Hierarchical Time Series Forecasting

by

M.S., Haoxi Ma

University of Connecticut, 2020

Project Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Statistics

in the

Department of Statistics

Advisor: Prof. Nalini Ravishanker

Autumn 2020

Content

Review paper: Forecast reconciliation in hierarchical time series

ABSTRACT

To forecast hierarchical time series that adhere to known linear constraints, the challenge is that we require forecasts that are coherent across the aggregation structure. Based on the matrix algebra, several methods have been developed. Among them, optimal reconciliation is particularly prominent compared to bottom-up or top-down methodologies. A geometric interpretation of so-called reconciliation facilitates the derivation of novel results that explain why and how reconciliation via projection is guaranteed to improve forecast accuracy with respect to a specific class of loss functions and results in unbiased forecasts provided the initial base forecasts are also unbiased. For dealing with biased base forecasts, according to an empirical study, doing bias-correcting before reconciliation is shown to outperform alternatives that only bias-correct or only reconcile forecasts

Introduction

Time series can often be naturally disaggregated by various attributes of interest. Lower levels are nested within the higher levels, and so the collection of time series follows a hierarchical aggregation structure called hierarchical time series. For example, the daily new cases of covid-19 in America can be disaggregated by state, within each state by county.

The past decade has seen rapid development in methodologies for forecasting this kind of time series involving two steps: first separate forecasts are produced for all series, then these are adjusted ex post to ensure coherence with aggregation constraints. These methods mainly consist of bottom-up, top-down, middle-out and optimal reconciliation.

A geometric interpretation and some related pretty nice properties of the last method has been proposed in Anastasios Panagiotelis's paper. Besides making a briefly review of Anastasios's accomplishments, we are going to demonstrate the thoughts of forecasting hierarchical time series logically combine with some matrix algebra.

The remainder of this paper is structured as follows. Section 2 deals with the concepts of hierarchical time series and coherence. Section 3 shows how reconciliation maps base forecasts $\hat{y}_h$ to reconciled forecast $\tilde{y}_h$. In Section 4, we demonstrate some approaches to forecast hierarchical time series with linear constraints. Then, we summarize the unbiasedness preserving property of reconciliation via certain project matrices and the way to do bias correction. At last, some discussions and thoughts are concluded in Section 6.

Preliminaries

Briefly, the hierarchical structure time series can be represented as the formula:

$$y_t = Sb_t$$

Where

$$y_t \in R^n \ a\ vector\ comprises\ observations\ of\ all\ variables\ at\ time\ t$$

$$S\ encodes\ the\ aggregation\ constraints$$

$$b_t \in R^m \ a\ vector\ comprises\ observations\ of\ bottom - level\ variables\ at\ time\ t$$

We will use the example in Anastasios's paper through this whole paper.

According to the formula, using the theory in algebra, the forecasts of y_t should always belongs to the column space of S. This leads to the definition of coherence: If the vector of point forecasts of all series, e.g. $\check{y}_{t+h}$, still lies in the column space of S, we say it's coherent.

However, in practice, we usually forecast all series independently, ignoring the linear constraints which will make forecasting vector out of the $span(S)$. Here, the mapping matrices were generated to deal with this problem.

Mapping matrices

As mentioned before, suppose we forecast all series independently regardless of the linear constraints. Then, we can get the base forecasts and denote them by $\hat{y}_h$ where h is the forecast horizon. They are stacked in the same order as the y_t.

Then, the forecast reconciliation can be represented as

$$\tilde{y}_h = SG\hat{y}_h$$

Where

$$SG\ called\ "projection"\ or\ "reconciliation"\ matrix$$

This is the crucial formula. Choosing different G will generate different reconciling approaches.

Approaches

In this part, we will first talk about two simple and intuitive methods: bottom-up and top-down approaches. Then, we are going to introduce the optimal reconciliation approach which is more complicated but accurate.

4.1 Bottom-up and top-down

The thought of these two methods is that first we can map the base forecasts into the bottom-level time series, then sum these up using the aggregation structure to produce a set of coherent forecasts.

