13.08.2020 16:25:46

Compare Results

Old File:		New File:
main_old.pdf	versus	**main_new.pdf**
33 pages (680 KB)		**37 pages (1.42 MB)**
09.04.2020 17:00:43		13.08.2020 16:12:21

Total Changes

468

Content

134 Replacements

144 Insertions

127 Deletions

Styling and Annotations

26 Styling

37 Annotations

Go to First Change (page 1)

Forecasting Swiss Exports using Bayesian Forecast Reconciliation

Florian Eckert*[†], Rob J. Hyndman[‡], and Anastasios Panagiotelis[‡]

[†]KOF Swiss Economic Institute, ETH Zurich
[‡]Department of Econometrics and Business Statistics, Monash University

August 2020

Abstract

This paper proposes a novel forecast reconciliation framework using Bayesian state-space methods. It allows for the joint reconciliation of all forecast periods and uses predictive distributions rather than past variation of forecast errors. Informative priors are used to assign weights to specific predictions, which makes it possible to reconcile forecasts such that they accommodate specific judgmental predictions or managerial decisions. The reconciled forecasts adhere to hierarchical constraints, which facilitates communication and supports aligned decision-making at all levels of complex hierarchical structures. An extensive forecasting study is conducted on a large collection of 13,118 time series that measure Swiss merchandise exports, grouped hierarchically by export destination and product category. We find strong evidence that in addition to producing coherent forecasts, reconciliation also leads to substantial improvements in forecast accuracy. The use of state-space methods is particularly promising for optimal decision-making under conditions with increased model uncertainty and data volatility.

JEL Classification: C32, C53, E17
Keywords: Forecasting, Hierarchical Reconciliation, Optimal Combination, Decision-making.

*Corresponding Author: Leonhardstrasse 21, 8092 Zürich, eckert@kof.ethz.ch, +41 44 632 29 80
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

1 Introduction

Forecasts are essential to the decision-making process in business analytics and macroeconomics. At an aggregate level, they are often strategic and, therefore, subject to judgmental adjustments and managerial decisions. At more disaggregate operational levels, forecasts usually rely on statistical methods. If the data is subject to linear hierarchical constraints, predictions generated from different methods and information sets are usually not coherent. In some instances, incoherent predictions are problematic because they may lead to contradictory conclusions and non-aligned decision-making. Furthermore, incoherent forecasts are often difficult to communicate. An ex post adjustment of forecasts to ensure coherence resolves these issues and has been shown to lead to substantial improvements in forecast accuracy (see Wickramasuriya et al., 2019, and references therein). This paper proposes a novel approach to jointly reconcile all forecasting periods using state-space methods. It allows for the identification and shrinkage of coherence errors, which means specific predictions can be assigned weights in the process of reconciliation. It may occur, for instance, that managerial decisions are reflected in scenario forecasts at the strategic level, but not in forecasts at the operational level. If the prediction at the strategic level is believed to be more accurate, the proposed method allows to reconcile the incoherent forecasts such that the entire hierarchy is consistent with the initial prediction for the strategic level. This supports aligned decision-making across all operational units while maintaining a high degree of flexibility.

Hierarchical data can be structured according to various characteristics such as geographical, organizational, societal or temporal features (Kourentzes and Athanasopoulos, 2019). Swiss merchandise exports, for example, can be disaggregated geographically into destination regions, such as Western Europe, North America or Australia. These regional aggregates can then be divided further by country. Total exports can also be disaggregated into product categories, such as precision instruments, textiles or vehicles and then further into subcategories, such as road, rail, air and water vehicles. As a result, the data has the structure of a so-called grouped hierarchy (see Hyndman and Athanasopoulos, 2018, and references therein). Figure 1 gives a simple example of grouped structure with $k = 3$ levels, $m = 9$ series in total and $q = 4$ series at the most disaggregate or 'bottom' level.

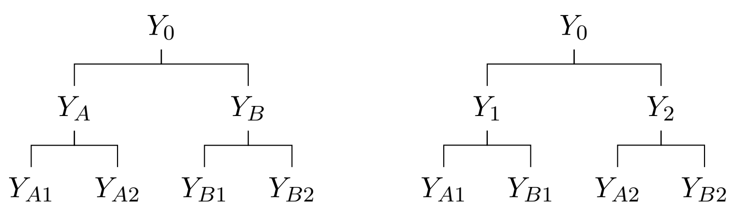

Figure 1: Simple Example of a Grouped Hierarchy. The data is structured into $k = 3$ levels, $m = 9$ time series in total and $q = 4$ time series at the bottom level.

Since it is known that all future realizations of the data will adhere to the constraints implied by the aggregation structure, a desirable property of any forecasts is that they also respect these constraints. Such forecasts are referred to as ‘coherent’. Earlier literature reduced the issue of producing coherent forecasts to one of predicting only a specific level of the hierarchy. For example, the ‘bottom-up’ approach (Gross and Sohl, 1990) achieves coherence by producing only forecasts for the bottom-level series and then summing these up according to the hierarchical structure. A major shortcoming of this approach is that disaggregate series tend to be noisy and there is a high risk of model misspecification. Features such as seasonality may be impossible to identify in the bottom-level data, despite being clearly present in the aggregate series. To address this shortcoming, a ‘top-down’ approach was proposed (see Athanasopoulos et al., 2009, and references therein), where the predicted top level series is disaggregated according to historical or forecasted proportions of lower levels. A compromise is given by the ‘middle-out’ approach, where the forecasts at an intermediate level of the hierarchy are summed up to get the higher levels and disaggregated to obtain lower-level predictions. A weakness of these single-level methods is information loss because the time series characteristics at other levels are not taken into account.

In response to these shortcomings, there has been a tendency over the past decade towards producing forecasts for all series in the hierarchy rather than only at a single level. These are referred to as ‘base’ forecasts and they generally do not adhere to aggregation constraints. Forecast reconciliation, introduced by Hyndman et al. (2011), performs an ex post adjustment to base forecasts in order to produce a new set of coherent forecasts. This adjustment effectively combines predictions from all levels and in doing so ‘hedges’ against misspecification error across the entire hierarchy. It has been shown repeatedly that linear combinations of prediction models lead to better and more robust forecasts (see, for instance, Stock and Watson, 2006; Conflitti et al., 2015). There is now substantial theoretical and empirical evidence that forecast reconciliation can significantly improve forecast accuracy for hierarchical data (see Wickramasuriya et al., 2019, and references therein).

In order to encode the aggregation constraints in a hierarchy, $\mathbf{y}_t$ is defined to be an m-vector that stacks observations at time t from all series, $\mathbf{b}_t$ to be a subvector of $\mathbf{y}_t$ containing only the q bottom-level series at time t and $\mathbf{S}$ to be an $m \times q$ aggregation matrix. In the simple grouped hierarchy shown in Figure 1, these are given by

$$\underset{(m\times 1)}{\mathbf{y}_t} = \begin{bmatrix} Y_0 \\ Y_A \\ Y_B \\ Y_1 \\ Y_2 \\ Y_{A1} \\ Y_{A2} \\ Y_{B1} \\ Y_{B2} \end{bmatrix} \qquad \underset{(m\times q)}{\mathbf{S}} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad \underset{(q\times 1)}{\mathbf{b}_t} = \begin{bmatrix} Y_{A1} \\ Y_{A2} \\ Y_{B1} \\ Y_{B2} \end{bmatrix}.$$

Here and in general, the matrix $\mathbf{S}$ is defined such that $\mathbf{y}_t = \mathbf{S}\mathbf{b}_t$ holds for all realized data. To reconcile these base forecasts, Hyndman et al. (2011) and later authors assumed the following regression structure:

$$\mathbf{y}_t(h) = \mathbf{S}\boldsymbol{\beta}_h + \mathbf{e}_t(h), \tag{1}$$

where $\mathbf{y}_t(h)$ is an m-vector containing the h-periods-ahead base forecasts at time t for each level in the hierarchy. The error term $\mathbf{e}_t(h)$ has mean zero and covariance matrix $\boldsymbol{\Sigma}_h$, and $\boldsymbol{\beta}_h$ represents the unknown mean of the bottom-level series which combines information about forecasts at all levels. It can be estimated using the regression equation

$$\boldsymbol{\beta}_h = \left(\mathbf{S}'\mathbf{W}_h^{-1}\mathbf{S}\right)^{-1}\mathbf{S}'\mathbf{W}_h^{-1}\mathbf{y}_t(h). \tag{2}$$

A vector of reconciled forecasts is then given by $\mathbf{S}\boldsymbol{\beta}_h$ and will adhere to the aggregation constraints by construction. There are several potential choices for $\mathbf{W}_h$. Letting $\mathbf{W}_h = \mathbf{I}_m$ corresponds to an ordinary least squares estimate, in the following also referred to as a 'no scaling' estimate. Alternatively, a high degree of heteroskedasticity in the error terms motivates a diagonal $\mathbf{W}_h$ or weighted least squares approach (Hyndman et al., 2016). Under so-called 'variance scaling', weights are the variances of in-sample h-step ahead forecast variances, and forecasts with less accurate historical performance are down-played in reconciliation. Another alternative is the 'structural scaling' approach suggested by Athanasopoulos et al. (2017), whereby weights are based on the number of series aggregated at each node. More recently, the 'MinT' approach was developed by Wickramasuriya et al. (2019) to allow for a $\mathbf{W}_h$ that is not diagonal and exploits the covariances between the h-step-ahead reconciled forecast errors. The nomenclature refers to the fact that this approach minimizes the trace of the covariance matrix of reconciliation errors. These methods reconcile each forecast period independently and sometimes use a scaled version of $\mathbf{W}_1$ for each h to simplify computations.

This paper contributes to the literature on forecast reconciliation and aligned decision-making in various ways. First, it introduces an explicit identification of the reconciliation errors and provides a representation in state-space form. This allows for the joint reconciliation of all forecast horizons, thereby extending the established literature where each horizon is reconciled independently and the reconciliation biases are treated as part of the error term. Second, the weights used in the reconciliation are derived from the predictive distribution rather than past variation of the base forecast errors. Our innovation is, therefore, particularly promising for forecasting models that allow for conditional heteroskedasticity. Third, we introduce an efficient Bayesian estimation algorithm that enables the inclusion of prior information. These informative priors can be used to shrink the influence of particular series irrespective of past forecasting performance. This is valuable if forecasters have strong judgmental reasons for believing that a particular model will work well in the future while other base forecasts will be less reliable. The proposed framework, therefore, explores also weighting options that don't necessarily increase the predictive accuracy but are very useful in operational forecasting. As a result, the proposed shrinkage approach not only extends the literature on optimal combination but also allows for the replication of existing single-level methods. It furthermore uses informative prior distributions to address some issues in the literature, namely the occurrence of negative reconciled forecasts and singular forecast error covariance matrices. Lastly, the proposed model and existing methods are evaluated using a comprehensive grouped hierarchy of Swiss merchandise trade. Apart from the comparative analysis, this application provides insights for public authorities as well as exporting firms. Merchandise forecasts often serve as inputs into projections of other quantities such as sales, inventories, currency reserves and production capacity.

