of each state or territory and tourism in other areas. The respective capital cities for the above states and territories are: Sydney, Brisbane (including the Gold Coast), Melbourne, Adelaide, Perth, Hobart and Darwin. The hierarchy is balanced, and so the ANOVA method of combination given by (12) and (13) is applicable.

Table 3
Hierarchy for Australian tourism.

Level	Number of series	Total series per level
Australia	1	1
Purpose of travel	4	4
States and territories	7	28
"Capital city" versus "other"	2	56

Table 4
MAPE for out-of-sample forecasting of the alternative hierarchical approaches applied to Australian tourism data.

	Forecast horizon (h)								
	1	2	3	4	5	6	7	8	Average
Top level: Australia									
Top-down	6.34	5.97	6.34	6.66	**7.16**	**7.03**	**7.06**	**6.93**	**6.69**
Bottom-up	**5.69**	6.00	6.39	**6.60**	7.70	7.77	7.45	7.42	6.88
Optimal	5.99	**5.81**	**6.23**	6.63	7.23	7.26	7.27	7.23	6.71
Level 1: *Purpose of travel*									
Top-down	16.07	15.14	15.56	16.18	16.76	16.09	16.66	17.19	16.21
Bottom-up	**10.96**	11.14	11.60	12.16	13.42	13.30	**13.52**	14.39	12.56
Optimal	11.03	**10.65**	**11.26**	**11.27**	**13.09**	**12.90**	13.73	**13.45**	**12.17**
Level 2: *States and Northern Territory*									
Top-down	40.16	39.41	40.50	40.68	41.68	41.10	42.38	43.22	41.14
Bottom-up	**28.41**	28.80	**29.52**	30.03	**31.64**	**32.81**	**33.18**	**33.91**	**31.04**
Optimal	31.31	**28.73**	30.08	**29.72**	35.53	34.12	36.76	36.32	32.82
Bottom level: Capital city versus other									
Top-down	49.44	48.94	50.33	50.27	51.41	49.80	51.25	52.35	50.48
Bottom-up	**38.60**	**38.99**	**40.25**	40.47	**42.72**	**43.37**	**43.92**	**45.14**	**41.68**
Optimal	44.15	40.11	40.57	**37.86**	50.68	46.78	50.73	50.93	45.22

For each series, we select an optimal ARIMA model using the automatic algorithm of Hyndman and Khandakar (2008). We then re-estimate the parameters of the model using a rolling window beginning with the model fitted using the first 24 observations (1998:Q1–2003:Q4). Forecasts from the fitted model are produced for up to 8 steps ahead. We iterate this process, increasing the sample size by one observation until 2009:Q3. This process produces 24 one-step-ahead forecasts, 23 two-step-ahead forecasts, and up to 17 eight-step-ahead forecasts. We use these to evaluate the out-of-sample forecast performance of each of the hierarchical methods we consider.

Table 4 contains the mean absolute percentage error (MAPE) for each forecast horizon yielded by our proposed optimal combination approach and the conventional top-down and bottom-up approaches. The bold entries identify the approach that performs best for the corresponding level and forecast horizon, based on the smallest MAPE. The last column contains the average MAPE across all forecast horizons.

From this empirical study, we find that the only level for which the top-down approach is best is the top level. As we move down the hierarchy our optimal combination approach and the bottom-up approach clearly outperform the top-down method with our optimal combination method performing best at level 1 and the bottom-up approach performing best at levels 2 and 3. The good performance of the bottom-up approach can be attributed to the fact that the data have strong seasonality and trends, even at the bottom level. With more noisy data, it is not so easy to detect the signal at the bottom level, and so the bottom-up approach would not do so well.

To formally test whether the forecasting performance of the alternative methods is different, we perform pairwise t-tests. The top-down method is significantly different from each of the other two methods ($p < 10^{-6}$ for each test); the bottom-up and optimal combination methods are not significantly different ($p = 0.655$).

8. Conclusions and discussion

We have proposed a new statistical method for forecasting hierarchical time series, which allows optimal point forecasts to be produced that are reconciled across the levels of a hierarchy. One useful feature of our method is that the base forecasts can come from any model, or can be judgmental forecasts. So ad hoc adjustments can be incorporated into the forecasts, information about individual series can be allowed for, and important covariates can be included. The method places no restrictions on how the original (base) forecasts are created. The procedure simply provides a means of optimally reconciling the forecasts so they aggregate appropriately across the hierarchy.

