Temporal hierarchies with autocorrelation for load forecasting

Peter Nystrup[a,*], Erik Lindström[b], Pierre Pinson[c], and Henrik Madsen[a]

December 23, 2018

[a] *Department of Applied Mathematics and Computer Science, Technical University of Denmark*
[b] *Centre for Mathematical Sciences, Lund University, Sweden*
[c] *Department of Electrical Engineering, Technical University of Denmark*

Abstract

We propose three different estimators that take into account the autocorrelation structure when reconciling forecasts in a temporal hierarchy. Combining forecasts from multiple temporal aggregation levels exploits information differences and mitigates model uncertainty, while reconciliation ensures a unified prediction that supports aligned decisions at different horizons. In previous studies, weights assigned to the forecasts were given by the structure of the hierarchy or the forecast error variances without considering potential autocorrelation in the forecast errors. Our first estimator considers the autocovariance matrix within each aggregation level. Since this can be difficult to estimate, we propose a second estimator that blends autocorrelation and variance information, but only requires estimation of the first-order autocorrelation coefficient at each aggregation level. Our third estimator facilitates information sharing between aggregation levels using a sparse representation of the inverse autocorrelation matrix. We demonstrate the usefulness of the proposed estimators through an application to short-term electricity load forecasting in different price areas in Sweden. We find that by taking account of the autocovariance when reconciling forecasts, accuracy can be significantly improved uniformly across all frequencies and areas.

Keywords: Forecasting; Forecast combination; Temporal Aggregation; Autocorrelation; Reconciliation.

1 Introduction

Temporal aggregation has been studied since the seminal work by Amemiya and Wu (1972) and Tiao (1972) (see Silvestrini and Veredas, 2008, for a literature review). Different temporal aggregations can reveal important information about the underlying data-generating process. When temporal aggregation is applied to a time series, it can strengthen or attenuate different features.

*Correspondence to: Peter Nystrup, Department of Applied Mathematics and Computer Science, Technical University of Denmark, Asmussens Allé, Building 303B, 2800 Kgs. Lyngby, Denmark.
Email: pnys@dtu.dk

Nonoverlapping temporal aggregation is a filter of high-frequency components. At an aggregate view, low frequency components, such as trend and cycle, will dominate. The opposite is true for disaggregate data, where short-term seasonality may be visible. Hence, temporal aggregation can be seen as a tool to better understand and model the data at hand.

Kourentzes et al. (2014) and Petropoulos and Kourentzes (2015) showed that combining forecasts from multiple aggregation levels can lead to improvements in forecast accuracy, while overcoming the need to select a single optimal level. The greatest improvements typically occur at the highest level of a hierarchy, where information from lower levels—i.e., higher-resolution data—is aggregated up (Athanasopoulos et al., 2011; Rostami-Tabar et al., 2013; Kourentzes et al., 2014; Athanasopoulos et al., 2017).

Temporal hierarchies for forecasting, as introduced by Athanasopoulos et al. (2017), can be constructed for any time series by means of nonoverlapping temporal aggregation. Rather than attempting to build one complex model that captures all temporal attributes (see, e.g., Ghysels et al., 2004; Livera et al., 2011; Nystrup et al., 2017; Sedoc et al., 2018), using temporal hierarchies forecasts for different horizons can be made with different (simple) methods. Since these forecasts are produced by different approaches and are based on different information sets, they will most likely not be consistent. This inconsistency can lead to decisions that are not aligned. Optimal decision making requires coherent forecasts; thus, reconciliation is necessary. In the framework proposed by Athanasopoulos et al. (2017), forecasts constructed at different aggregation levels can be combined to yield temporally reconciled, accurate, and robust forecasts, independently of forecasting models.

1.1 Related work

Traditionally, either top-down or bottom-up approaches are used to produce forecasts for a hierarchy. According to the former, forecasts are generated for the time series at the top level and then disaggregated down all the way to the bottom level, while for the latter forecasts are generated at the very bottom level and then aggregated up. Such approaches ensure that forecasts add up across a hierarchy. The advantages and disadvantages of these traditional approaches are not complementary. The top-down approach requires forecasts for only one time series at the very aggregate level; however, aggregation implies a large loss of information, and it is challenging to disaggregate the forecasts down the hierarchy (see Gross and Sohl, 1990; Athanasopoulos et al., 2009, for a summary of top-down approaches). By contrast, bottom up implies no loss of information; but it requires many and possibly very noisy time series to be forecast.

Hyndman et al. (2011) formulated the forecast reconciliation problem for a structural hierarchy as a linear regression model. In order to circumvent the problem of estimating the covariance of the base forecasts, they proposed the use of the ordinary least-squares (OLS) estimator for computing the reconciled forecasts. Hyndman et al. (2016) suggested using weighted least squares (WLS), taking account of the variances on the diagonal of the covariance matrix but ignoring the offdiagonal covariance elements. Later, Wickramasuriya et al. (2018) considered the generalized least-squares (GLS) estimator and found the incorporation of correlation information into the reconciliation procedure to be beneficial for forecast accuracy, when combined with a simple Stein-type shrinkage estimator (Ledoit and Wolf, 2004).

Van Erven and Cugliari (2015) proposed a game-theoretically optimal reconciliation method that guarantees that the total weighted quadratic error of the reconciled forecasts is never greater than that of the base forecasts. Their approach is fundamentally different in that they formulate the forecast reconciliation problem as an optimization problem rather than a regression model. Their formulation included only a diagonal weight matrix and, thus, no information sharing between the base forecasts. Van Erven and Cugliari (2015) proved that the reconciled forecasts are guaranteed to be at least as good as the base forecasts for any loss function that is based on a Bregman divergence. Wickramasuriya et al. (2018) reiterated the proof for their reconciliation method.

A crucial reason for this improvement is that the implied combination mitigates model uncertainty. Forecast combination is widely regarded as beneficial, leading to a reduction of forecast error variance (see, e.g., Clemen, 1989; Timmermann, 2006; Hall and Mitchell, 2007). Ways to best combine forecasts have been widely investigated, resulting in various sophisticated weighting methods; yet, simple approaches, such as the unweighted average, are often found to perform as well as more sophisticated methods (Timmermann, 2006).

Taieb et al. (2017a,b) considered reconciliation of density forecasts, as opposed to point forecasts. They introduced an algorithm that does not require distributional assumptions and imposes dependencies between forecast distributions using samples from the empirical copulas. A promising, although computationally demanding approach. Following van Erven and Cugliari (2015), they used the reconciliation approach to produce load forecasts for individual electricity consumers at the bottom to the total grid at the top of a structural hierarchy.

Temporally aggregated time series can be represented as a hierarchical time series. As a consequence, Athanasopoulos et al. (2017) showed that it is possible to use the reconciliation framework proposed by Hyndman et al. (2011) to produce coherent forecasts. They considered three diagonal estimators as approximations to the sample covariance matrix. By construction, these estimators share the property that they ignore any autocorrelation in the forecast errors. In most cases the simplest of the three, which is based purely on the structure of the hierarchy and requires no estimation of forecast errors, performed as well as the more complicated ones.

The work of Athanasopoulos et al. (2017) was extended by Taieb (2017), who considered load forecasting for individual households. He introduced regularization terms in order to obtain sparse and smooth adjustments that satisfy the aggregation constraints and minimize forecast errors. Imposing sparsity means that some base forecasts remain unaffected by the adjustments. Smoothness provides additional regularization by exploiting the fact that adjustments are applied to consecutive observations of a time series. The reconciled forecasts can be found by solving a sparse fused LASSO (least absolute shrinkage and selection operator) problem (Tibshirani et al., 2005).