The remainder of the paper is structured as follows. Section 2 introduces our novel Bayesian state-space reconciliation framework and an efficient estimation algorithm. Section 3 introduces in detail the data on exports of Swiss goods, using modern techniques for exploring and visualizing high-dimensional time series. Section 4 conducts an extensive forecast evaluation that compares our proposed method with existing reconciliation techniques. In addition, it highlights the usefulness of bias shrinkage for applications in operational forecasting. Section 5 concludes.

2 Reconciliation using Bayesian State-Space Methods

This section proposes a novel approach to forecast reconciliation. We show how to explicitly identify the reconciliation errors and estimate them as latent states that evolve over the forecasting horizon. Our method involves the use of predictive distributions instead of historical forecast errors and uses prior information to address several issues in the literature.

2.1 Model

An integrated reconciliation of all forecasted periods has the advantage of combining information across the entire forecasting horizon. If a base forecast in any given forecasted period is revised downwards as a result of the reconciliation, it is likely to be revised downwards as well in the next period. This dependency can be taken into account using state-space methods, which could further improve forecasting accuracy. Pennings and van Dalen (2017) pioneered their use in order to integrate the reconciliation of incoherent information across time periods. They assume the bottom-level series to be the underlying states and use elaborate state equations to capture their stochastic properties. While this takes the intertemporal dependencies nicely into account, it requires the estimation of initial states and restricts the number of usable models for the base forecasts. We propose the explicit identification of state-dependent reconciliation errors $\boldsymbol{\alpha}_h$, which leaves the coherent bottom-level forecasts $\boldsymbol{\beta}_h$ completely free of any assumptions or restrictions:

$$E\left[\mathbf{y}_{T+h}\right] = \mathbf{S}\boldsymbol{\beta}_h = E\left[\mathbf{y}_T(h)\right] - \boldsymbol{\alpha}_h \tag{3}$$

The expected h-step-ahead reconciled forecasts are given by $\mathbf{S}\boldsymbol{\beta}_h$. The m-dimensional vector $\boldsymbol{\alpha}_h = E\left[\mathbf{y}_T(h)\right] - \mathbf{S}\boldsymbol{\beta}_h$ contains the reconciliation biases. It can be interpreted as a fixed effect that is unique to each forecasted variable. $\mathbf{S}$ is a summation matrix of order $m \times q$ and the q-vector $\boldsymbol{\beta}_h$ estimates the unknown mean of the reconciled bottom-level forecasts. The measurement equation is then given as

$$\mathbf{y}_T(h) = \boldsymbol{\alpha}_h + \mathbf{S}\boldsymbol{\beta}_h + \mathbf{e}_T(h), \qquad \mathbf{e}_T(h) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_h). \tag{4}$$

The m-vector $\mathbf{e}_T(h)$ consists of h-step ahead base forecast errors that follow a normal distribution with mean zero and covariance matrix $\boldsymbol{\Sigma}_h$. It can be estimated by taking a sample of n prediction errors $\hat{\mathbf{e}}_T(h)$ from the predictive distribution of $\mathbf{y}_T(h)$. Sampling from the predictive distribution allows also to obtain an estimate for the mean of the incoherent base forecasts $E\left[\mathbf{y}_T(h)\right] = \hat{\mathbf{y}}_T(h)$. These draws may originate from posterior predictive distributions resulting from Bayesian forecasting models (Amisano and Geweke, 2017), bootstrap aggregating (Bergmeir et al., 2016), model pooling (Timmermann, 2006; Kapetanios et al., 2015), or sampling from a fitted model (Hyndman and Athanasopoulos, 2018).

In order to reconcile all forecasting horizons jointly, the coherence errors $\boldsymbol{\alpha}_h$ are modeled to be state-dependent. They are assumed to follow a random walk, which is very common in the literature on time-varying parameters (see, for instance, Primiceri, 2005, and references therein). This implies that the best guess for a coherence error in any given forecasting period is the error in the preceding period. The state equation is, therefore, specified as

$$\boldsymbol{\alpha}_h = \boldsymbol{\alpha}_{h-1} + \mathbf{v}_h, \qquad \mathbf{v}_h \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}). \tag{5}$$

The initial state $\boldsymbol{\alpha}_{\mathbf{0}}$ is the coherence error in the last observation $\mathbf{y}_T$, which is conveniently known to be zero. It is however necessary to impose some restrictions in order to identify the parameters, which is a result of multicollinearity in equation (3). This is quite intuitive since there is more than one unique way to reconcile incoherent forecasts. To show this formally, the coherence errors $\boldsymbol{\alpha}_h$ can be expressed equivalently by concentrating out $\boldsymbol{\beta}_h$ using the projection matrix $\mathbf{P}_h = \mathbf{S}(\mathbf{S}'\boldsymbol{\Sigma}_h^{-1}\mathbf{S})^{-1}\mathbf{S}'\boldsymbol{\Sigma}_h^{-1}$. This leads to the following identity: $(\mathbf{I}_m - \mathbf{P}_h)\,\boldsymbol{\alpha}_h = (\mathbf{I}_m - \mathbf{P}_h)\,\hat{\mathbf{y}}_T(h)$. It is useful to define the idempotent residual maker $\mathbf{M}_h = \mathbf{I}_m - \mathbf{P}_h$. Since $\mathbf{M}_h$ is not invertible due to the presence of multicollinearity, the identity cannot be solved for $\boldsymbol{\alpha}_h$. Our identifying assumption is that $\boldsymbol{\alpha}_h$ lies in the span of $\mathbf{M}_h$, in which case $\mathbf{M}\boldsymbol{\alpha}_h = \boldsymbol{\alpha}_h$. This solves the identification problem and leaves the reconciliation biases as a function of the data and the residual maker $\mathbf{M}_h$. This result is also intuitive since the reconciliation biases are the residuals from a regression of the base forecasts on the aggregation matrix.[1]

2.2 Estimation

The latent states are sampled jointly using the efficient state smoothing and simulation algorithm proposed by Chan and Jeliazkov (2009). We get the marginal distributions by approximating the joint posterior distribution via Gibbs sampling from the conditional distributions (Ando and Zellner, 2010). Convergence is achieved very quickly, irrespective of the starting values. We take a sample of size 1000 from the joint posterior distribution after a burn-in of 100 draws. The measurement equation (4) is stacked over the H forecasting periods in order to reconcile all reconciliation error states jointly.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta} + \mathbf{e}, \quad \mathbf{e} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \tag{6}$$

where the parameters of interest are given by

$$\underset{((H+1)m\times 1)}{\boldsymbol{\alpha}} = \begin{bmatrix} \boldsymbol{\alpha}_0 \\ \vdots \\ \boldsymbol{\alpha}_H \end{bmatrix}, \quad \underset{(Hq\times 1)}{\boldsymbol{\beta}} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_H \end{bmatrix}, \quad \underset{(Hm\times Hm)}{\boldsymbol{\Sigma}} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & & \\ & \ddots & \\ & & \boldsymbol{\Sigma}_H \end{bmatrix}.$$

The unreconciled base forecast means $\hat{\mathbf{y}}_T(h)$, the summation matrices and some further identities are stacked accordingly into

$$\underset{(Hm\times 1)}{\mathbf{y}} = \begin{bmatrix} \hat{\mathbf{y}}_T(1) \\ \vdots \\ \hat{\mathbf{y}}_T(H) \end{bmatrix}, \quad \underset{(Hm\times (H+1)m)}{\mathbf{X}} = \begin{bmatrix} & \mathbf{I}_m & & \\ \mathbf{0} & & \ddots & \\ & & & \mathbf{I}_m \end{bmatrix}, \quad \underset{(Hm\times Hq)}{\mathbf{Z}} = \begin{bmatrix} \mathbf{S} & & \\ & \ddots & \\ & & \mathbf{S} \end{bmatrix}.$$

The state equation (5) needs to be written correspondingly as

$$\mathbf{F}\boldsymbol{\alpha} = \mathbf{v}, \quad \mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}) \tag{7}$$

where

$$\underset{(Hm\times (H+1)m)}{\mathbf{F}} = \begin{bmatrix} \mathbf{I}_m & & & \\ -\mathbf{I}_m & \mathbf{I}_m & & \\ & \ddots & \ddots & \\ & & -\mathbf{I}_m & \mathbf{I}_m \end{bmatrix}, \quad \underset{(Hm\times Hm)}{\mathbf{G}} = \begin{bmatrix} \boldsymbol{\Omega}_0 & & & \\ & \boldsymbol{\Omega} & & \\ & & \ddots & \\ & & & \boldsymbol{\Omega} \end{bmatrix}.$$

The initial state $\boldsymbol{\alpha}_\mathbf{0}$ is known to be zero since it corresponds to the coherence error in the last observation $\mathbf{y}_T$. It is sufficient to choose $\boldsymbol{\Omega}_0$ very small in order shrink the initial state $\boldsymbol{\alpha}_0$ towards zero. Furthermore, the stacked residual maker can then be calculated as $\mathbf{M} = \mathbf{I}_{Hm} - \mathbf{Z}(\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\Sigma}^{-1}$. Using the identification described in section 2, it is

[1] See Appendix A.1 for a detailed derivation.

then straightforward to rewrite the stacked measurement equation (6) as

$$\mathbf{My} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{e}, \quad \mathbf{e} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}). \tag{8}$$

Following Chan and Jeliazkov (2009), the conditional posterior distribution of $\boldsymbol{\alpha}$ is then given by

$$\begin{aligned} \boldsymbol{\alpha} \sim \mathcal{N}(\mathbf{a}_1, \mathbf{A}_1) \quad \text{where} \quad \mathbf{A}_1 &= (\mathbf{F}'\mathbf{G}^{-1}\mathbf{F} + \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1} \\ \mathbf{a}_1 &= \mathbf{A}_1(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{My}). \end{aligned}$$

This algorithm is computationally very efficient if block-banded matrices and sparse matrix algorithms are used. Following Chan and Jeliazkov (2009), it is even faster to compute the banded Cholesky factor of $\mathbf{A}_1$ and solve for $\mathbf{a}_1$ by forward- and backward substitution. The bottom-level means $\boldsymbol{\beta}$ are retrieved from

$$\begin{aligned} \boldsymbol{\beta} \sim \mathcal{N}(\mathbf{b}_1, \mathbf{B}_1) \quad \text{where} \quad \mathbf{B}_1 &= \left(\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z} + \mathbf{B}_0^{-1}\right)^{-1} \\ \mathbf{b}_1 &= \mathbf{B}_1\left(\mathbf{Z}'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}) + \mathbf{B}_0^{-1}\mathbf{b}_0\right). \end{aligned}$$

The priors $\mathbf{b}_0$ and $\mathbf{B}_0$ should be chosen to be as uninformative as possible. An inconvenient effect of most reconciliation methods is the occurrence of negative reconciled bottom-level forecasts. This might be a concern since many applications such as sales or exports do not allow for negative observations. Using a truncated normal prior, this issue can be resolved in an uncomplicated fashion by simply discarding draws of $\boldsymbol{\beta}$ that contain negative entries during the sampling process.