The G matrix is defined corresponding to the approach implemented. For example, if the bottom-up approach is used to forecast the hierarchy of the example, then

$$G = [0_{7\times4}\ I_7\ ]$$

If the top-down approach is used, then

$$G = [P_7, 0_{7\times10}]$$

Where

$P_7 = (p_1, p_2, \ldots\ldots p_7)^T$ be a set of disaggregation proportions which dictate how the forecasts of the Total series are to be distributed to obtain forecasts for each series at the bottom-level of the structure. (For more detail, refer to the book Forecasting: Principles and Practice authored by Rob.J.Hyndman.)

These two methods have been based on forecasts from a single level of the aggregation structure. However, in general, we want to combine and reconcile all the base forecasts in order to produce the most accurate reconciled forecasts.

4.2 Optimal reconciliation approach

Optimal forecast reconciliation will occur if we can find the G matrix which minimizes the forecast error of the set of coherent forecasts:

$$G = (S^T W_h^{-1} S)^{-1} S^T W_h^{-1}$$

Where

$$W_h \text{ is the covariance matrix of the corresponding base forecast errors}$$

Then

$$\tilde{y}_h = S(S^T W_h^{-1} S)^{-1} S^T W_h^{-1} \hat{y}_h$$

We refer to this as the Minimum Trace estimator.

To use this in practice, there are two kind of "projection" when we provide simplifying approximations of W_h.

- Orthogonal projection (OLS and MLS):

In Euclidean distance, figure 1 clearly shows that $\tilde{y}_h$, $\hat{y}_h$ and y_h form a right-angled triangle with $\tilde{y}_h$ at the right- angled vertex. In this triangle the line between y_h and $\hat{y}_h$ is the hypotenuse and therefore must be longer than the distance between y_h and $\tilde{y}_h$. Therefore, reconciliation is guaranteed to reduce the squared error of the forecast.

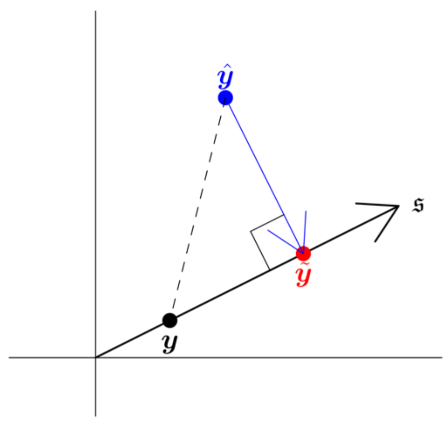


Figure 1: Orthogonal Projection

However, this measure is the square root of the sum of squared forecast errors of all variables in the hierarchy. Consequently, while forecast improvement is guaranteed for the hierarchy overall, reconciliation can lead to less accurate forecasts for individual series. Second, although orthogonal projections are guaranteed to improve on base forecasts, they are not necessarily the projection that leads to the greatest improvement in forecast accuracy. Therefore, we turn our attention to oblique projection.

- Oblique projection (MinT):

If we consider the objective function to be some weighted sum of squared errors, or a Mahalanobis distance, then the oblique projection matrix is guaranteed to improve forecast accuracy over base forecasts, for an appropriately selected W_h. Figure 2 below shows the geometric interpretation of improving the performance compared to the orthogonal projection.

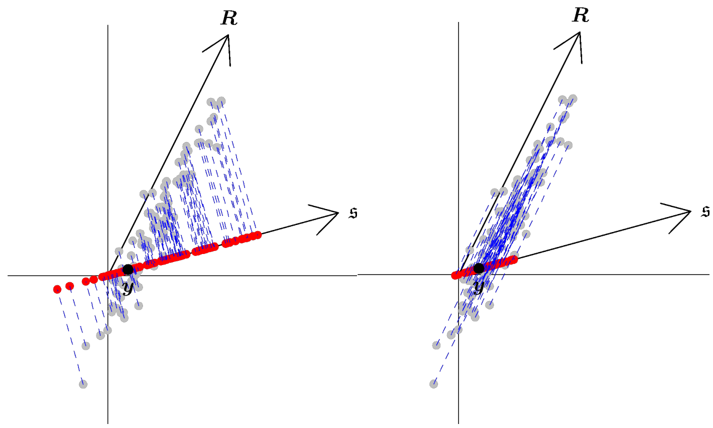


Figure 2: Orthogonal Projection and Oblique Projections

Bias in forecast reconciliation

In general, when doing forecasting, we want to make sure we have unbiased forecasts. If the base forecasts are unbiased, then the coherent forecasts will be unbiased provided $SGS = S$, e.g. SG is a projection onto the column space of S. It is easy to prove that the optimal reconciliation method satisfies this constraint while all top-down methods do not.