In two successive articles, Yang et al. (2017a,b) applied a structural and temporal hierarchy, respectively, for reconciling solar-power forecasts. In the structural case, they followed Wickramasuriya et al. (2018) by considering both diagonal and nondiagonal, regularized and nonregularized covariance estimators. The largest forecast accuracy improvements occurred when including correlations in the reconciliation process. In the temporal case, they followed Athanasopoulos et al. (2017) by considering only diagonal estimators, thus disregarding potential information in the autocorrelation structure. In another case study, Zhang and Dong (2018) documented the benefit from taking into account correlations when reconciling short-term wind-power forecasts across several wind farms in a structural hierarchy.

1.2 Contribution

We consider a general and flexible formulation of the forecast reconciliation problem based on convex optimization. Given the well-documented benefits to incorporating correlation information when reconciling forecasts in a structural hierarchy (Yang et al., 2017a; Wickramasuriya et al., 2018; Zhang and Dong, 2018), we propose an estimator that considers the full autocovariance matrix within each aggregation level when reconciling forecasts in a temporal hierarchy. With the purpose of temporal aggregation being to exploit important information in a time series at different frequencies, it does not make sense to disregard the potentially most important information, namely its autocorrelation structure; at least not when there is enough data available that it can be estimated with reasonable precision. This is often the case in high-frequency settings. Hence, it should be possible to improve accuracy by considering autocorrelation information when reconciling forecasts.

Even with high-frequency data available, it can be difficult to estimate the full covariance matrix without assuming that it has some special form (see, e.g., Bien et al., 2016). Therefore, we propose a second estimator that blends autocorrelation and variance information, but only requires estimation of the first-order autocorrelation coefficient at each aggregation level. This estimator is based on decomposing the autocovariance matrix into two diagonal variance matrices and a block-diagonal autocorrelation matrix that imposes a first-order Markov process on the reconciliation errors.

In order to facilitate information sharing between aggregation levels, we propose a third estimator that uses the graphical LASSO (GLASSO) for estimating a sparse representation of the inverse autocorrelation matrix across aggregation levels (Friedman et al., 2007; Banerjee et al., 2008). This estimator overcomes the problem of inverting a potentially singular autocovariance matrix and is robust to noise and high dimensionality due to the imposed sparsity.

We document the usefulness of our approach through an application to short-term electricity load forecasting in different price areas in Sweden. We will not consider advanced forecasting techniques, nor include explanatory variables such as weather data (see, e.g., Fan and Hyndman, 2012; Clements et al., 2016), rather we will only use exponential-smoothing methods to generate base forecasts. We show that incorporating information about the autocorrelation structure significantly improves forecast accuracy, both compared to traditional approaches, such as bottom up and top down, and compared to the diagonal estimators proposed by Athanasopoulos et al. (2017). Improvements in forecast accuracy are in several cases greatest out of sample where the accuracy of the base forecasts, on average, is lower. In other words, reconciliation increases robustness by increasing accuracy the most when needed the most.

This article is organized as follows. In Section 2, we outline the forecast reconciliation problem and its relation to ordinary, weighted, and generalized least-squares estimation. In Section 3, we propose three different estimators for taking account of the autocorrelation structure. Results from the application to short-term load forecasting are presented in Section 4. Finally, we discuss the results and limitations of the forecast reconciliation approach in Section 5.

2 Forecast reconciliation

2.1 Temporal hierarchies

Given n individual *base* forecasts stacked in a column vector $\hat{y} \in \mathbb{R}^n$, where $\hat{y}_n$ is a forecast of the aggregate, we want to find *reconciled* forecasts $\tilde{y} \in \mathbb{R}^n$, which are aggregate consistent, so that $\sum_{i=1}^{n-1} \tilde{y}_i = \tilde{y}_n$. For example, $\hat{y}_1, \ldots, \hat{y}_{n-1}$ could be sales forecasts for $n-1$ individual stores and $\hat{y}_n$ the aggregate sales forecast for the entire chain of stores. This is an example of a *structural* hierarchy. An example of a *temporal* hierarchy would be if $\hat{y}_1, \ldots, \hat{y}_4$ were quarterly inflation forecasts and $\hat{y}_5$ was the annual inflation forecast.

The framework can easily be extended to multiple aggregation levels; for example, quarterly forecasts should reconcile to half-year forecasts, which should reconcile to annual forecasts, as illustrated in Figure 1. This is most easily done by introducing a summation matrix S, which for the hierarchy illustrated in Figure 1 would be

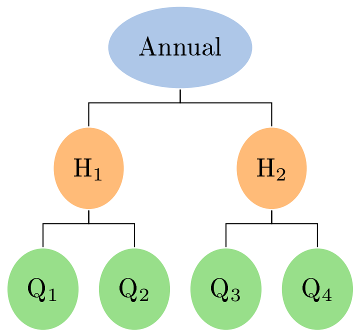


Figure 1: Temporal hierarchy for quarterly series.

$$S = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

It is even possible to have several different forecasts for each quarter, year, etc.

Letting $y^{\text{bottom}} = (y_{\text{Q}_1}, y_{\text{Q}_2}, y_{\text{Q}_3}, y_{\text{Q}_4})^T$ denote the vector of observations at the most disaggregate level $k = 1$, the reconciliation constraint(s) can be written as

$$\tilde{y} = S\tilde{y}^{\text{bottom}}. \tag{1}$$

In general, there are $k \in \{k_1, \ldots, k_K\}$ aggregation levels, where k is a factor of m, with $k_1 = 1$ and $k_K = m$, and m/k is the number of observations at aggregation level k. Then the summation matrix is given by

$$S = \begin{bmatrix} I_{k_K} \otimes 1_{m/k_K} \\ \vdots \\ I_{k_1} \otimes 1_{m/k_1} \end{bmatrix}, \tag{2}$$

where I_k is an identity matrix of order k and 1_m is an m-vector of ones.

2.2 Optimal reconciliation

Van Erven and Cugliari (2015) proposed to formulate the forecast reconciliation problem as

$$\begin{aligned} &\text{minimize} \quad \left\| W^{1/2} (\tilde{y} - \hat{y}) \right\|_2^2 \\ &\text{subject to} \quad \tilde{y} = S\tilde{y}^{\text{bottom}}, \end{aligned} \tag{3}$$

where $\tilde{y} \in \mathbb{R}^n$ is the variable, the parameter $W \in \mathbb{R}^{n \times n}_{++}$ is a diagonal matrix with weights w_i along its diagonal, and $\|y\|_2 = \left(\sum_{i=1}^n y_i^2\right)^{1/2}$ denotes the ℓ_2 norm of a vector $y \in \mathbb{R}^n$ of dimension n. The squared error is the most common choice of loss function. The reconciled forecasts are optimal in that the base forecasts are adjusted by the least amount (in the sense of least squares) so that these become *coherent*. Formulation (3) is convex optimization problem that can readily be solved (Boyd and Vandenberghe, 2004).

2.3 Relation to generalized least squares

Hyndman et al. (2011) and Athanasopoulos et al. (2017) formulated the structural and temporal reconciliation problems, respectively, as linear regression models. The reconciled forecasts can be found using the *generalized* least-squares estimate:

$$\begin{array}{ll} \text{minimize} & (\tilde{y} - \hat{y})^T \Sigma^{-1} (\tilde{y} - \hat{y}) \\ \text{subject to} & \tilde{y} = S\tilde{y}^{\text{bottom}}, \end{array} \tag{4}$$

where $\tilde{y} \in \mathbb{R}^n$ is the variable and the parameter $\Sigma \in \mathbb{R}^{n \times n}_{++}$ is the covariance matrix for the *coherency errors* $\varepsilon = \tilde{y} - \hat{y}$, which are assumed to be multivariate Gaussian and unbiased, i.e., have zero mean.