The covariance matrix of the state equation errors $\boldsymbol{\Omega}$ is chosen to be diagonal because the reconciliation errors are assumed to be unique for each base forecast model. The diagonal elements $\omega_1, \dots, \omega_m$ can be retrieved from an inverse-gamma distribution:

$$\begin{aligned} \omega_i \sim \mathcal{IG}(c_1/2, d_1/2), \quad \text{where} \quad c_1 &= c_0 + H \\ d_1 &= d_0 + (\boldsymbol{\alpha}_{h,i} - \boldsymbol{\alpha}_{h-1,i})'(\boldsymbol{\alpha}_{h,i} - \boldsymbol{\alpha}_{h-1,i}) \end{aligned}$$

where $\boldsymbol{\alpha}_{h,i}$ denotes the reconciliation error of series $i \in \{1, \dots, m\}$ at forecasting period $h \in \{1, \dots, H\}$. We choose a weakly informative proper prior distribution with $c_0 = 3$ and $d_0 = 0.01$ in order to restrict the movement of the reconciliation biases slightly.

The covariance matrix of the base forecast errors $\boldsymbol{\Sigma}_h$ is assumed to be diagonal as well since it can be very cumbersome to estimate in large hierarchies and for longer forecasting horizons. The diagonal elements $\sigma_{1,h}, \dots, \sigma_{m,h}$ are, therefore, drawn from an inverse-gamma distribution according to

$$\begin{aligned} \sigma_{i,h} \sim \mathcal{IG}(k_1/2, l_1/2), \quad \text{where} \quad k_1 &= k_0 + n \\ l_1 &= l_0 + \hat{\mathbf{e}}_{i,T}(h)'\hat{\mathbf{e}}_{i,T}(h). \end{aligned}$$

The n-dimensional vector $\hat{\mathbf{e}}_{i,T}(h)$ represents the ex-ante known prediction errors for variable $i \in \{1, \ldots, m\}$, which have been obtained from the predictive distribution of the base forecasts. While this parsimonious approach has advantages when it comes to computational speed and more accurate forecasts at lower levels of the hierarchy, it is possible to estimate the full covariance matrix of historical base forecast prediction errors in the spirit of Wickramasuriya et al. (2019) or by ordering the draws from the independent predictive distributions following Jeon et al. (2019). Both approaches would then require sampling from an inverse-Wishart distribution and are likely to require additional shrinkage priors. For our parsimonious approach using an inverse-gamma distribution, we choose a weakly informative prior distribution with $k_0 = 3$ and $l_0 = 1$. This has negligible impact on the posterior distribution, but ensures that $\mathbf{\Sigma}_h$ is nonsingular in the case where a base forecast has no variation.

2.3 Bias Shrinkage

In order to align decisions with a specific base forecast, it may be of interest to selectively shrink some reconciliation biases towards zero. This is especially useful when there exists some prior knowledge that a particular base forecast contains additional information that is not reflected in other levels of the hierarchy. Since the covariance matrix of the prediction errors enters the model as a weighting matrix, it can be useful to impose prior restrictions on $\mathbf{\Sigma}_h$. This can be achieved using a time-invariant diagonal matrix $\mathbf{\Lambda}$ with m weights on its diagonal. For the conjugate inverse-gamma prior described in section 2.2, this implies a parameter choice of $k_0 = n$ and $l_0 = (\hat{\mathbf{e}}_{i,T}(h)'\hat{\mathbf{e}}_{i,T}(h))\lambda_i$, where λ_i is the weight of the ith variable according to the corresponding entry in $\mathbf{\Lambda}$. This can be interpreted as an empirical Bayes prior because it is defined using the known base forecast errors. In order to shrink the reconciliation bias towards zero, the corresponding weight $\mathbf{\Lambda}$ has to be less than one. At the same time, it is necessary to increase the remaining elements above one such that they are able to account for the higher reconciliation biases at their level of the hierarchy. This is achieved by constructing the weights such that the product of the diagonal elements of $\mathbf{\Lambda}$ remains constant at unity. As a result, the determinant and therefore the generalized variance of $\mathbf{\Lambda}\mathbf{\Sigma}_h\mathbf{\Lambda}'$ remains the same for each $\mathbf{\Lambda}$ (Mustonen, 1997).

Figure 2 demonstrates the impact of different prior assumptions on the estimated reconciliation biases. It features identical unreconciled forecasts of a simple hierarchy with $m = 3$ series, where $Y_A + Y_B = Y_0$. For each series a sample is drawn from the predictive forecasts density, assumed to be $\mathcal{N}(4, 2)$ for Y_A, $\mathcal{N}(6, 1)$ for Y_B, and $\mathcal{N}(16, 3)$ for Y_0. The horizontal axis shows draws from the unreconciled base forecasts, which are clearly incoherent. The vertical axis, on the other hand, shows the means of the reconciled forecasts. The diagonal line shows values where the means of base and reconciled forecasts are equal. For boxes above this diagonal line, reconciliation adjusts forecasts upwards. For boxes below the diagonal line, forecasts are adjusted downwards.

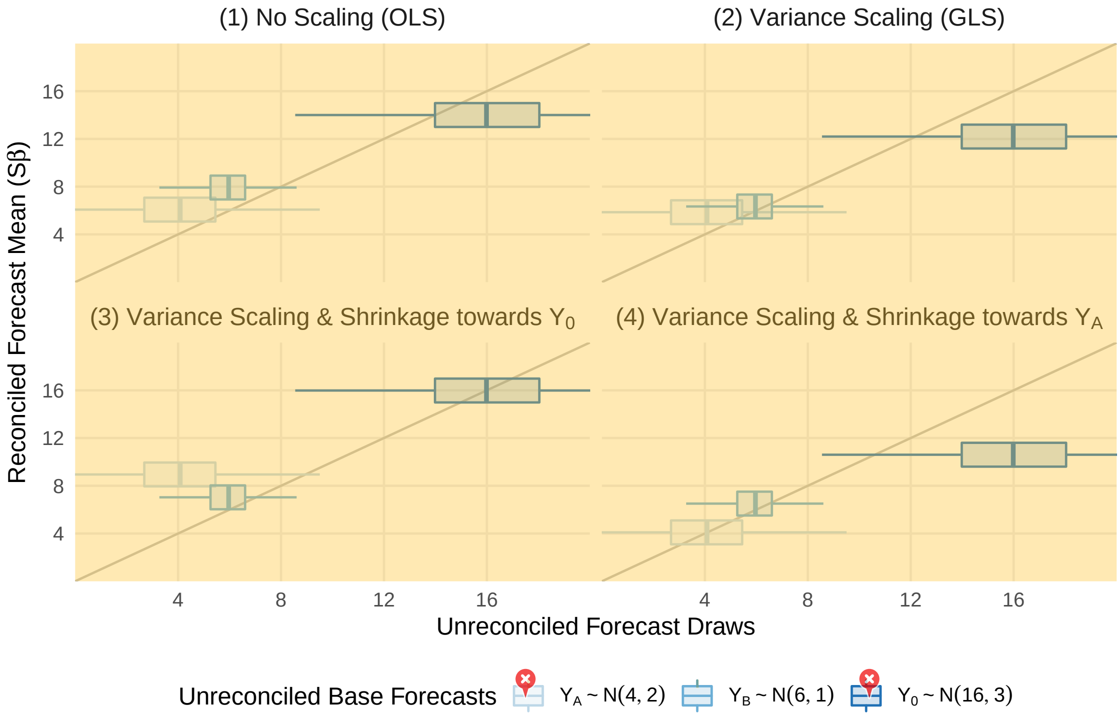

Figure 2: Prior Weighting Schemes. The 45° line indicates where the unreconciled base forecasts on the x-axis are equal to reconciled forecast means on the ordinate.

Each panel corresponds to a different prior choice for $\mathbf{\Sigma}_h$. Subfigure (1) shows that the forecast biases for each margin are treated equally in an ordinary least squares regression, consequently the means of Y_A and Y_B are adjusted upwards while the mean of Y_0 is adjusted downwards. Subfigure (2) shows reconciliation biases that are weighted with the inverse of their corresponding forecast variances. This leads to a smaller adjustment in Y_B (the reconciled and base means are close) relative the others since it is more accurate. Subfigures (3) and (4) shrink the reconciled forecasts of Y_0 and Y_A towards their base forecasts. There may exist prior information on the reliability of certain models or the requirement to fix some forecasts at specific values. This could be due to better data availability, higher suitability of a particular model or subjective judgment of the forecaster.

Besides the shrinkage of specific reconciliation biases towards zero, there are several other weighting methods conceivable. Section 4.4 provides empirical applications and highlights the usefulness of shrinkage reconciliation in operational forecasting.

3 Data

We use a comprehensive dataset containing exports of Swiss goods, collected by the Swiss Federal Customs Administration. All time series cover a period from 1988 to 2018 in monthly frequency and are denominated in Swiss francs. They are not adjusted for seasonalities or calendar effects and data revisions are usually very insignificant. The data can be grouped by export destination and product category. The geographic hierarchy consists of 8 regions, aggregated from 245 countries and dependent territories. The categorical hierarchy follows a national nomenclature covering 12 main economic groups and 48 subgroups. This leads to a grouped hierarchy with $m = 13,118$ series containing at least one nonzero entry of which $q = 9,483$ series are at the bottom level. Table 1 provides summary statistics on the series at each level. Further visualizations on the change in composition and a detailed statement of all categorical and geographical classifications can be found in Appendix A.3.

Table 1: Summary Statistics for Hierarchical Levels

Levels		No. of Series	Mean	Std. Dev.	Min	Max	IQR
geographical	categorical						
World	Total	1	12,129	-	-	-	-
World	Category	12	1,011	1,340	66	4,292	1,023
World	Subcategory	48	253	619	0	3,844	137
Region	Total	8	1,516	2,493	136	7,498	1,384
Region	Category	96	126	344	0	2,571	60
Region	Subcategory	377	32	146	0	2,271	9
Country	Total	245	50	212	0	2,527	14
Country	Category	2,848	4	29	0	683	0
Country	Subcategory	9,483	1	13	0	567	0

Notes: Summary statistics refer to the average monthly export volume in million Swiss francs of each series between 1988 and 2018.