A remarkable feature of our results is that the point forecasts are independent of the correlations between the series. While this may initially seem counter-intuitive, it is a natural consequence of assuming that the forecast errors across the hierarchy aggregate in the same way that the observed data aggregate, which is a reasonable approximation to reality. Of

course, the forecast variances will depend crucially on the correlations among series, and so the covariance matrix is still required in order to produce prediction intervals. We have not discussed the computation of the covariance matrix in this paper, as we will address that in a future paper.

Another surprising result is that the optimal combination weights depend only on the hierarchical structure and not on the observed data. Again, this arises from the assumption concerning the aggregation of forecast errors. Because of this result, it is possible to determine the combination weights once only, and then apply them as each new set of observations are available. This saves a great deal of computational time.

Our simulations and empirical example demonstrate that the optimal combination method proposed in this paper outperforms existing methods for forecasting hierarchical data. So we recommend that it be adopted for routine use in business and industry, whenever hierarchical time series data need to be forecasted.

Acknowledgements

We would like to thank Blyth Archibald for some helpful comments on how the simulated data should be constructed, and the editor, associate editor and two referees for their insightful comments which improved the paper. Athanasopoulos acknowledges support from the Australian Research Council grant DP0984399.

References

Athanasopoulos, G., Ahmed, R.A., Hyndman, R.J., 2009. Hierarchical forecasts for Australian domestic tourism. International Journal of Forecasting 25 (1), 146–166.
Bernstein, D.S., 2005. Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory. Princeton University Press, Princeton, New Jersey.
Collins, D.W., 1976. Predicting earnings with sub-entity data: some further evidence. Journal of Accounting Research 14 (1), 163–177.
Dangerfield, B.J., Morris, J.S., 1992. Top–down or bottom–up: aggregate versus disaggregate extrapolations. International Journal of Forecasting 8 (2), 233–241.
Duff, I.S., Heroux, M.A., Pozo, R., 2002. An overview of the sparse basic linear algebra subroutines: the new standard from the BLAS technical forum. ACM Transactions on Mathematical Software 28 (2), 239–267.
Dunn, D.M., Williams, W.H., DeChaine, T.L., 1976. Aggregate versus subaggregate models in local area forecasting. Journal of the American Statistical Association 71 (353), 68–71.
Edwards, J.B., Orcutt, G.H., 1969. Should aggregation prior to estimation be the rule? The Review of Economics and Statistics 51 (4), 409–420.
Espasa, A., Senra, E., Albacete, R., 2002. Forecasting inflation in the European monetary union: a disaggregated approach by countries and by sectors. European Journal of Finance 8 (4), 402–421.
Fair, R.C., Shiller, R.J., 1990. Comparing information in forecasts from econometric models. The American Economic Review 80 (3), 375–389.
Fliedner, E.B., Lawrence, B., 1995. Forecasting system parent group formation: an empirical application of cluster analysis. Journal of Operations Management 12 (2), 119–130.
Fliedner, G., 1999. An investigation of aggregate variable time series forecast strategies with specific subaggregate time series statistical correlation. Computers and Operations Research 26 (10–11), 1133–1149.
Fliedner, G., 2001. Hierarchical forecasting: issues and use guidelines. Industrial Management and Data Systems 101 (1), 5–12.
Fogarty, D.W., Blackstone, J.H., Hoffman, T.R., 1990. Production and Inventory Management, 2nd ed. South-Western Publication Co., Cincinnati, OH.
Golub, G., Kahan, W., 1965. Calculating the singular values and pseudo-inverse of a matrix. Journal of the Society for Industrial and Applied Mathematics, Series B: Numerical Analysis 2 (2), 205–224.
Gross, C.W., Sohl, J.E., 1990. Disaggregation methods to expedite product line forecasting. Journal of Forecasting 9 (3), 233–254.