If Σ were known, the solution to (4) would be given by the GLS estimator

$$\tilde{y} = S\left(S^T \Sigma^{-1} S\right)^{-1} S^T \Sigma^{-1} \hat{y}. \tag{5}$$

When the summation matrix S is very large, it can be faster to solve the optimization problem (4) than to evaluate the closed-form solution (5) (Boyd and Vandenberghe, 2004). Hyndman et al. (2016) showed how the computations required to evaluate (5) can be handled efficiently by exploiting the sparse structure of the summation matrix, while Wickramasuriya et al. (2018) derived an alternative representation that is significantly less demanding in terms of computation.

The close correspondence between (3) and (4) is evident, as the two coincide when $\Sigma^{-1} = W$. This provides a benchmark for selecting the weights. In formulation (4), the precision matrix Σ^{-1} is used to scale deviations from the base forecasts; hence, it is often referred to as a weight matrix. It is more expensive to adjust base forecasts with a higher precision. Another option is to select weights based on prior knowledge about the precision or importance of the base forecasts. The machine-learning approach would be to run a number of tests to find the combination of weights that gives the best (in-sample) result. Hyndman et al. (2011) and van Erven and Cugliari (2015) argued for selecting uniform weights to increase the importance of forecasting the aggregate. The unweighted case $\Sigma = I$ corresponds to *ordinary* least-squares estimation.

Wickramasuriya et al. (2018) showed that, in general, Σ is not known and is not identifiable. By minimizing the variances of the reconciled forecast errors, they proposed an estimator which results in unbiased reconciled forecasts given by the GLS estimator (5), but with a different covariance matrix. As a proxy for the unidentifiable covariance matrix for the coherency errors, they proposed to use the covariance matrix for the reconciled *forecast errors* $e = y - \tilde{y}$. Although it does not suffer from a lack of identifiability, it can still be challenging to estimate. Wickramasuriya et al. (2018) proposed different estimators based on the in-sample base forecast errors.

2.4 Weighted least-squares estimators

Athanasopoulos et al. (2017) proposed three diagonal estimators Λ of increasing simplicity that approximate Σ. For the temporal hierarchy illustrated in Figure 1, the three estimators are

$$\begin{aligned}\Lambda_{\text{struc}} &= \text{diag}\left(1,1,1,1,2,2,4\right),\\ \Lambda_{\text{svar}} &= \text{diag}\left(\sigma^2_{\text{Q}},\sigma^2_{\text{Q}},\sigma^2_{\text{Q}},\sigma^2_{\text{Q}},\sigma^2_{\text{H}},\sigma^2_{\text{H}},\sigma^2_{\text{A}}\right),\\ \Lambda_{\text{hvar}} &= \text{diag}\left(\sigma^2_{\text{Q}_1},\sigma^2_{\text{Q}_2},\sigma^2_{\text{Q}_3},\sigma^2_{\text{Q}_4},\sigma^2_{\text{H}_1},\sigma^2_{\text{H}_2},\sigma^2_{\text{A}}\right).\end{aligned}$$

By definition, these ignore correlations across and within aggregation levels and lead to alternative *weighted* least-squares estimators. They referred to the simplest of the three estimators as *structural* scaling. As base forecast errors at each level of a temporal hierarchy are associated with a single time series, they argued that it is reasonable to assume that the variances at each level are approximately equal. Assuming that the variance of each bottom-level base forecast error is σ^2 and that they are uncorrelated between nodes, they set $\Sigma = \sigma^2\Lambda_{\text{struc}}$, where Λ_{struc} is a diagonal matrix with each element containing the number of forecasts errors contributing to that aggregation level:

$$\Lambda_{\text{struc}} = \text{diag}\left(S1_m\right).$$

This estimator has several desirable properties. First, it depends only on the seasonal period m of the most disaggregated observations and is independent of both data and forecasting model. Second, it permits forecasts which originate from any forecasting method or even predictions from human experts that are not described by a formal model, since no estimation of the variance of the forecast errors is needed. In their empirical evaluation, Athanasopoulos et al. (2017) found that structural scaling often performed at least as well as the more complicated estimators.

The second estimator proposed by Athanasopoulos et al. (2017) is referred to as *series variance* scaling. This estimator includes separate variance estimates for each aggregation level. That is, it assumes homogeneous error variance within a level, but not across levels. They argued that this is a reasonable assumption given that base forecast errors within the same aggregation level are for the same time series.

Generally, the further out in time a forecast is made, the more uncertain it is (see, e.g., Hyndman et al., 2008, Chapter 6.2). Thus, the third estimator proposed by Athanasopoulos et al. (2017) includes separate variance estimates for each base forecast. This estimator is referred to as *hierarchy variance* scaling.

3 Accounting for autocorrelation

3.1 Autocovariance scaling

In a temporal hierarchy, WLS corresponds to ignoring autocorrelation. As the purpose of temporal aggregation is to exploit important information about a time series at different frequencies, we argue that potential information in the autocorrelation structure should be included. Therefore, our first proposal is to estimate the full autocovariance matrix within each aggregation level, while ignoring correlations between aggregation levels.

We refer to this estimator as *hierarchy autocovariance* scaling. For example, for the temporal hierarchy illustrated in Figure 1, the estimator is

$$\Sigma_{\text{hacov}} = \begin{bmatrix} \sigma^2_{\text{Q}_1} & \sigma^2_{\text{Q}_1,\text{Q}_2} & \sigma^2_{\text{Q}_1,\text{Q}_3} & \sigma^2_{\text{Q}_1,\text{Q}_4} & 0 & 0 & 0 \\ \sigma^2_{\text{Q}_1,\text{Q}_2} & \sigma^2_{\text{Q}_2} & \sigma^2_{\text{Q}_2,\text{Q}_3} & \sigma^2_{\text{Q}_2,\text{Q}_4} & 0 & 0 & 0 \\ \sigma^2_{\text{Q}_1,\text{Q}_3} & \sigma^2_{\text{Q}_2,\text{Q}_3} & \sigma^2_{\text{Q}_3} & \sigma^2_{\text{Q}_3,\text{Q}_4} & 0 & 0 & 0 \\ \sigma^2_{\text{Q}_1,\text{Q}_4} & \sigma^2_{\text{Q}_2,\text{Q}_4} & \sigma^2_{\text{Q}_3,\text{Q}_4} & \sigma^2_{\text{Q}_4} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_{\text{H}_1} & \sigma^2_{\text{H}_1,\text{H}_2} & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_{\text{H}_1,\text{H}_2} & \sigma^2_{\text{H}_2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{\text{A}} \end{bmatrix}.$$

Note that if we did not assume zero correlation between aggregation levels, Σ_{hacov} might not have full rank and consequently not be invertible, because of the linear relationship between forecast errors from different aggregation levels.

3.2 Markov process

In high-dimensional hierarchies, even with high-frequency data available, it can be difficult to estimate the full autocovariance matrix within each aggregation level without assuming that it has some special form. Therefore, we also propose an estimator that blends autocorrelation and variance information, but only requires estimation of the first-order autocorrelation coefficient at each aggregation level. The estimator is based on decomposing the autocovariance matrix into two diagonal variance matrices Λ and a block-diagonal autocorrelation matrix that imposes a first-order autoregressive structure on the reconciliation errors.