In actual cases, if the in-sample errors are centered around the origin, it would be expected from an unbiased forecast. If not, we should do some transformation until they are centered at the origin before doing reconciliation which is shown to outperform alternatives that only bias-correct or only reconcile forecasts.

Discussion

All forecast reconciliation approaches have their advantages and disadvantages. When dealing with an actual case, we need to decide which model to use for fitting time series, how to do forecasting reconciliation, whether to do bias-correcting and so on. Drawing some time series plots will help with these stuffs. In a word, try different models and choose the "best" one.

Lastly, besides the possibilities of future research mentioned in Anastasios's paper, we want to argue that it should be possible to construct a statistical test for testing whether the base forecast is biased. What's more, paying more attention to the situation where some nodes have different depths will help improve the practical performance.

Empirical Study

Introduction

After the review part, we are familiar with the hierarchical time series. Now, an empirical study has been given to show more details. Hierarchical time series often arise due to geographic divisions. There is a really vivid example around us: the daily new cases of COVID-19 in America can be disaggregated by state, then within each state by county. In this case study, we will construct the "best" model to forecast the daily new cases. The data are provided by Johns Hopkins Whiting School of Engineering and the link is https://github.com/CSSEGISandData/COVID-19.

This dataset contains daily accumulative confirmed cases updated once a day around 23:59 (UTC). We choose to use the latest one. According to a natural geographical hierarchy, we disaggregate total daily new cases into 52 states and 3334 counties which means that there are 3387 series across the hierarchy with 3334 bottom-level series. The data span the period 01/22/2020 to 12/16/2020, which gives a total of 329 observations per series.

In order to acknowledge the properties of the time series intuitively, we draw some plots.

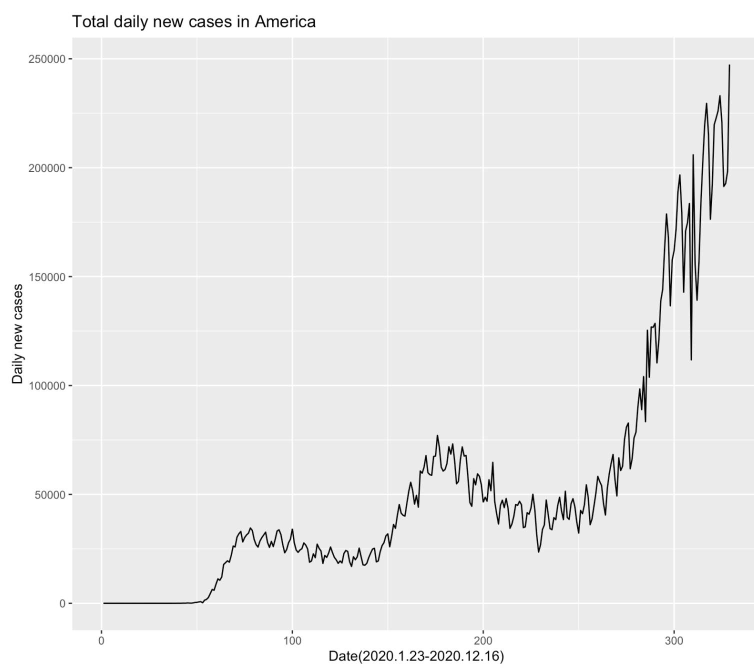


Figure 1: Time series plot of the overall daily new cases in America

From the figure 1, we know that the top-level time series has an increasing trend with some fluctuation but no obvious seasonality can be observed. Then, we plot the new cases by different states.