All values shown refer to the invoiced price of the goods in Swiss francs, including transport and insurance costs as well as other expenditure up to the Swiss border. If the invoice is in a foreign currency, the invoiced amounts are converted using the previous day's exchange rate. As a result, the figures are affected by exchange rate fluctuations. However, Bonadio et al. (2020) document a quick exchange rate pass-through.

Figure 3 shows the historical development of the regional and categorical hierarchies. As a result of its status as a small open economy in a rapidly globalizing world, Swiss exports have increased significantly since the late 1980s. Accounting for more than half of total exports, Western Europe is a key market for Swiss goods. Increasingly larger shares of exports also go to North America and East Asia, with around 17% each in 2018. Exports to Africa and the Middle East, Latin America and the Caribbean, Central Asia and Eastern Europe, South Asia, Australia and Oceania account only for about 10% combined. The hierarchical grouping by categories is more evenly distributed, but has been subject to greater shifts in its composition. The most important categories are 'Chemicals and Pharmaceuticals', 'Precision Instruments' and 'Machines and Electronics'. The two hierarchical groupings are quite different. The geographic hierarchy widens towards the bottom, but with a majority of the export volume going to European countries it is nevertheless highly concentrated. The categorical hierarchy, on the other hand, has fewer subgroups and therefore remains narrow towards the bottom. Compared to the regional hierarchy, the export volume is however more evenly distributed.

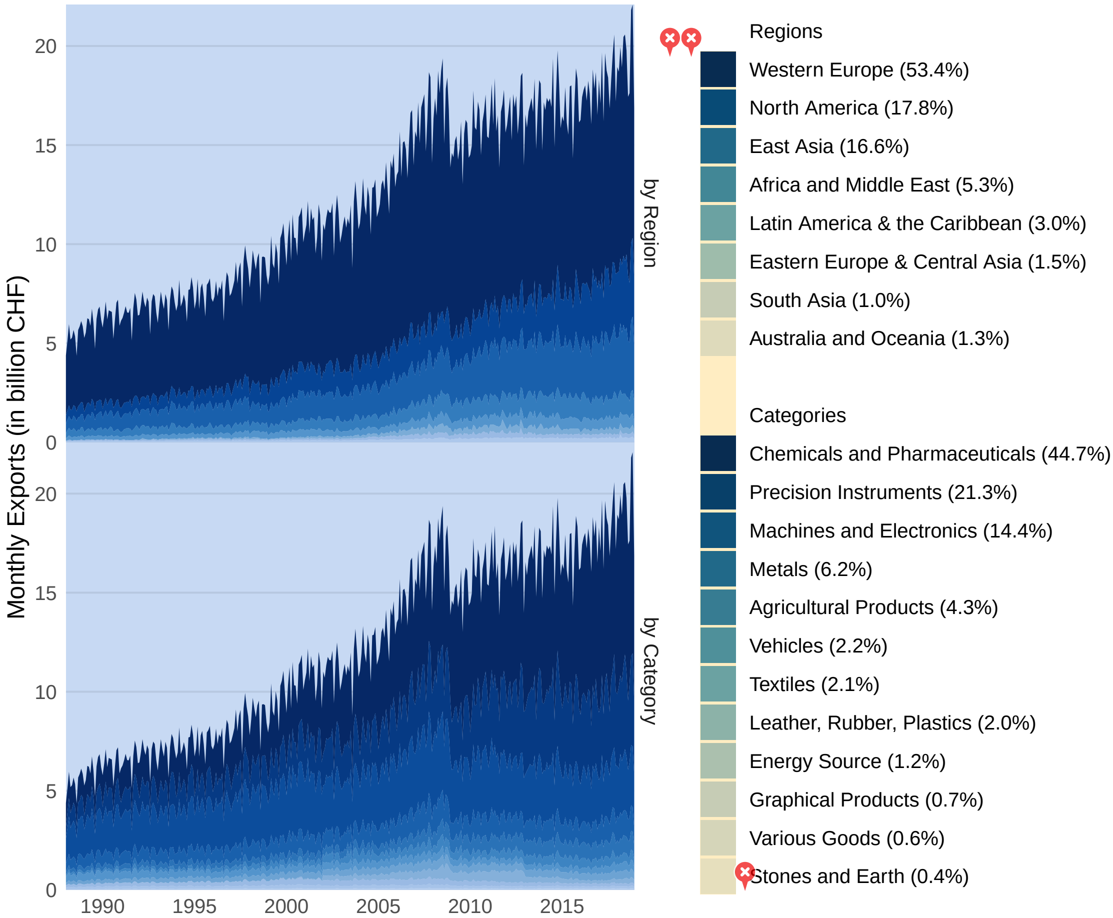

Figure 3: Contribution to Swiss Merchandise Exports of Goods. Monthly Swiss merchandise exports, denominated in billion Swiss francs, not adjusted for seasonalities or calendar effects. Average export shares of the year 2018 in parentheses.

A common assumption is that series at the top level of a hierarchy are easier to forecast. Due to the aggregation involved, they are usually less noisy and exhibit more predictable characteristics such as the strength of seasonality, trend, spectral entropy, and serial correlation. Kang et al. (2017) measure this by extracting a number of time series features from the data that are commonly associated with better predictability. They then construct a measure of predictability for each time series by estimating principal components from these features. Figure 4 shows the first principal component, which accounts for a large share of the variation in these predictability features. It is evident that there exists a strong correlation between predictability and export volume.

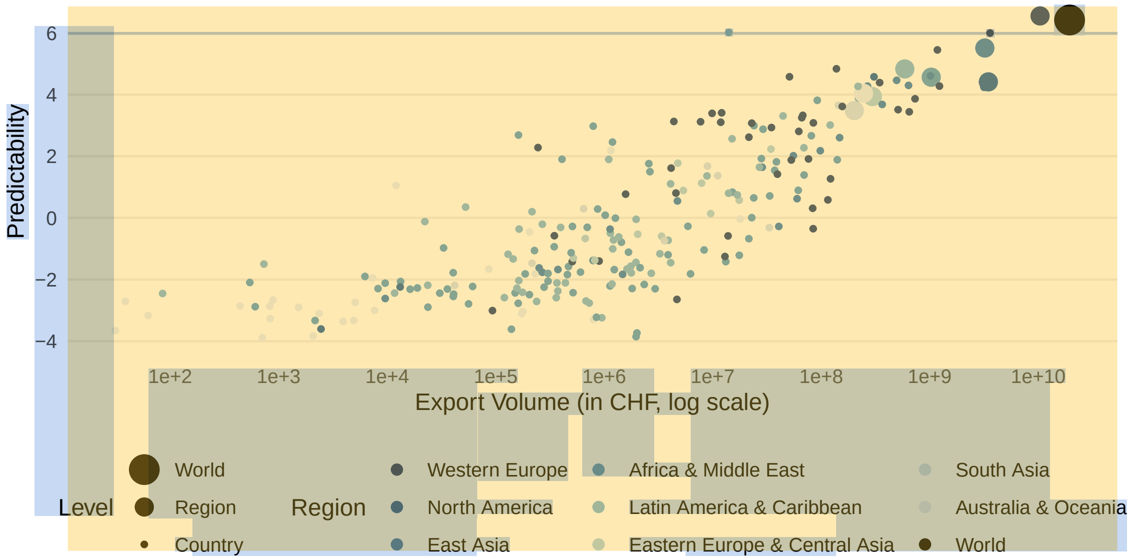

Figure 4: Predictability of Different Levels in a Hierarchy. Predictability is defined as the first principal component of a large number of time series features described in Kang et al. (2017).

4 Reconciliation of Export Forecasts

This section provides empirical evidence for the benefits of forecast reconciliation. It compares the performance of different reconciliation methods and explores which data characteristics profit in particular from hierarchical combination. The setup of our comprehensive pseudo-realtime forecasting exercise is described in the following.

A large hierarchy of Swiss goods exports is used to test the Bayesian reconciliation framework and various competing methods. In each month from 1995 to 2015, forecasts for all series in the hierarchy are calculated for the next 36 months. For each of the 13,118 series, we use a few univariate benchmark models to get the base forecasts. It should be noted that this is not necessarily the best model choice since we do not take into account important explanatory variables such as exchange rates, relative prices or global economic developments. Our focus is to show the advantages of reconciliation for forecasting accuracy, which we do by comparing reconciled predictions to unreconciled base forecasts. Therefore, the method used to create the base forecast is not of importance. We use an autoregressive integrated moving average model (ARIMA), an exponential smoothing state-space model (ETS), and a seasonal random walk model (RW). As described in Hyndman and Khandakar (2008), the model for each series is parameterized automatically based on the Akaike Information Criterion. In order to get samples from the predictive densities, $n = 1000$ sample paths are simulated from each fitted model using Gaussian errors. With the exception of the volatile period during the Great Recession, the ARIMA and ETS approaches outperform the random walk on average for series at every level and forecasting horizon. All results in the following subsection will, therefore, rely on ARIMA forecasts.[2]

These incoherent forecasts are then reconciled using several basic single-level and optimal combination methods. The single-level techniques include bottom-up, top-down and middle-out methods. The latter two can only be used for non-grouped time series and are, therefore, tested on the regional and categorical hierarchies separately. The combination methods used are ordinary least squares (no scaling), weighted least squares with variance scaling, weighted least sqared with structural scaling, MinT and the Bayesian state-space reconciliation framework (BSR). If aggregation of the prediction errors is necessary, they are weighted with their respective export share. The reconciled forecasts are then compared to the corresponding unreconciled predictions using log relative root mean squared forecast errors (Hyndman and Koehler, 2006).

4.1 Comparison of Reconciliation Methods

Figure 5 shows the accuracy of all forecasts, defined as the log of the root mean squared forecasting error of the base forecasts relative to the mean squared errors of the coherent forecasts from each method. Values above zero indicate, therefore, a better forecast performance. It is worth noting that reconciliation methods and bottom-up forecasts are the only techniques that allow for coherence across all levels of a grouped hierarchy. Top-down and middle-out reconciliations are not applicable in the case of grouped time series.

[2] A detailed description of the methodology and comparisons of forecasting methods, horizons and accuracy measures can be found in Appendix A.2.