We refer to this estimator as either *structural*, *series*, or *hierarchy Markov* scaling, depending on which diagonal variance matrix is used. For example, for the temporal hierarchy illustrated in Figure 1, the series Markov estimator is

$$\Sigma_{\text{sMarkov}} = \Lambda^{1/2}_{\text{svar}} \begin{bmatrix} 1 & \rho_{\text{Q}} & \rho^2_{\text{Q}} & \rho^3_{\text{Q}} & 0 & 0 & 0 \\ \rho_{\text{Q}} & 1 & \rho_{\text{Q}} & \rho^2_{\text{Q}} & 0 & 0 & 0 \\ \rho^2_{\text{Q}} & \rho_{\text{Q}} & 1 & \rho_{\text{Q}} & 0 & 0 & 0 \\ \rho^3_{\text{Q}} & \rho^2_{\text{Q}} & \rho_{\text{Q}} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \rho_{\text{H}} & 0 \\ 0 & 0 & 0 & 0 & \rho_{\text{H}} & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \Lambda^{1/2}_{\text{svar}}.$$

If the models used to generate the base forecasts were perfect, then there would not be any residual autocorrelation. This is seldom the case in practical applications when forecasting multiple steps ahead. Yet, if the residual autocorrelation structure is complex and, for example, features seasonality, then it is probably worthwhile revisiting the model before considering forecast reconciliation. Thus, it seems reasonable to assume that the reconciliation errors can be approximated by a first-order Markov process.

3.3 Inverse correlation estimation using GLASSO

A higher-order autoregressive structure could be imposed by considering the inverse of the autocorrelation matrix—i.e., the partial autocorrelation—which has a very simple structure for autoregressive processes (Madsen, 2008, Chapter 5). After all, we only need to estimate the inverse in order to solve (4). Our third proposal is to estimate a sparse representation of the inverse autocorrelation matrix using the graphical LASSO (Friedman et al., 2007; Banerjee et al., 2008). Because the LASSO penalty is sensitive to the scale of variables, we estimate the inverse autocorrelation rather than the inverse autocovariance matrix to avoid problems caused by the inherent heteroscedasticity.

We refer to this estimator as either *series* or *hierarchy GLASSO* scaling, depending on which diagonal variance matrix is used. It will become clear from the discussion in Section 4.4 why we do not introduce a structural version of this estimator. With the hierarchy variances the estimator is

$$\Sigma_{\text{hGLASSO}}^{-1} = \Lambda_{\text{hvar}}^{-1/2} \Theta \Lambda_{\text{hvar}}^{-1/2},$$

where the inverse autocorrelation matrix Θ is found by maximizing the penalized Gaussian log-likelihood

$$\log\det\Theta - \operatorname{tr}(R\Theta) - \lambda \|\Theta\|_1 . \tag{6}$$

The parameter R is the empirical autocorrelation matrix, $\|X\|_1 = \sum_{i=1}^{n}\sum_{j=1}^{n} |X_{ij}|$ denotes the ℓ_1 norm of a square matrix $X \in \mathbb{R}^{n\times n}$ of order n, and λ is a regularization parameter that controls the degree of sparsity in the solution. For small values of λ, GLASSO scaling is equivalent to scaling by the full autocovariance matrix (which is not feasible), while for large values of λ, it is equivalent to scaling by the diagonal variance matrix. There exist several specialized algorithms for finding the solution to the maximization problem (6); however, the problem is convex and can be solved using standard software for convex optimization, such as CVXPY (Diamond and Boyd, 2016; Akshay Agrawal and Boyd, 2018).

The GLASSO estimator has several desirable properties. First, it overcomes the problem of inverting a potentially singular autocovariance matrix. Second, it is robust to noise and high dimensionality due to the imposed sparsity. Third, contrary to autocovariance and Markov scaling, this estimator allows for *information sharing* between aggregation levels. For example, the forecast errors for Q_1 and Q_2 should reasonably be correlated with the forecast error for H_1.

4 Results

4.1 Load forecasting

Automated short-term load forecasting is needed for efficient operation of power systems and to support transactions by participants in deregulated electricity markets (Hahn et al., 2009). For day-ahead prediction, a weather-based model is typically used. However, weather-based models are less important for intraday horizons, as weather variables tend to change relatively smoothly over short intervals of time. Moreover, weather forecasts are sometimes not available or only available with a delay. This prompts consideration of modeling approaches that use only historical load data (Taylor, 2010, 2012).

In short-term load forecasting, the seasonal Holt–Winters exponential smoothing is a common choice for modeling seasonality (Taylor, 2003; Gould et al., 2008; Taylor, 2010; Livera et al., 2011; Taylor, 2012). Holt–Winters exponential smoothing was extended by Taylor (2003) to accommodate intraday and intraweek cycles in intraday data. It is well-suited for electricity-demand forecasting, as demand has both daily and weekly seasonalities.

Letting p_1 and p_2 denote the periods of the two seasons, the additive double-seasonal Holt–Winters method from Taylor (2012) can be presented as the following state-space model:

$$y_t = l_{t-1} + s^{(1)}_{t-p_1} + s^{(2)}_{t-p_2} + \phi e_{t-1} + \varepsilon_t, \tag{7a}$$

$$e_t = y_t - \left(l_{t-1} + s^{(1)}_{t-p_1} + s^{(2)}_{t-p_2}\right), \tag{7b}$$

$$l_t = l_{t-1} + \alpha e_t, \tag{7c}$$

$$s^{(1)}_t = s^{(1)}_{t-p_1} + \gamma_1 e_t, \tag{7d}$$

$$s^{(2)}_t = s^{(2)}_{t-p_2} + \gamma_2 e_t, \tag{7e}$$

where $\varepsilon_t \sim N\left(0, \sigma^2\right)$ and σ^2 is a constant variance; y_t is the load; l_t and $s^{(1)}_t$ are the state variables for the level and intraday cycle, respectively; $s^{(2)}_t$ is the state variable for the intraweek cycle remaining after $s^{(1)}_t$ is removed; α, γ_1, and γ_2 are smoothing parameters; and the term involving ϕ is an autoregressive adjustment for first-order residual autocorrelation. The inclusion of this term greatly improves forecast accuracy (Taylor, 2010). The method assumes the same intraday cycle for all days of the week and does not include a trend component, as short-term electricity demand mostly does not have a trend.

4.2 Data

We consider hourly load data for 2016 and 2017 from Nord Pool.[1] Nord Pool Spot is the power market for Sweden, Norway, Denmark, Finland, Estonia, Latvia, and Lithuania. The day-ahead market is an auction, where power is traded for delivery each hour of the next day. The Nord Pool markets are divided into several bidding areas, as shown in Figure 2. We consider the load in GWh in the four Swedish areas SE1–SE4. We do not apply a log transformation to the load data, as it is custom to do, because it would ruin the summation property of the hierarchy. We use data from 2016 for in-sample training, with the first two weeks being used for initialization, and data from 2017 for out-of-sample testing.

Figure 2: Map showing the situation in the Nordic region. Sweden is divided into four electricity price areas: Malmö (SE4), Stockholm (SE3), Sundsvall (SE2) and Luleå (SE1). Norway currently has five electricity price areas. Denmark is split into Eastern and Western Denmark. Finland, Estonia, Lithuania, and Latvia remain undivided.