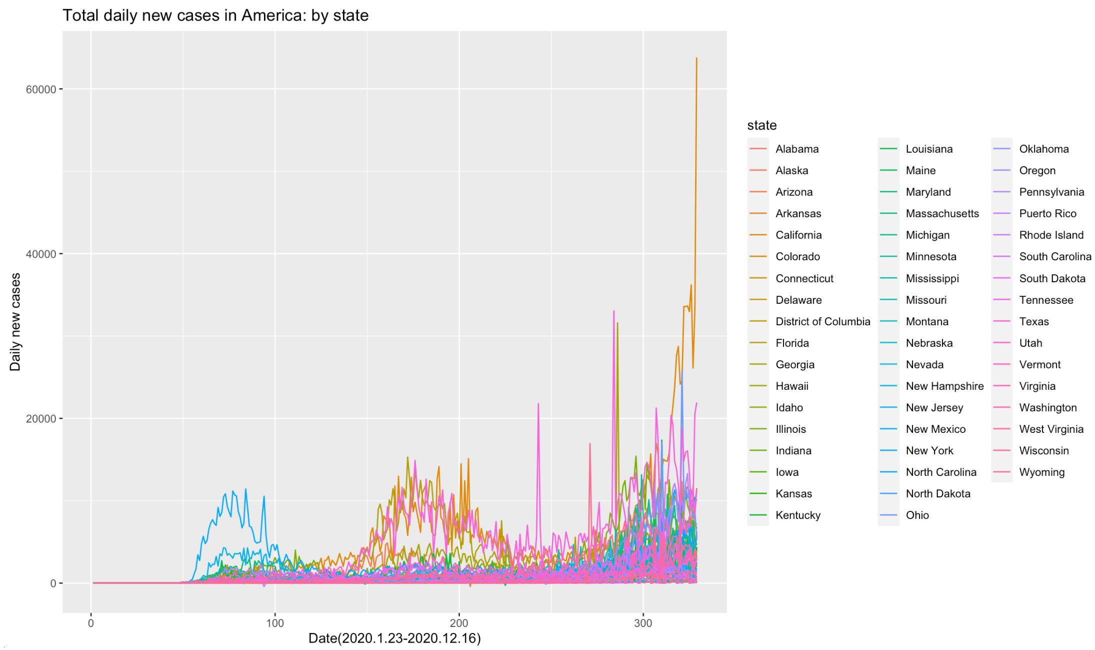


Figure 2: Time series plot of daily new cases by state

According to figure 2, the cases in some states increase by a large margin while some keep low at all times. And some of them have similar patterns while some of them are really special. After that, we draw a plot for Connecticut by county.

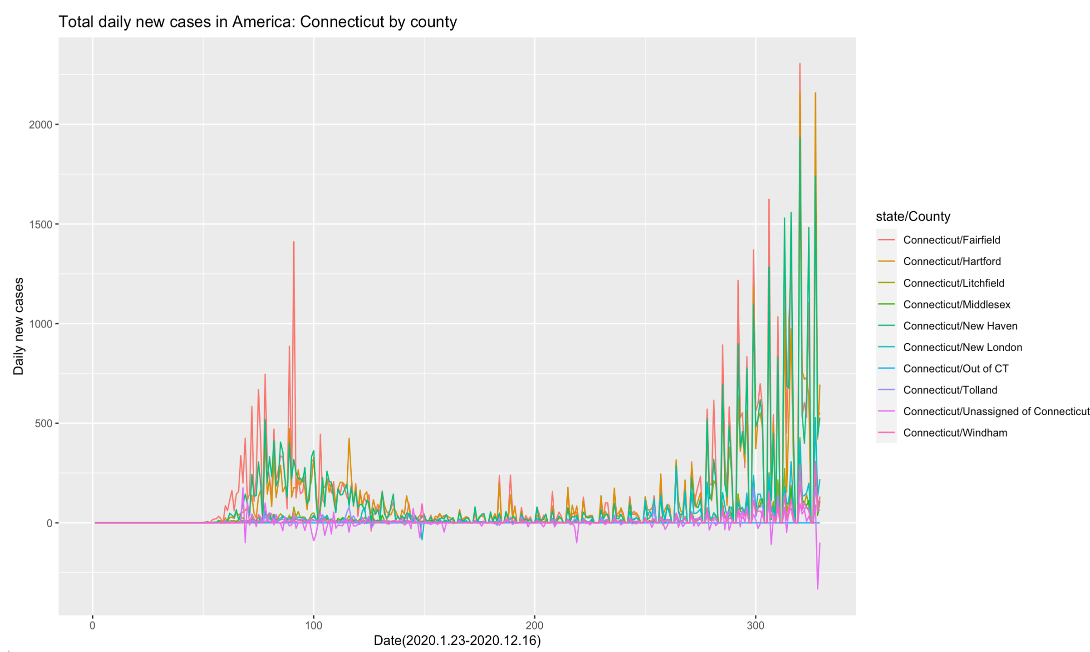


Figure 3: Time series plot of daily new cases of Connecticut by county

Here, we can find that most of them have a similar pattern: controlled fairly well at first, but broke out at last. The time series plot of the unassigned place in Connecticut looks a little bit strange-smaller than zero at some points. We think this may result from the recording methods of unassigned places so also take them into account. At last, we would like to plot the data of the nearest city Hartford.