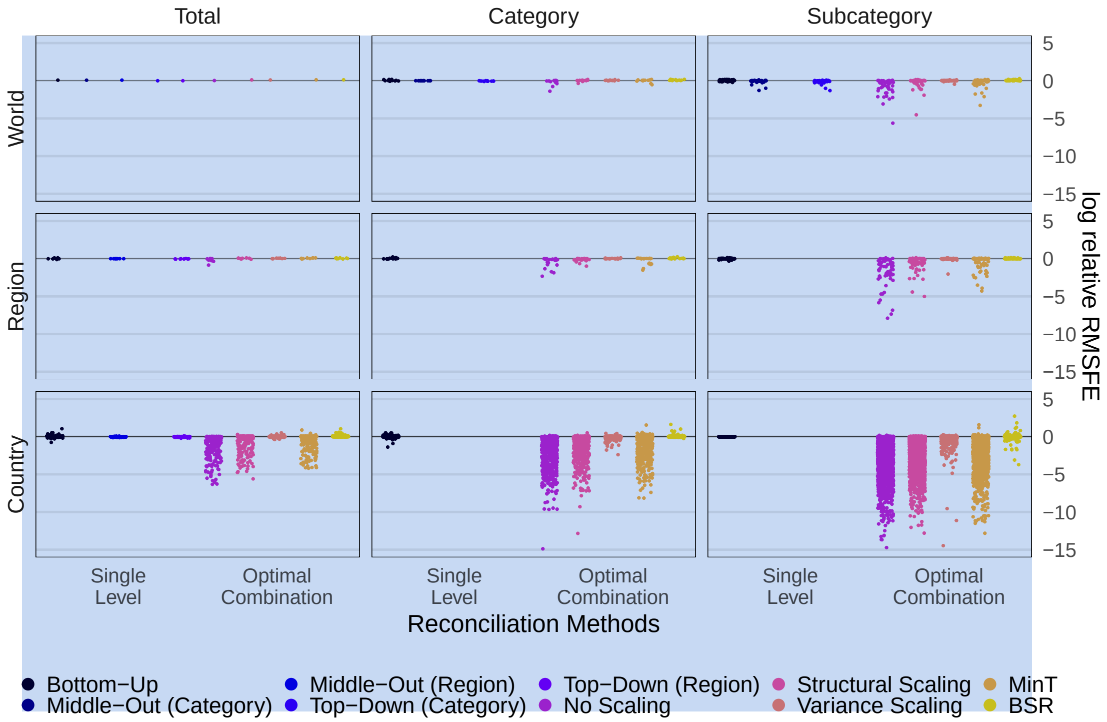

Figure 5: Relative Accuracy of Reconciliation Methods. Values above zero indicate higher forecast accuracy relative to the unreconciled case. Forecast accuracy is given by the log of the root mean squared forecast errors of the unreconciled forecasts relative to the reconciled forecasts. Average of all forecast dates and horizons.

Figure 6 shows that some forecasts do not benefit from reconciliation. Especially for the bottom-level series, combination methods appear to decrease forecasting performance. Variance scaling and BSR are less affected by this deterioration of forecast accuracy at lower levels, probably as a result of their parsimonious parameter choice. In order to demonstrate the benefits of hierarchical combination, Figure 6 shows average log relative RMSFEs by weighting the prediction errors at intermediate and lower levels using the corresponding share in total export volume. It is evident that single-level methods do not consistently improve forecasting accuracy. The bottom-up and middle out methods fare reasonably well for the top level series, but fail to outperform the unreconciled forecasts at lower levels and are sometimes even significantly worse. Optimal combination, on the other hand, tends to outperform the base forecasts especially for top and intermediate-level series. Especially variance scaling, MinT and BSR work well at all levels.

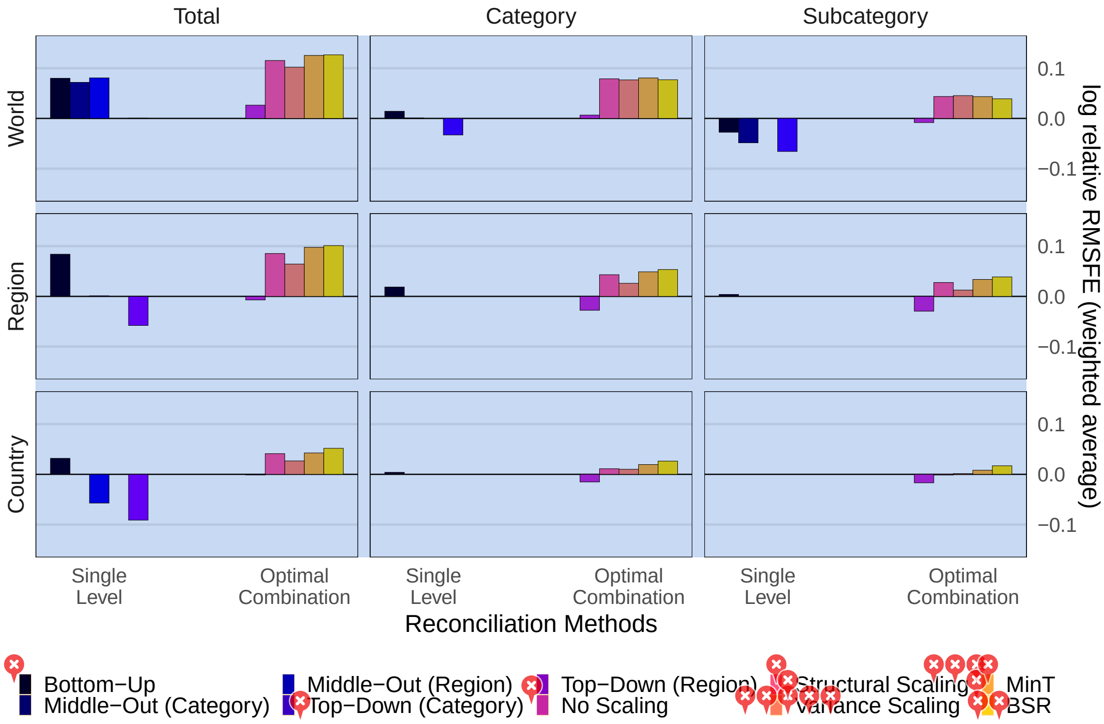

Figure 6: Average Relative Accuracy of Reconciliation Methods. Values above zero indicate higher forecast accuracy relative to the unreconciled case. Forecast accuracy is given by the log of the root mean squared forecast errors of the unreconciled forecasts relative to the reconciled forecasts. Errors at intermediate and lower levels are weighted using their corresponding share in total export volume. Average of all forecast dates and horizons.

It is also instructive to look at the development of the relative forecasting accuracy over time in Figure 7. Even though the combination methods are more accurate on average, they do not consistently outperform the unreconciled forecasts. While the combination using no scaling does not beat the unreconciled benchmark, the remaining methods perform better and fairly similar over time. For the top level series, the benefits of reconciliation accrue mostly during times of global economic distress and corresponding appreciations of the Swiss franc. This is due to the fact that the simpler models at lower levels provide stability at times when the top level model is biased. The biggest gains can be observed during the early 2000s recession following the burst of the dot-com bubble, the global financial crisis and the following sovereign debt crisis in Europe, and the sudden appreciation of the Swiss franc after the Swiss National Bank stopped supporting the currency peg to the Euro in early 2015. Interesting is the forecasting accuracy after January 2002, when electrical energy was reclassified as a good instead of a service. The structural break in the time series leads to misspecified models, but the rigid structure imposed by the hierarchy increases forecast accuracy substantially relative to the unreconciled case.

Figure 7: Relative Accuracy of Combination Methods over Time. Values above zero indicate higher forecast accuracy relative to the unreconciled case. Forecast accuracy is given by the log of the root mean squared forecast errors of the unreconciled forecasts relative to the reconciled forecasts. Errors at intermediate and lower levels are weighted using their corresponding share in total export volume. Average of all forecast horizons.

4.2 Comparison of Forecasting Horizons

In order to check whether the accuracy improvements are significant, we test for equality of the mean squared errors of reconciled and unreconciled forecasts. Since it is not possible to get unreconciled forecasts by constraining the parameter space of the reconciliation procedure, the comparison involves non-nested models. Following Clark and McCracken (2013), we therefore rely on the test statistic proposed by Diebold and Mariano (1995), using the variance correction suggested by Harvey et al. (1997). It accounts for serial correlation in the squared error loss while testing for significance in the difference between two squared forecast errors at various forecasting horizons. Table 2 shows the p-values for the one-sided test, where the alternative hypothesis is that the accuracy of reconciliation methods is greater.

Table 2: Tests for Predictive Accuracy

	Entire Sample 1998 - 2018			Moderate Period 1998 - 2006			Crisis and Recovery 2007 - 2018		
	12	24	36	12	24	36	12	24	36
Single-Level									
Bottom Up	0.49	0.28	0.29	0.84	0.86	0.67	0.10	0.12	0.14
Middle Out (Category)	0.14	0.05*	0.17	0.28	0.13	0.09	0.19	0.16	0.33
Middle Out (Region)	0.02*	0.04*	0.01*	0.08	0.15	0.05	0.08	0.10	0.02*
Top Down (Category)	0.11	0.14	0.13	0.16	0.15	0.14	0.16	0.14	0.08
Top Down (Region)	0.14	0.14	0.84	0.07	0.15	0.16	0.14	0.12	0.92
Optimal Combination									
No Scaling	0.00**	0.00**	0.01*	0.05	0.08	0.03*	0.01*	0.01**	0.02*
Structural Scaling	0.05*	0.04*	0.09	0.37	0.40	0.19	0.00**	0.02*	0.06
Variance Scaling	0.02*	0.01*	0.05*	0.22	0.15	0.07	0.00**	0.01**	0.05*
MinT	0.03*	0.01**	0.08	0.14	0.10	0.13	0.04*	0.03*	0.11
BSR	0.14	0.08	0.12	0.57	0.62	0.25	0.01**	0.04*	0.09

Notes: Table shows p-values of one-sided Diebold Mariano tests. They are retrieved by testing differences between reconciled and unreconciled forecast errors, using the top level series and forecasting horizons of 12, 24 and 36 months. Significance indicated by *** $p < 0.001$, ** $p < 0.01$ and * $p < 0.05$.

With the exception of the middle-out approaches, single-level methods are not significantly more accurate than the unreconciled forecasts at all horizons. Optimal combinations are associated with lower p-values, in particular for the period of increased economic volatility after the Great Recession. Parsimonious approaches, in particular no scaling or variance scaling, appear to perform best even at longer forecasting horizons. There is however no evidence that the use of predictive distributions and state-dependent reconciliation errors leads to significant gains.

4.3 Comparison of Hierarchical Levels

Another way to dissect the results is to identify which time series see the greatest gains in forecast accuracy from using reconciliation. Figure 8 provides an overview of the relative forecast accuracy by geographic classification, using the Bayesian reconciliation framework. It is again obvious that reconciled forecasts are on average more accurate than in the unreconciled case, but not in every instance. It appears that series with a larger export volume benefit most from reconciliation. Forecasts of exports to countries in Europe, North America and East Asia are almost entirely better off than in the unreconciled case, whereas forecasts of exports to countries with a lower share, such as the islands in Oceania, tend to be worse off. In addition, a top-level time series doesn't necessarily benefit more from reconciliation than a time series at lower levels. The same results also hold true for the relative forecast accuracy by categories, as shown in Figure 9. Because the export shares in the categorical hierarchy are more evenly distributed, the pattern of smaller export volumes being worse off due to reconciliation is less pronounced. Results for other variance scaling methods such as weighted least squares and MinT are similar.