Figure 3 shows the autocorrelation functions (ACFs) for three different frequencies of the load data for the SE areas in 2016. Across all frequencies and areas, the load data is strongly autocorrelated. In Figure 3a, the ACFs for the daily load reveal a clear weekly pattern, which is much stronger for SE4 than SE1. In Figure 3b, the ACFs for the six-hourly load display both a daily and weekly variation. Once again, the magnitude of the seasonal variation follows the order of the areas with both seasons being significantly more pronounced in SE4 com-

[1] https://www.nordpoolgroup.com/historical-market-data/

	In sample (2016)				Out of sample (2017)			
	SE1	SE2	SE3	SE4	SE1	SE2	SE3	SE4
Daily								
RMSE	1.07	2.22	8.44	2.77	1.14	2.46	9.15	3.17
RMSPE	4.17	4.95	3.61	4.27	4.32	5.39	3.97	4.92
Six-hourly								
RMSE	0.36	0.72	2.51	0.85	0.33	0.78	2.62	0.93
RMSPE	5.64	6.18	4.07	5.08	4.94	6.39	4.20	5.47
Hourly								
RMSE	0.09	0.14	0.46	0.16	0.08	0.14	0.44	0.16
RMSPE	8.31	7.46	4.52	5.62	7.39	7.43	4.37	5.94

Table 1: In- and out-of-sample RMSE and RMSPE for base forecasts of one-day-ahead power consumption in SE areas at different data frequencies.

pared to SE1. The daily and weekly cycles are also evident from Figure 3c, although the weekly cycle is less pronounced at the hourly frequency.

In order to limit the size of the hierarchy, we choose to focus on the hourly, six-hourly, and daily frequencies and not include the remaining intermediate aggregation levels. This increases clarity and reduces computational complexity. There is no particular reason why we choose six instead of four or eight hours as the intermediate frequency. Base forecasts at the daily frequency are generated using the automatic forecasting procedure based on single exponential smoothing implemented by Hyndman and Khandakar (2008). The six-hourly and hourly base forecasts are generated using the additive double-seasonal Holt–Winters method from Taylor (2012). All model parameters are fitted by minimizing the mean squared error of the in-sample one-step-ahead forecasts using the data from 2016.

4.3 Base forecasts

Table 1 shows the root mean squared error

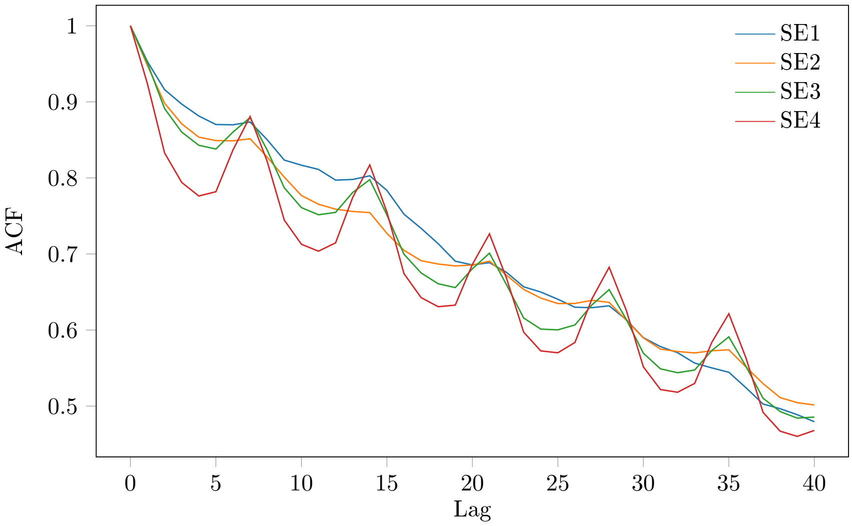
$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2} \tag{8}$$

and the root mean squared percentage error

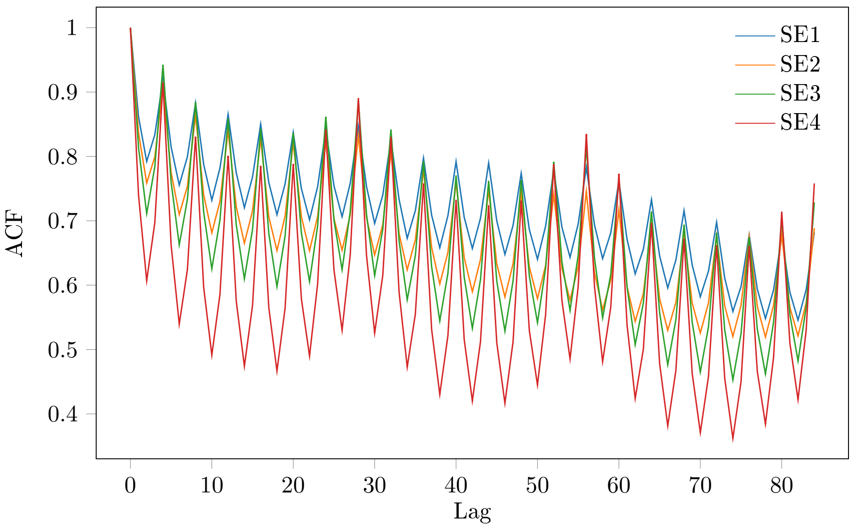
$$\text{RMSPE} = 100\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{y_t - \hat{y}_t}{y_t}\right)^2} \tag{9}$$

for base forecasts of one-day-ahead power consumption in the SE areas at the different data frequencies. At the daily frequency, this corresponds to one-step-ahead predictions, whereas at the six-hourly and hourly frequencies, it corresponds to four- and 24-step-ahead predictions, respectively, once per day at the end of each day. For example, at the hourly frequency the RMSE is an average across the 24 steps each day across all days in the sample.

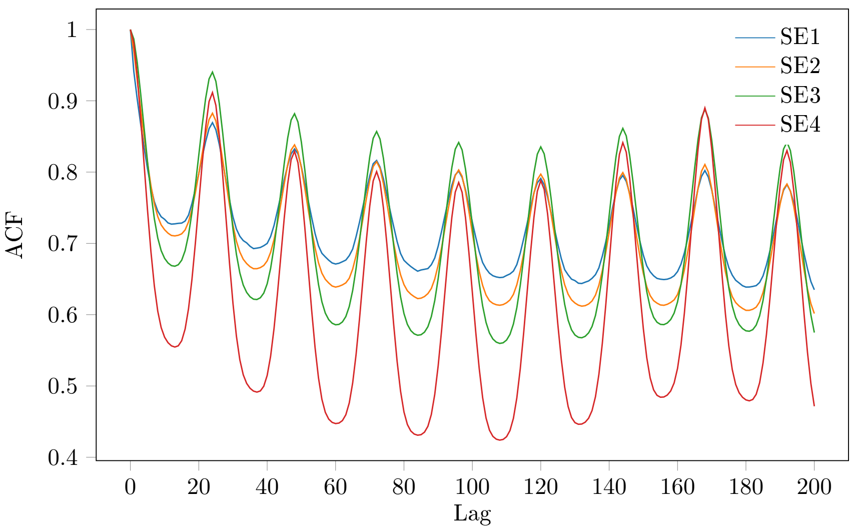
The RMSE is much larger for SE3 compared to the other areas, because consumption in this area is larger. The RMSPE is better suited for comparing the forecast accuracy across the different areas, since it has the advantage of being scale-independent. Looking at the RMSPE, it is evident that load forecasts for SE3 are more accurate, relatively speaking, compared to the other areas.



(a) Daily



(b) Six-hourly



(c) Hourly

Figure 3: Autocorrelation functions for different frequencies of load data for SE areas in 2016.

This is true both in and out of sample. The RMSPE for SE3 is relatively similar across the different data frequencies, whereas it doubles, for example in SE1, when going from the daily to the hourly frequency. It is important to emphasize that this does not mean that the one-day-ahead forecasts based on hourly data are worse than those based on daily data, as it will become evident from the next table.