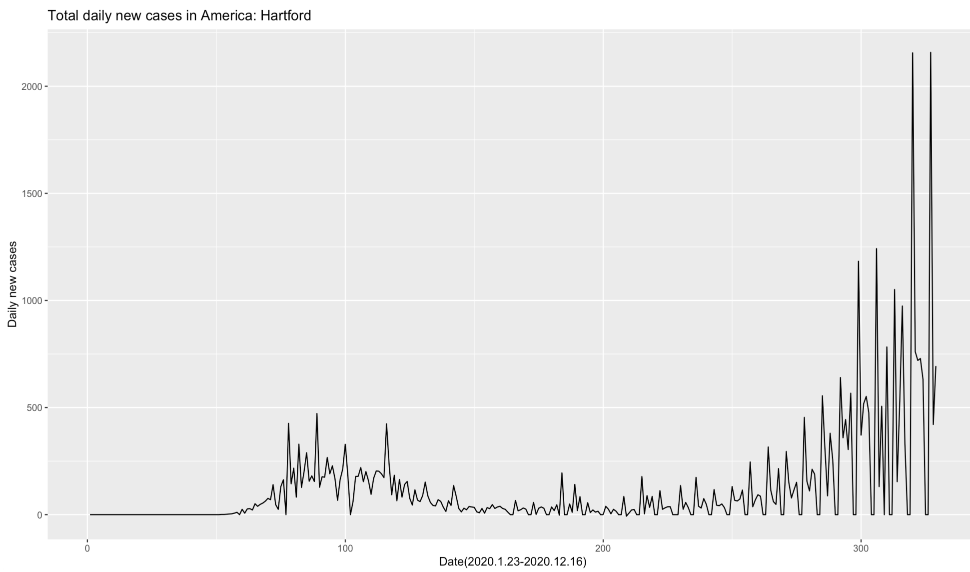


Figure 4: Time series plot of daily new cases of Hartford

Again, looks good at first but out of control recently.

All of these plots indicate that just forecasting from a single level of the aggregation structure is not enough. Instead, we need to combine and reconcile all the base forecasts in order to produce the most accurate reconciled forecasts. Therefore, fitting a single separate model for total new cases, bottom-up and top-down would not perform as well as optimal reconciliation.

Goal of the analysis

Our main goal is to construct the "best" time-series model to forecast the daily new cases of COVID-19 in America which includes deciding which model to use when fitting each time series and using which method for reconciling.

Comprehensive data analysis

3.1 Data preprocessing

In this part, we mainly use the hts and $forecast$ packages to fit our model. Because the build-in function in hts can not deal with the hierarchical time series with different node's depths, we delete the state not having county (there are just a few of them which will not influence the forecasting a lot). Besides, there are some constant columns in the dataset which means it is impossible to estimate covariances when we perform MinT reconciliation. Therefore, we need to add some small white noise (here we choose variance equal to 0.01) to the constant column. Then, when we apply the MinT reconciliation, we use a shrinkage estimator which shrinks the sample covariance to a diagonal matrix because the number of bottom-level series is large compared to the length of the series. After finishing all adjustments, split the hierarchical structured data into train and test parts with 300/29.

3.2 Models

A single separate ARIMA model has been constructed for total new cases. For fitting each time series within the hierarchical time series, we can choose ARIMA or ETS model in the function $forecast :: forecast.gts()$. Then, we apply the bottom-up and top-down methods first. After that, try 4 different reconciliation, e.g. OLS, WLS, structural scaling and MinT. Finally, we can select the "best" model based on MAPE criteria.

3.3 Results

Table 1 shows the MAPE of the single separate ARIMA model of the overall daily new cases. Compared to many models in Table 2, its performance is not good enough.