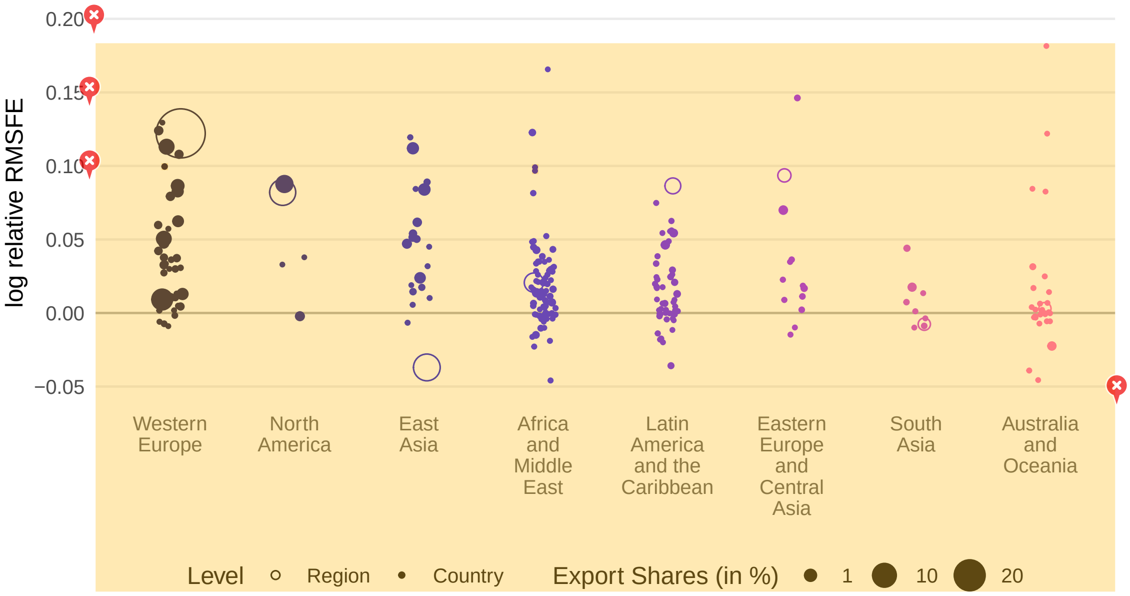

Figure 8: Relative Accuracy of Reconciliation Methods by Regions. Values above zero indicate higher forecast accuracy relative to the unreconciled case. Forecast accuracy is given by the log of the root mean squared forecast errors of the unreconciled forecasts relative to the reconciled forecasts. Average of all forecast dates and horizons. Reconciliation using unweighted Bayesian state-space reconciliation.

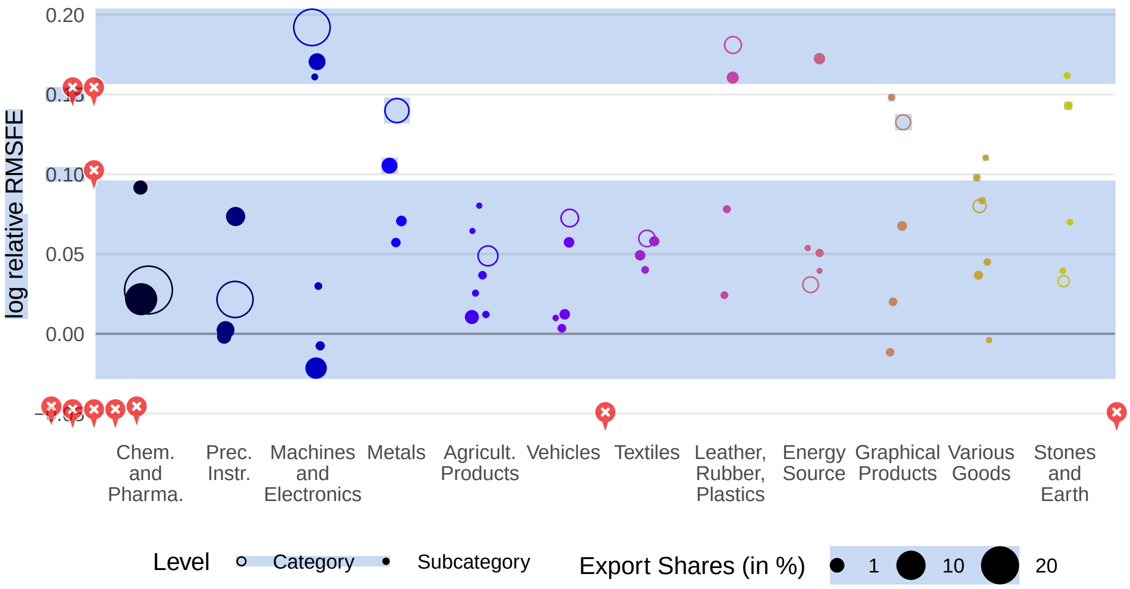

Figure 9: Relative Accuracy of Reconciliation Methods by Categories. Values above zero indicate higher forecast accuracy relative to the unreconciled case. Forecast accuracy is given by the log of the root mean squared forecast errors of the unreconciled forecasts relative to the reconciled forecasts. Average of all forecast dates and horizons. Reconciliation using unweighted Bayesian state-space reconciliation.

4.4 Benefits for Aligned Decision-Making

Reconciliation techniques are useful for complex operational structures because base forecasts can be generated independently at all hierarchical levels. However, it is usually costly to incorporate specific adjustments into all base forecasts as they rely on different methods and information sets. For instance, some predictions may contain managerial decisions that are difficult to incorporate into models at other levels of the hierarchy. Because reconciliation procedures minimize the distance between coherent and incoherent forecasts, it often occurs that the judgmental adjustments to a specific forecast are diluted because they are not reflected in other base forecasts. The reconciled forecasts allow then for aligned decision-making, but do not fully reflect the judgmental adjustments to a few specific base forecasts. An advantage of the generalized weighting scheme proposed in section 2.3 is that the reconciliation errors can be targeted directly. This allows the forecaster to shrink certain reconciled forecasts towards their base forecast.

An example for the usefulness of shrinkage reconciliation is given by the reclassification of electricity as a good instead of a service in foreign trade statistics (see also section 4.1). Starting in early 2002, this change in accounting standards increased the share of energy sources from almost 0% to more than 2% of total merchandise exports. Figure 10 shows various forecasting scenarios for exports of energy sources. All predictions are based on a training sample that includes historical data up the first observation with the new accounting standard.

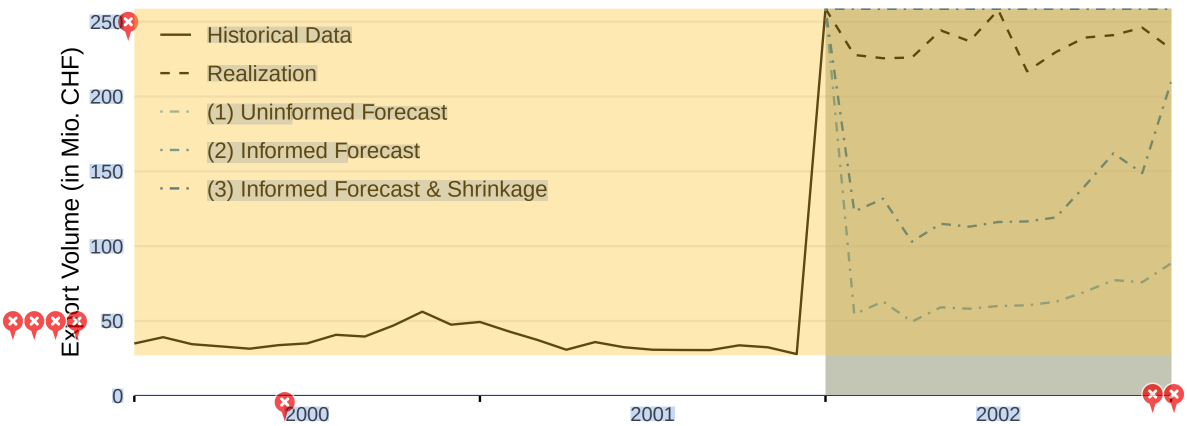

Figure 10: Forecasting Scenarios for Exports of Energy Sources. Monthly exports of energy sources from 2000 to 2002 in million Swiss francs. The structural break in January 2002 is caused by the reclassification of electrical energy as a good instead of a service.

Scenario (1) shows the forecast for exports of energy sources using the same model as before the structural change. It is evident that it fails to capture the structural break and adjusts very slowly as more observations pour in. Scenario (2) assumes a judgmental decision to use a random walk forecast after observing the increase in export volumes in January. Even though the forecaster has prior knowledge that a random walk forecast is appropriate for this series, the models at other levels assume the structural break to be an outlier. They dominate any information from the random walk forecast and it is shifted downwards during the reconciliation procedure. The only solution would be an adjustment of the base forecasts for all other series, which is cumbersome given the complex categorical and geographical data structure. Scenario (3) relies on a random walk forecast as well, but uses shrinkage reconciliation to put more weight on this particular forecast. This forces the reconciled prediction to stay close to its base forecast, which provides a reasonable projection for exports of energy sources in 2002.

Since the reconciliation procedure distributes the coherency errors across the entire hierarchy, this also leads to some accuracy gains at other levels. More importantly, it allows managers to align their decisions with other units without sacrificing forecast accuracy. Shrinkage reconciliation thus improves alignment of decisions across complex hierarchies by allocating more weight to predictions that managers are confident in.

Besides the shrinkage of specific reconciliation biases towards zero, there are several other weighting methods conceivable. Possible approaches include the weighting of each series by its level in the hierarchy or by to the number of series at each node in the hierarchy. This allows for the emulation of the 'structural scaling', 'bottom-up', 'middle-out' and 'top-down' approaches. This is particularly relevant for judgmental adjustments of forecasts in complex operational structures. An important example is the case where forecasts at the top level are assumed to be correct because they contain information on managerial decisions that is not available to forecasters at lower units. As a result, all remaining base forecasts should be reconciled such that they are coherent with the base forecasts at the strategic level. Another common example is the case where forecasts at the most disaggregate, operational level are trusted more and predictions are aggregated from the bottom level.

We illustrate these two cases by reconciling a set of predictions using the appropriate single-level methods, namely the top-down and bottom-up reconciliation, and the Bayesian shrinkage approach. Table 3 shows the deviations from the base forecasts in mean absolute percentage errors for each of the three methods. Deviations between base forecasts and state-space reconciliations without any shrinkage serve as comparison.