4.4 Reconciled forecasts

We follow Athanasopoulos et al. (2017) and the recommendation of Hyndman and Koehler (2006) by considering the relative root mean squared error

$$\mathrm{RRMSE} = \frac{\mathrm{RMSE}}{\mathrm{RMSE}^{\mathrm{base}}} - 1 \tag{10}$$

when comparing the accuracy of reconciled and base forecasts. Table 2 compares the RRMSE for the different reconciliation approaches applied to forecasts of one-day-ahead power consumption in the SE areas. The table shows the RRMSE in and out of sample for base forecasts based on different data frequencies. A negative entry shows a percentage decrease in RMSE relative to the base forecast, i.e., improved accuracy. Bold entries identify the best performing approaches in each area.

We use the two-sided Diebold–Mariano test (Diebold and Mariano, 1995) with the modification suggested by Harvey et al. (1997) to compare the accuracy of the reconciled and base forecasts. The null hypothesis is that they have the same accuracy. The results of the test are included in Table 2 with * and ** denoting significance at the 0.05 and 0.01 level, respectively. Due to the length of the sample periods, most of the differences in accuracy are statistically significant at the tested levels.

Traditional approaches The bottom-up approach, where forecasts made at the hourly frequency are aggregated to six-hourly and daily forecasts, in general, is not very successful, though it improves out-of-sample forecasts in some areas. The middle-out approach, where forecasts based on six-hourly data are aggregated to daily forecasts, works well with improvements ranging from 8% to 20%. This means that it is much better to model six-hourly rather than hourly or daily load data, when interested in forecasting the load one day ahead. The six-hourly frequency is a compromise between the loss of information at the daily frequency and the extra noise in hourly data. Disaggregating daily or six-hourly forecasts to hourly forecasts, on the other hand, leads to significantly worse forecasts in most cases.

Ordinary least squares The unweighted OLS approach leads to modest, yet significant, improvements of the daily base forecasts, but in most cases lower accuracy at the six-hourly and hourly frequencies. Hyndman et al. (2011) and van Erven and Cugliari (2015) argued for selecting uniform weights; however, this does not work well for a temporal hierarchy, because of the inherent heteroscedasticity.

Structural scaling Structural scaling, where the variance is assumed to be proportional to the number of forecast errors contributing to each aggregation level, leads to significant improvements of between 7% and 16% at the daily frequency, but only minor improvements, and in a few cases slightly lower accuracy, at the six-hourly and hourly frequencies.

	In sample (2016)				Out of sample (2017)			
	SE1	SE2	SE3	SE4	SE1	SE2	SE3	SE4
Daily								
Bottom up	0.18**	-0.04	0.02	-0.01	-0.09*	-0.09**	-0.11**	-0.10**
Middle out	-0.14**	-0.13*	-0.08*	-0.09*	-0.20**	-0.09*	-0.14**	-0.13**
Unweighted	-0.05**	-0.05**	-0.04**	-0.04**	-0.06**	-0.05**	-0.05**	-0.05**
Structural	-0.08**	-0.10**	-0.07**	-0.08**	-0.16**	-0.10**	-0.12**	-0.11**
Series variance	0.02	-0.07*	-0.02	-0.04	-0.16**	-0.11**	-0.13**	-0.11**
Hierarchy variance	0.02	-0.08*	-0.03	-0.05	-0.16**	-0.11**	-0.14**	-0.12**
Structural Markov	-0.08**	-0.09**	-0.05**	-0.06**	-0.12**	-0.08**	-0.08**	-0.07**
Series Markov	-0.07**	-0.11**	-0.07**	-0.08**	-0.17**	-0.11**	-0.13**	-0.11**
Hierarchy Markov	-0.07**	-0.11**	-0.13**	-0.13**	-0.17**	-0.12**	-0.18**	-0.15**
Hierarchy autocovariance	-0.10**	-0.13**	-0.15**	-0.16**	-0.16**	-0.13**	-0.20**	-0.18**
Series GLASSO	-0.13**	-0.25**	-0.24**	-0.24**	-0.22**	-0.27**	-0.30**	-0.27**
Hierarchy GLASSO	**-0.17****	**-0.26****	**-0.34****	**-0.32****	**-0.23****	**-0.28****	**-0.37****	**-0.34****
Six-hourly								
Bottom up	0.14**	0.02	0.04	0.01	0.06**	-0.08*	-0.05	-0.03
Top down	0.04*	0.07	0.22**	0.22**	0.16**	0.04	0.24**	0.21**
Unweighted	0.04**	0.03	0.01	0.01	0.09**	0.01	0.05*	0.04*
Structural	0.00	-0.02	-0.03	-0.04**	0.01	-0.06*	-0.04*	-0.03
Series variance	0.05*	-0.01	0.00	-0.02	0.01	-0.09**	-0.06*	-0.05
Hierarchy variance	0.05*	-0.01	0.00	-0.03	0.01	-0.09**	-0.07*	-0.05
Structural Markov	0.01	0.00	-0.01	-0.02	0.04**	-0.03	0.02	0.01
Series Markov	0.00	-0.03	-0.04*	-0.05**	0.00	-0.07**	-0.05**	-0.04*
Hierarchy Markov	0.00	-0.03*	-0.08*	-0.08**	0.00	-0.08**	-0.10**	-0.07**
Hierarchy autocovariance	-0.03*	-0.05**	-0.10**	-0.11**	0.00	-0.10**	-0.11**	-0.10**
Series GLASSO	-0.09**	-0.13**	-0.12	-0.15**	-0.06**	-0.17**	-0.16**	-0.13**
Hierarchy GLASSO	**-0.11****	**-0.14****	**-0.25****	**-0.23****	**-0.07****	**-0.19****	**-0.25****	**-0.21****
Hourly								
Middle out	-0.03	0.06*	0.35**	0.36**	0.02	0.14**	0.43**	0.35**
Top down	-0.06**	0.02*	0.20**	0.22**	0.05**	0.10**	0.33**	0.25**
Unweighted	-0.06**	0.01	-0.03	0.00	0.01	0.07**	0.09**	0.06**
Structural	-0.08**	-0.03**	-0.07**	-0.04**	-0.03**	0.01	0.00	0.00
Series variance	-0.05**	-0.02**	-0.03**	-0.03**	-0.03**	-0.01	-0.01**	-0.01**
Hierarchy variance	-0.05**	-0.02**	-0.04**	-0.03**	-0.03**	-0.01	-0.02**	-0.02**
Structural Markov	-0.07**	-0.01	-0.05*	-0.02	-0.01	0.04**	0.06**	0.04*
Series Markov	-0.08**	-0.03**	-0.07**	-0.05**	-0.03**	0.00	-0.01	-0.01
Hierarchy Markov	-0.08**	-0.04**	-0.11**	-0.08**	-0.03**	0.00	-0.05**	-0.04**
Hierarchy autocovariance	-0.09**	-0.05**	-0.13**	-0.11**	-0.03**	-0.01	-0.07**	-0.06**
Series GLASSO	-0.13**	-0.11**	-0.15**	-0.14**	-0.06**	-0.07**	-0.11**	-0.09**
Hierarchy GLASSO	**-0.14****	**-0.12****	**-0.26****	**-0.22****	**-0.07****	**-0.09****	**-0.20****	**-0.17****

Table 2: In- and out-of-sample percentage difference in RMSE between the reconciled and base forecasts of one-day-ahead power consumption in SE areas for different data frequencies. Bold entries identify the best performing approaches. Differences that are statistically significant at the 0.05 and 0.01 level, as evaluated by the Diebold–Mariano test, are marked with * and **, respectively.