Grunfeld, Y., Griliches, Z., 1960. Is aggregation necessarily bad? The Review of Economics and Statistics 42 (1), 1–13.
Hubrich, K., 2005. Forecasting euro area inflation: does aggregating forecasts by HICP component improve forecast accuracy? International Journal of Forecasting 21 (1), 119–136.
Hyndman, R.J., Khandakar, Y., 2008. Automatic time series forecasting: the forecast package for R. Journal of Statistical Software 26 (3).
Hyndman, R.J., Koehler, A.B., Snyder, R.D., Grose, S., 2002. A state space framework for automatic forecasting using exponential smoothing methods. International Journal of Forecasting 18 (3), 439–454.
Kahn, K.B., 1998. Revisiting top–down versus bottom–up forecasting. The Journal of Business Forecasting 17 (2), 14–19.
Kinney Jr., W.R., 1971. Predicting earnings: entity versus subentity data. Journal of Accounting Research 9 (1), 127–136.
Koenker, R., Ng, P., 2010. SparseM: Sparse linear algebra. R package version 0.85. http://CRAN.R-project.org/package=SparseM.
Kohn, R., 1982. When is an aggregate of a time series efficiently forecast by its past? Journal of Econometrics 18 (3), 337–349.
Lanczos, C., 1950. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. Journal of Research of the National Bureau of Standards 45 (4), 255–282.
Lanczos, C., 1952. Solution of system of linear equations by minimized iteration. Journal of Research of the National Bureau of Standards 49 (1), 33–53.
Marcellino, M., Stock, J.H., Watson, M.W., 2003. Macroeconomic forecasting in the Euro area: country specific versus area-wide information. European Economic Review 47 (1), 1–18.
Narasimhan, S.L., McLeavey, D.W., Billington, P.J., 1995. Production Planning and Inventory Control, 2nd ed. Prentice Hall, Englewood Cliffs, New Jersey.
Ng, E.G., Peyton, B.W., 1993. Block sparse Cholesky algorithms on advanced uniprocessor computers. SIAM Journal on Scientific Computing 14 (5), 1034–1056.
Orcutt, G.H., Watts, H.W., Edwards, J.B., 1968. Data aggregation and information loss. The American Economic Review 58 (4), 773–787.
Paige, C.C., Saunders, M.A., 1982. Algorithm 583 LSQR: sparse linear equations and least squares problems. ACM Transactions on Mathematical Software 8 (2), 195–209.
Schwarzkopf, A.B., Tersine, R.J., Morris, J.S., 1988. Top–down versus bottom–up forecasting strategies. International Journal of Production Research 26 (11), 1833–1843.
Shing, N.K., 1993. A study of bottom–up and top–down forecasting methods. M.Sc. Thesis. Royal Melbourne Institute of Technology.
Shlifer, E., Wolff, R.W., 1979. Aggregation and proration in forecasting. Management Science 25 (6), 594–603.
Solomou, S., Weale, M., 1991. Balanced estimates of UK GDP 1870-1913. Explorations in Economic History 28 (1), 54–63.
Solomou, S., Weale, M., 1993. Balanced estimates of national accounts when measurement errors are autocorrelated: the UK, 1920-38. Journal of the Royal Statistical Society. Series A 156 (1), 89–105.
Solomou, S., Weale, M., 1996. UK national income, 1920-1938: the implications of balanced estimates. The Economic History Review 49 (1), 101–115.
Stone, R., Champernowne, D.G., Meade, J.E., 1942. The precision of national income estimates. The Review of Economic Studies 9 (2), 111–125.

Tian, Y., Wiens, D.P., 2006. On equality and proportionality of ordinary least squares, weighted least squares and best linear unbiased estimators in the general linear model. Statistics and Probability Letters 76 (12), 1265–1272.
Tiao, G.C., Guttman, I., 1980. Forecasting contemporal aggregates of multiple time series. Journal of Econometrics 12 (2), 219–230.
Tourism Reseach Australia, 2005. Travel by Australians. September Quarter 2005, Tourism Australia. Canberra.
Weale, M., 1985. Testing linear hypothesis on national account data. The Review of Economics and Statistics 67 (4), 685–689.
Weale, M., 1988. The reconciliation of values, volumes and prices in the national accounts. Journal of the Royal Statistical Society. Series A 151 (1), 211–221.
Zellner, A., Tobias, J., 2000. A note on aggregation, disaggregation and forecasting performance. Journal of Forecasting 19 (5), 457–469.