Table 3: Deviations from Base Forecasts

Target Level	Evaluation Level	Appropriate Methods			Benchmark
		Top-Down	Bottom-Up	Shrinkage	
Top Level	Total / World	0		0	0.6
	Total / Region	37.2		2.4	1.0
	Category / World	17.4		2.8	1.1
	Category / Region	-		3.0	11.6
Bottom Level	Total / World		2.5	2.5	0.6
	Total / Region		2.8	2.8	1.0
	Category / World		3.6	3.6	1.1
	Category / Region		0	0	11.6

Notes: Deviations from the base forecast are measured using mean absolute percentage errors. Predictions are generated from random walk forecasts with a horizon of 12 months for each month from 1998 to 2018. Results from state-space reconciliation without shrinkage serve as benchmark.

Table 3 highlights a key contribution of the proposed framework, namely that it can gear the reconciled predictions towards any desired base forecast. It replicates all results from the 'bottom-up' approach, which is intuitive given that the upper levels are aggregated from the bottom-level base forecasts. The shrinkage approach extends the scope of the 'middle-out' and 'top-down' methods because they cannot be used for grouped hierarchies. Table 3 also shows that lower reconciliation biases for some forecasts comes at the cost of reconciled predictions being further away from their base forecasts at other levels.

5 Conclusion

This paper extends the literature on hierarchical forecast combination and aligned decision-making by introducing an explicit definition of state-dependent reconciliation biases. This allows for the joint reconciliation of all forecast periods and combines information on the coherence errors across the entire forecasting horizon. The Bayesian framework furthermore allows for the incorporation of prior information on the parameters, which enables the forecaster to introduce subjective judgment into the reconciliation. In addition, informative priors avoid some issues such as the occurrence of negative reconciled forecasts and singular forecast error covariance matrices. The use of predictive densities instead of past forecast errors allows for greater flexibility in the choice of the base forecast models, taking, for instance, conditional heteroskedasticity into account when weighting the forecasts at different horizons. However, the approach tends to be slower than established reconciliation techniques because it requires simulation of the joint posterior distribution using Gibbs sampling.

Using a comprehensive hierarchical dataset of Swiss goods exports, we demonstrate that optimal combination methods improve the forecasting accuracy significantly compared to the unreconciled case and simpler reconciliation methods. While the 'MinT' approach shows the largest improvements at aggregate levels, more parsimonious approaches such as variance scaling or Bayesian state-space reconciliation improve the accuracy also at more disaggregate levels. Even though reconciled forecasts are significantly more accurate on average, no reconciliation method consistently outperforms the unreconciled forecasts across the hierarchy or over time. Forecasts at aggregate levels tend to benefit more from reconciliation than noisy series at the bottom of a hierarchy. At the same level, forecasts that account for a larger share of the total are on average more accurate after reconciliation. Optimal combination methods are shown to be particularly useful in the case of misspecified models and during periods of high volatility in the time series. In addition, state-space methods increase the robustness of forecasts at the operational level, thus increasing the usefulness of bottom-level forecasts for decision-making.

The proposed method is useful for operational forecasting. Instead of using predictive accuracy as the only objective, it enables the practictioner to introduce judgmental adjustments into the reconciliation procedure. This allows, for instance, to assign more weight to forecasts that managers are confident in. Furthermore, the shrinkage approach also works for complex hierarchical data structures which are common in operational forecasting. It provides, for example, coherent predictions for sales data that can be aggregated according to product category, market segment, sales outlet, trade mark and other characteristics. Possible applications extend, of course, beyond demand forecasting to any hierarchical data that organizations base their decisions on.

References

Amisano, G. and Geweke, J. (2017). Prediction Using Several Macroeconomic Models. *Review of Economics and Statistics*, 99(5):912–925.

Ando, T. and Zellner, A. (2010). Hierarchical Bayesian Analysis of the Seemingly Unrelated Regression and Simultaneous Equations Models Using a Combination of Direct Monte Carlo and Importance Sampling Techniques. *Bayesian Analysis*, 5(1):65–96.

Athanasopoulos, G., Ahmed, R. A., and Hyndman, R. J. (2009). Hierarchical Forecasts for Australian Domestic Tourism. *International Journal of Forecasting*, 25(1):146–166.

Athanasopoulos, G., Hyndman, R. J., Kourentzes, N., and Petropoulos, F. (2017). Forecasting with Temporal Hierarchies. *European Journal of Operational Research*, 262(1):60–74.

Bergmeir, C., Hyndman, R. J., and Benítez, J. M. (2016). Bagging Exponential Smoothing Methods using STL Decomposition and Box-Cox Transformation. *International Journal of Forecasting*, 32(2):303–312.

Bonadio, B., Fischer, A. M., and Sauré, P. (2020). The Speed of Exchange Rate Pass-Through. *Journal of the European Economic Association*, 18(1):506–538.

Chan, J. and Jeliazkov, I. (2009). Efficient Simulation and Integrated Likelihood Estimation in State Space Models. *International Journal of Mathematical Modelling and Numerical Optimisation*, 1(1-2):101–120.

Clark, T. and McCracken, M. (2013). *Advances in Forecast Evaluation*, volume 2. Elsevier B.V.

Conflitti, C., Mol, C. D., and Giannone, D. (2015). Optimal Combination of Survey Forecasts. *International Journal of Forecasting*, 31(4):1096–1103.

Diebold, F. X. and Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Farebrother, R. W. (1978). Partitioned Ridge Regression. *Technometrics*, 20(2):121–122.

Gross, C. W. and Sohl, J. E. (1990). Disaggregation Methods to Expedite Product Line Forecasting. *Journal of Forecasting*, 9(3):233–254.

Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the Equality of Prediction Mean Squared Errors. *International Journal of Forecasting*, 13(2):281–291.

Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., and Shang, H. L. (2011). Optimal Combination Forecasts for Hierarchical Time Series. *Computational Statistics and Data Analysis*, 55(9):2579–2589.

Hyndman, R. J. and Athanasopoulos, G. (2018). Forecasting Hierarchical or Grouped Time Series. In *Forecasting: Principles and Practice*.

Hyndman, R. J. and Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software*, 27(3):1–22.

Hyndman, R. J. and Koehler, A. B. (2006). Another Look at Measures of Forecast Accuracy. *International Journal of Forecasting*, 22(4):679–688.

Hyndman, R. J., Lee, A. J., and Wang, E. (2016). Fast Computation of Reconciled Forecasts for Hierarchical and Grouped Time Series. *Computational Statistics and Data Analysis*, 97:16–32.

Jeon, J., Panagiotelis, A., and Petropoulos, F. (2019). Probabilistic forecast reconciliation with applications to wind power and electric load. *European Journal of Operational Research*, 279(2):364–379.

Kang, Y., Hyndman, R. J., and Smith-Miles, K. (2017). Visualising Forecasting Algorithm Performance Using Time Series Instance Spaces. *International Journal of Forecasting*, 33(2):345–358.

Kapetanios, G., Mitchell, J., Price, S., and Fawcett, N. (2015). Generalised Density Forecast Combinations. *Journal of Econometrics*, 188(1):150–165.

Kourentzes, N. and Athanasopoulos, G. (2019). Cross-Temporal Coherent Forecasts for Australian Tourism. *Annals of Tourism Research*, 75:393–409.

Mustonen, S. (1997). A Measure for Total Variability in Multivariate Normal Distribution. *Computational Statistics and Data Analysis*, 23(3):321–334.

Pennings, C. L. and van Dalen, J. (2017). Integrated Hierarchical Forecasting. *European Journal of Operational Research*, 263(2):412–418.

Primiceri, G. E. (2005). Time Varying Structural Vector Autoregressions and Monetary Policy. *Review of Economic Studies*, 2005(72):821–852.

Stock, J. H. and Watson, M. W. (2006). Forecasting with Many Predictors. *Handbook of Economic Forecasting*, 1(05):515–554.

Timmermann, A. (2006). Forecast Combinations. *Handbook of Economic Forecasting*, 1(05):135–196.

Wickramasuriya, S. L., Athanasopoulos, G., and Hyndman, R. J. (2019). Forecasting Hierarchical and Grouped Time Series Through Trace Minimization. *Journal of the American Statistical Association*, 114(526):804–819.

A Appendix

A.1 Identification of the Reconciliation Errors

The regression in equation (3) is essentially an ill-posed problem since the predictor variables are multicollinear design matrices constructed from ones and zeros. It is therefore necessary to impose additional restrictions on the reconciliation biases $\boldsymbol{\alpha}_h$ in order to achieve identification.

Following Farebrother (1978), the regression is partitioned in order to separate the parameters that cause multicollinearity. The conditional distribution of $\boldsymbol{\alpha}_h$ can be expressed equivalently by concentrating out $\boldsymbol{\beta}_h$ in the following reconciliation identity.

$$\boldsymbol{\alpha}_h = \hat{\mathbf{y}}_T(h) - \mathbf{S}\boldsymbol{\beta}_h. \tag{9}$$

In order to eliminate $\boldsymbol{\beta_h}$ from equation (9), both sides are multiplied by the standard generalized least squares projection matrix $\mathbf{P}_h = \mathbf{S}(\mathbf{S}'\boldsymbol{\Sigma}_h^{-1}\mathbf{S})^{-1}\mathbf{S}'\boldsymbol{\Sigma}_h^{-1}$, which results in

$$\mathbf{P}_h\,\boldsymbol{\alpha}_h = \mathbf{P}_h\,\hat{\mathbf{y}}_T(h) - \mathbf{S}\boldsymbol{\beta}_h. \tag{10}$$

This implies an orthogonal projection onto the coherent subspace subject to the weighting matrix $\boldsymbol{\Sigma_h}$. The resulting terms are then subtracted from both sides of equation (9), which gets rid of $\boldsymbol{\beta}_h$:

$$(\mathbf{I}_m - \mathbf{P}_h)\,\boldsymbol{\alpha}_h = (\mathbf{I}_m - \mathbf{P}_h)\,\hat{\mathbf{y}}_T(h). \tag{11}$$

It is useful to define the idempotent residual maker $\mathbf{M}_h = \mathbf{I}_m - \mathbf{P}_h$. Since $\mathbf{M}_h$ is not invertible due to the presence of multicollinearity, equation (11) cannot be solved for $\boldsymbol{\alpha}_h$. Our identifying assumption is that $\boldsymbol{\alpha}_h$ lies in the span of $\mathbf{M}_h$ in which case $\mathbf{M}_h\boldsymbol{\alpha}_h = \boldsymbol{\alpha}_h$. This solves the identification problem and leaves the reconciliation biases as a function of the data and the residual maker $\mathbf{M}_h$:

$$\boldsymbol{\alpha}_h = \mathbf{M}\hat{\mathbf{y}}_T(h). \tag{12}$$

This result is again intuitive since the reconciliation biases are the residuals from a regression of the base forecasts on the aggregation matrix.