Variance scaling Scaling by the estimated pooled one-step-ahead variance at each aggregation level of the series in most cases yields slightly worse results than structural scaling across all frequencies. The results for hierarchy variance scaling, i.e., scaling by the nonpooled one-step-ahead variance at each aggregation level of the hierarchy, are almost identical to series variance scaling.

Markov scaling Hierarchy Markov scaling improves accuracy compared to hierarchy variance scaling in nearly all cases. In most cases it is better than series Markov scaling, and it is never worse. The results for series and hierarchy Markov scaling are much better than those for structural Markov scaling. Structural Markov scaling leads to worse results than structural scaling, especially out of sample. Besides its simplicity, the advantage of structural scaling compared to series and hierarchy variance scaling is that it does not misestimate the aggregated forecast error variance by failing to account for autocorrelation. When accounting explicitly for autocorrelation, the structural assumption appears to be a poor estimate of the ratio between the forecast error variances at different aggregation levels.

Autocovariance scaling Hierarchy autocovariance scaling improves forecast accuracy compared to structural, series variance, hierarchy variance, structural Markov, series Markov, and hierarchy Markov scaling across all frequencies and areas. It also improves accuracy compared to bottom up, middle out, and top down in all but two cases, which suggests that the combination of forecasts from multiple aggregation levels helps decrease forecast errors. Improvements in RMSE relative to the base forecasts range from 0 to 20%, and it is the first approach that reduces RMSE both in and out of sample for all frequencies and areas.

Graphical LASSO The regularization parameter in (6) is found by doing a grid search in sample over the values $\lambda = 10^{-10}, 10^{-9}, \ldots, 10^{0}$ using the hierarchy variances. Results are very stable over the range from 10^{-10} to 10^{-2}. When $\lambda = 1$, results are the same as for hierarchy variance scaling. In Table 2, results are shown for $\lambda = 10^{-3}$. Even though it is suboptimal to use the same regularization parameter for all price areas and both series and hierarchy variances, this increases robustness of the results.

The GLASSO estimators outperform all previously considered reconciliation approaches. While in some cases the series and hierarchy version yield similar results, in many cases the latter significantly outperforms the former. This is remarkable given that both estimators better all other reconciliation approaches by a good margin. Hierarchy GLASSO scaling yields the most accurate reconciled forecasts both in and out of sample across all frequencies and areas, with the greatest improvements occurring at the daily frequency. Improvements in RMSE range from 7% to 37%.

Similar to the general tendency across all the implemented reconciliation approaches, hierarchy GLASSO scaling often leads to greater improvements in forecast accuracy out of sample compared to in sample. There are very few forecasting methods that produce better results out of sample than in sample. The explanation is most likely that greater model and parameter uncertainty causes lower forecast accuracy out of sample, cf. Table 1. The potential improvement that can be achieved through reconciliation and forecast combination is typically greater when base forecasts are worse. Nonetheless, it is an encouraging result that emphasizes the robustness of the reconciliation approach.

5 Discussion

We have proposed to consider forecast error autocorrelation when reconciling forecasts in a temporal hierarchy. Our first proposal was to consider the autocovariance matrix within each aggregation level. Even with high-frequency data available, however, it can be difficult to estimate the full autocovariance matrix. Therefore, we proposed a second estimator that blends autocorrelation and variance information in a Markov structure that only requires estimation of the first-order autocorrelation coefficient at each aggregation level. In order to facilitate information sharing between aggregation levels, we proposed a third estimator that uses the GLASSO for estimating a sparse representation of the inverse autocorrelation matrix across aggregation levels.

We demonstrated the usefulness of the proposed estimators through an application to short-term load forecasting in four price areas in Sweden. By taking account of the autocovariance when reconciling forecasts, accuracy could be improved uniformly across all frequencies and areas. The most accurate forecasts were obtained when allowing information sharing between aggregation levels, with improvements in RMSE of between 7% and 37% in and out of sample.

Our initial idea was to explore a crosstemporal hierarchy, which, in addition to consistency across the hourly, six-hourly, and daily forecasts, requires the sum of forecasts for the four price areas to be consistent with aggregate forecasts. The intention was to exploit information in the correlation between price areas to improve forecasts in each of the areas. When we found that we had to choose between including correlation and autocorrelation information in the reconciliation step, it was evident that autocorrelation was most important in our case of load forecasting.

There are many advantages to reconciling forecasts for hierarchical time series, yet the approach also has its limitations. There are limits to the size of hierarchies that it is computationally feasible to consider. The greatest improvements in forecast accuracy are typically achieved at the highest levels of a hierarchy, where information from the lower levels is aggregated up. The computational constraint limits the frequency of data that it is feasible to consider, though the use of estimators such as GLASSO can alleviate the problem. Another limitation is the inability to apply nonlinear transformations to the data. For example, in the case of load forecasting, it is custom to apply a log transformation to make the distribution less skewed, but doing so would ruin the summation property of the hierarchy and make the reconciliation problem intractable.

The forecast reconciliation problem is fundamentally an optimization problem, which was our motivation for considering a general and flexible formulation based on convex optimization. This provides the flexibility necessary to explore other loss functions than the squared error, to add regularization penalties, to combine structural and temporal aggregation constraints in a crosstemporal hierarchy, or to relax the aggregation constraints. All of which are matters that should be explored in future research, but that it is hardly possible to do in the framework of linear regression.

Acknowledgment

The work of Henrik Madsen was supported by the Centre for IT-Intelligent Energy Systems (CITIES) project funded in part by Innovation Fund Denmark under Grant No. 1305-00027B.

References

Akshay Agrawal, S. D., Robin Verschueren and S. Boyd. "A rewriting system for convex optimization problems." *Journal of Control and Decision*, vol. 5, no. 1 (2018), pp. 42–60.

Amemiya, T. and R. Y. Wu. "The effect of aggregation on prediction in the autoregressive model." *Journal of the American Statistical Association*, vol. 67, no. 339 (1972), pp. 628–632.

Athanasopoulos, G., R. A. Ahmed, and R. J. Hyndman. "Hierarchical forecasts for australian domestic tourism." *International Journal of Forecasting*, vol. 25, no. 1 (2009), pp. 146–166.

Athanasopoulos, G., R. J. Hyndman, N. Kourentzes, and F. Petropoulos. "Forecasting with temporal hierarchies." *European Journal of Operational Research*, vol. 262, no. 1 (2017), pp. 60–74.

Athanasopoulos, G., R. J. Hyndman, H. Song, and D. C. Wu. "The tourism forecasting competition." *International Journal of Forecasting*, vol. 27, no. 3 (2011), pp. 822–844.

Banerjee, O., L. E. Ghaoui, and A. d'Aspremont. "Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data." *Journal of Machine Learning Research*, vol. 9 (2008), pp. 485–516.

Bien, J., F. Bunea, and L. Xiao. "Convex banding of the covariance matrix." *Journal of the American Statistical Association*, vol. 111, no. 514 (2016), pp. 834–845.

Boyd, S. and L. Vandenberghe. *Convex Optimization*. Cambridge University Press: New York (2004).

Clemen, R. T. "Combining forecasts: A review and annotated bibliography." *International Journal of Forecasting*, vol. 5, no. 4 (1989), pp. 559–583.

Clements, A., A. Hurn, and Z. Li. "Forecasting day-ahead electricity load using a multiple equation time series approach." *European Journal of Operational Research*, vol. 251, no. 2 (2016), pp. 522–530.

Diamond, S. and S. Boyd. "CVXPY: A Python-embedded modeling language for convex optimization." *Journal of Machine Learning Research*, vol. 17, no. 83 (2016), pp. 1–5.

Diebold, F. X. and R. S. Mariano. "Comparing predictive accuracy." *Journal of Business & Economic Statistics*, vol. 13, no. 3 (1995), pp. 253–263.