Model	MAPE
ARIMA	16.457

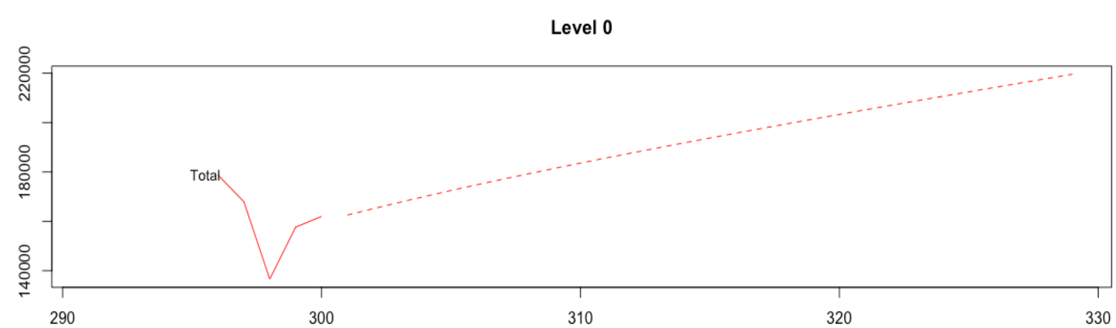
Table 1: ARIMA model of the total daily new cases

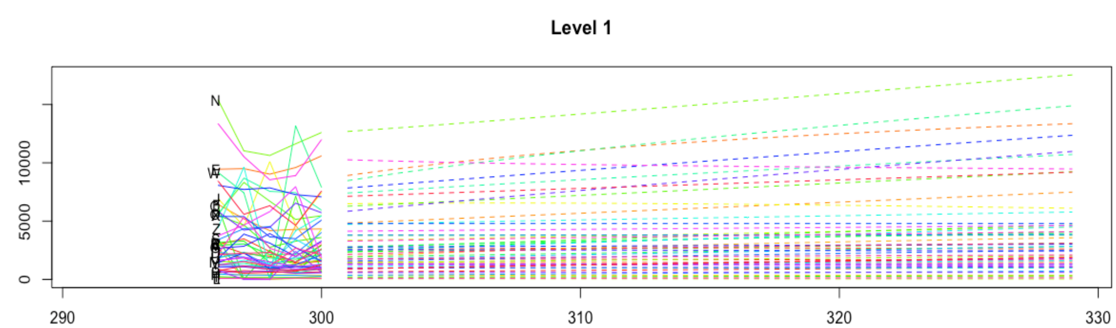
According to Table 2, by selecting the model with smallest MAPE, we choose to use the ETS based on MinT reconciliation hierarchical time series model.

Models		MAPE
Bottom-up	ARIMA	16.686
	ETS	12.281
Top-down	ARIMA	12.304
	ETS	18.973
OLS	ARIMA	12.307
	ETS	18.691
WLS	ARIMA	15.788
	ETS	12.251
Structural scaling	ARIMA	14.106
	ETS	12.146
MinT	ARIMA	12.905
	ETS	12.207

Table 2: Forecasting Reconciliation with ARIMA and ETS

We also plot the forecasting plots of the "best" model (just show level 0 and level 1):





Discussion

In this case study, in order to forecast the daily new cases of COVID-19 in America, we construct a hierarchical time series. Then, because there are many methodologies to figure out this problem, we need to select the "best" one based on the MAPE criteria. After exploratory data analysis, we decide to use the MinT reconciliation with each time series fitted by the ETS model.

From the model, the forecasts of American new cases from 12/17/2020 to 12/20/2020 are 162519.5, 165130.4, 167654.7 and 170102.3 which indicates the worst situation. Therefore, the government should take more effort to control the epidemic and people need to raise more awareness of prevention.

For deeper study, we should do some transformations to make the forecast unbiased before reconciliation because the MPE of the selected model is -3 not very close to 0. Then, which transformation and retransformation should do is an important part to identify. And how to use the state that does not have a county is also a good point to figure out.

In the future, go and forecast the global daily new cases of COVID-19. In this situation, if some countries do not have the corresponding level to the state, we can construct some structure by ourselves. For example, if some of European countries lack the state, we combine them as the European Union, then, the countries have the same levels as the states in America.