Fan, S. and R. J. Hyndman. "Short-term load forecasting based on a semi-parametric additive model." *IEEE Transactions on Power Systems*, vol. 27, no. 1 (2012), pp. 134–141.

Friedman, J., T. Hastie, and R. Tibshirani. "Sparse inverse covariance estimation with the graphical lasso." *Biostatistics*, vol. 9, no. 3 (2007), pp. 432–441.

Ghysels, E., P. Santa-Clara, and R. Valkanov. "The MIDAS touch: Mixed data sampling regression models." Working paper, UNC and UCLA (2004).

Gould, P. G., A. B. Koehler, J. K. Ord, R. D. Snyder, R. J. Hyndman, and F. Vahid-Araghi. "Forecasting time series with multiple seasonal patterns." *European Journal of Operational Research*, vol. 191, no. 1 (2008), pp. 207–222.

Gross, C. W. and J. E. Sohl. "Disaggregation methods to expedite product line forecasting." *Journal of Forecasting*, vol. 9, no. 3 (1990), pp. 233–254.

Hahn, H., S. Meyer-Nieberg, and S. Pickl. "Electric load forecasting methods: Tools for decision making." *European Journal of Operational Research*, vol. 199, no. 3 (2009), pp. 902–907.

Hall, S. G. and J. Mitchell. "Combining density forecasts." *International Journal of Forecasting*, vol. 23, no. 1 (2007), pp. 1–13.

Harvey, D., S. Leybourne, and P. Newbold. "Testing the equality of prediction mean squared errors." *International Journal of Forecasting*, vol. 13, no. 2 (1997), pp. 281–291.

Hyndman, R. J., R. A. Ahmed, G. Athanasopoulos, and H. L. Shang. "Optimal combination forecasts for hierarchical time series." *Computational Statistics & Data Analysis*, vol. 55, no. 9 (2011), pp. 2579–2589.

Hyndman, R. J. and Y. Khandakar. "Automatic time series forecasting: The forecast package for R." *Journal of Statistical Software*, vol. 27, no. 3 (2008), pp. 1–22.

Hyndman, R. J. and A. B. Koehler. "Another look at measures of forecast accuracy." *International Journal of Forecasting*, vol. 22, no. 4 (2006), pp. 679–688.

Hyndman, R. J., A. B. Koehler, J. K. Ord, and R. D. Snyder. *Forecasting with Exponential Smoothing: The State Space Approach*. Springer: Berlin (2008).

Hyndman, R. J., A. J. Lee, and E. Wang. "Fast computation of reconciled forecasts for hierarchical and grouped time series." *Computational Statistics & Data Analysis*, vol. 97 (2016), pp. 16–32.

Kourentzes, N., F. Petropoulos, and J. R. Trapero. "Improving forecasting by estimating time series structural components across multiple frequencies." *International Journal of Forecasting*, vol. 30, no. 2 (2014), pp. 291–302.

Ledoit, O. and M. Wolf. "A well-conditioned estimator for large-dimensional covariance matrices." *Journal of Multivariate Analysis*, vol. 88, no. 2 (2004), pp. 365–411.

Livera, A. M. D., R. J. Hyndman, and R. D. Snyder. "Forecasting time series with complex seasonal patterns using exponential smoothing." *Journal of the American Statistical Association*, vol. 106, no. 496 (2011), pp. 1513–1527.

Madsen, H. *Time Series Analysis*. Chapman & Hall: London (2008).

Nystrup, P., H. Madsen, and E. Lindström. "Long memory of financial time series and hidden Markov models with time-varying parameters." *Journal of Forecasting*, vol. 36, no. 8 (2017), pp. 989–1002.

Petropoulos, F. and N. Kourentzes. "Forecast combinations for intermittent demand." *Journal of the Operational Research Society*, vol. 66, no. 6 (2015), pp. 914–924.

Rostami-Tabar, B., M. Z. Babai, A. Syntetos, and Y. Ducq. "Demand forecasting by temporal aggregation." *Naval Research Logistics*, vol. 60, no. 6 (2013), pp. 479–498.

Sedoc, J., J. Rodu, D. Foster, and L. Ungar. "Multiscale hidden Markov models for covariance prediction." (2018).

Silvestrini, A. and D. Veredas. "Temporal aggregation of univariate and multivariate time series models: A survey." *Journal of Economic Surveys*, vol. 22, no. 3 (2008), pp. 458–497.

Taieb, S. B. "Sparse and smooth adjustments for coherent forecasts in temporal aggregation of time series." In *Proceedings of the Time Series Workshop at NIPS 2016*, edited by O. Anava, A. Khaleghi, M. Cuturi, V. Kuznetsov, and A. Rakhlin, vol. 55 of *Proceedings of Machine Learning Research* (2017), pp. 16–26.

Taieb, S. B., J. W. Taylor, and R. J. Hyndman. "Coherent probabilistic forecasts for hierarchical time series." In *Proceedings of the 34th International Conference on Machine Learning*, edited by D. Precup and Y. W. Teh, vol. 70 of *Proceedings of Machine Learning Research* (2017a), pp. 3348–3357.

———. "Hierarchical probabilistic forecasting of electricity demand with smart meter data." (2017b).

Taylor, J. W. "Short-term electricity demand forecasting using double seasonal exponential smoothing." *Journal of the Operational Research Society*, vol. 54, no. 8 (2003), pp. 799–805.

———. "Exponentially weighted methods for forecasting intraday time series with multiple seasonal cycles." *International Journal of Forecasting*, vol. 26, no. 4 (2010), pp. 627–646.

———. "Short-term load forecasting with exponentially weighted methods." *IEEE Transactions on Power Systems*, vol. 27, no. 1 (2012), pp. 458–464.

Tiao, G. C. "Asymptotic behaviour of temporal aggregates of time series." *Biometrika*, vol. 59, no. 3 (1972), pp. 525–531.

Tibshirani, R., M. Saunders, S. Rosset, J. Zhu, and K. Knight. "Sparsity and smoothness via the fused lasso." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 1 (2005), pp. 91–108.

Timmermann, A. "Forecast combinations." In *Handbook of Economic Forecasting*, edited by G. Elliott, C. W. J. Granger, and A. Timmermann, vol. 1, chap. 4. Elsevier: Amsterdam (2006), pp. 135–196.

van Erven, T. and J. Cugliari. "Game-theoretically optimal reconciliation of contemporaneous hierarchical time series forecasts." In *Modeling and Stochastic Learning for Forecasting in High Dimensions*, edited by A. Antoniadis, J.-M. Poggi, and X. Brossat, vol. 217 of *Lecture Notes in Statistics*. Springer: Cham (2015), pp. 297–317.

Wickramasuriya, S. L., G. Athanasopoulos, and R. J. Hyndman. "Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization." *Journal of the American Statistical Association* (2018), pp. 1–16. URL `http://dx.doi.org/10.1080/01621459.2018.1448825`.

Yang, D., H. Quan, V. R. Disfani, and L. Liu. "Reconciling solar forecasts: Geographical hierarchy." *Solar Energy*, vol. 146 (2017a), pp. 276–286.

Yang, D., H. Quan, V. R. Disfani, and C. D. Rodríguez-Gallegos. "Reconciling solar forecasts: Temporal hierarchy." *Solar Energy*, vol. 158 (2017b), pp. 332–346.

Zhang, Y. and J. Dong. "Least squares-based optimal reconciliation method for hierarchical forecasts of wind power generation." *IEEE Transactions on Power Systems* (2018), pp. 1–1. URL http://dx.doi.org/10.1109/tpwrs.2018.2868175.