A.2 Robustness

This section provides additional robustness checks and contrasts the results found in the main part with the results obtained from different base forecast models and other forecasting metrics, as suggested in Hyndman and Koehler (2006).

Figure 11 shows the mean absolute scaled forecast errors (MASE) for all methods, averaged across all forecast dates and horizons. It mirrors several results obtained from relative RMSFEs seen in the main part. For instance, optimal combination methods seem to have, on average, lower forecast errors than the unreconciled benchmark, indicated by the horizontal line. This holds true especially for higher levels. Single-level methods are often less accurate than the benchmark. In addition, top-down and middle-out methods are not able to reconcile grouped hierarchies. At the bottom level, optimal combination methods are, on average, not as accurate as the unreconciled forecasts when using the MASE as a metric for forecasting accuracy. This stands in contrast with the results obtained from the RMSFE, where most combination methods improve upon the accuracy of unreconciled forecasts at the lowest level.

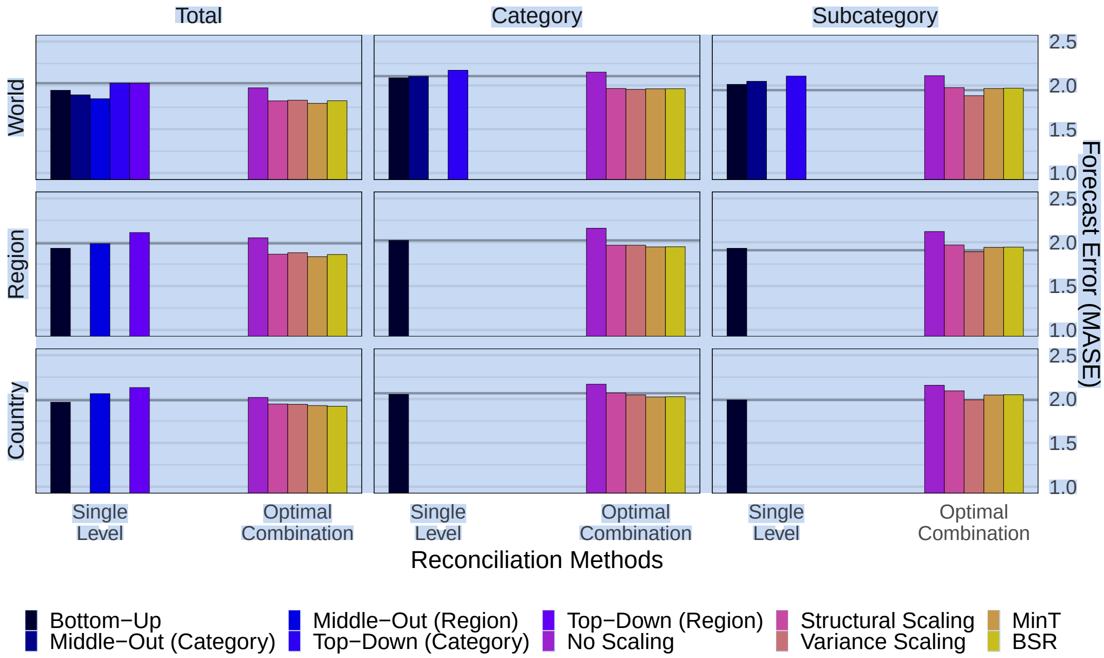

Figure 11: Mean Absolute Scaled Error of Reconciliation Methods. Lower bars indicate more accurate forecasts. Horizontal lines show the mean absolute scaled error of unreconciled predictions. Errors at intermediate and bottom levels are weighted using their corresponding share in total export volume. Average of all forecast dates and horizons.

Figure 12 evaluates the mean absolute percentage error (MAPE) of reconciled forecasts and compares the results to the unreconciled benchmark, again highlighted using a horizontal line. While the results are here less obvious than in the RMSFE or MASE case, they still point towards the same conclusions: Single-level methods, on average, do not outperform the accuracy of unreconciled predictions. Combination methods on the other hand, and in particular Mint and BSR, have lower forecast errors.

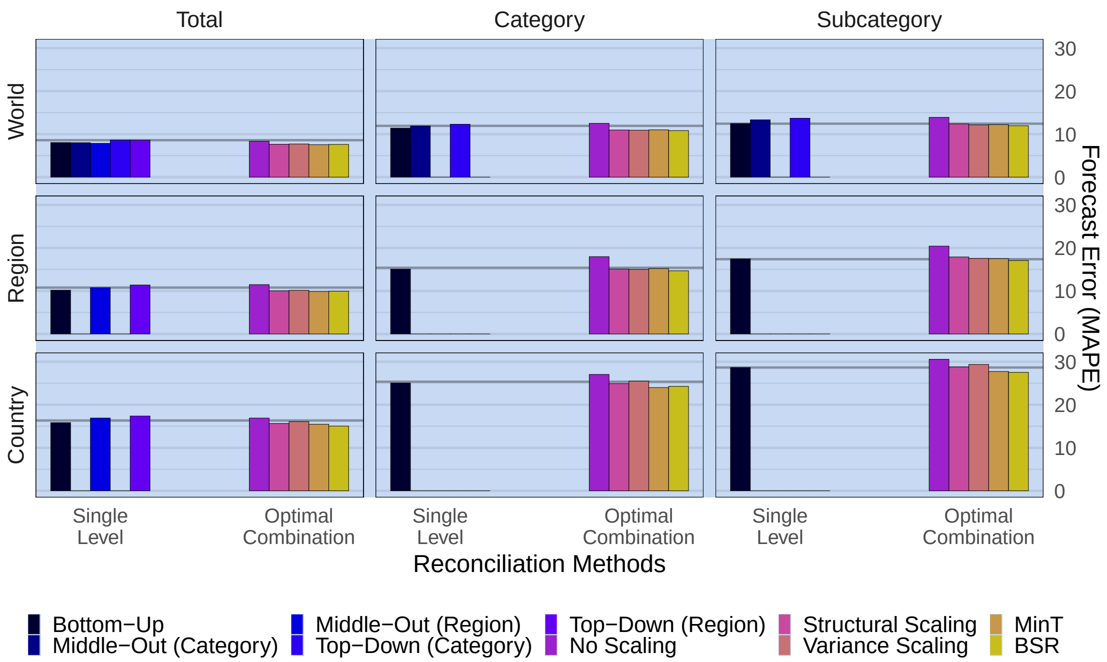

Figure 12: Mean Absolute Percentage Error of Reconciliation Methods. Lower bars indicate more accurate forecasts. Horizontal lines show the mean absolute percentage error of unreconciled predictions. Errors at intermediate and bottom levels are weighted using their corresponding share in total export volume. Average of all forecast dates and horizons.

As the forecasting exercises in Section 4 rely on base forecasts from ARIMA models, it is necessary to also consider alternative forecasting methods. However, as the focus of this paper is on the advantages of reconciliation rather than export forecasts, the quality of the underlying base forecasts is not of great importance. We use a few established methods that are already implemented in the 'hts' and 'forecast' packages for R (Hyndman and Khandakar, 2008; Hyndman et al., 2011, see for instance).

Figure 13 shows the forecast accuracy of ARIMA models as well as exponential smoothing state-space models (ETS) and seasonal random walk models (RW). It shows the average accuracy at each forecasted periods from 1995 to 2015 for the top level series, measured using all previously established metrics. While there are some differences in the accuracy of the different forecasting methods, the accuracy measures show a very similar picture. The random walk forecast is very volatile and generally doesn't outperform the ARIMA and ETS predictions, with the notable exception of the period during the Great Recession. ARIMA and ETS perform very similarly.

Figure 14 shows the corresponding forecast accuracy measures for the bottom level, where the prediction errors are weighted using their corresponding export shares. There appear to be no significant differences in the accuracy of the forecasting methods at the bottom level. There is a noticable shift in the root mean squared (weighted) errors, which can be explained with the shifting composition of export volumes at the lowest level. However, this does not pose a problem for the analysis in the main text as we rely on the logarithmized ratio of reconciled and unreconciled RMSFEs. Interestingly, there is no strong deterioration of forecast accuracy noticeable during the Great Recession. This might indicate that prediction of bottom level series are always rather poor and not just during crises.

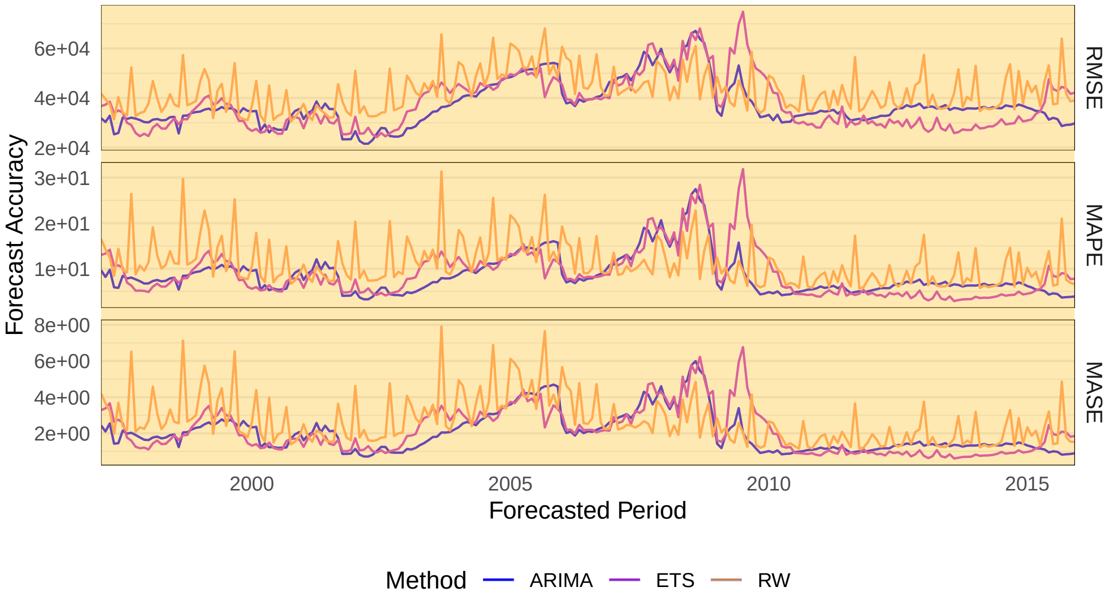

Figure 13: Accuracy of Forecasting Methods at the Top Level. Figure provides several measures of forecast accuracy for the top-level series from 1995 to 2015. Average of all forecast horizons.

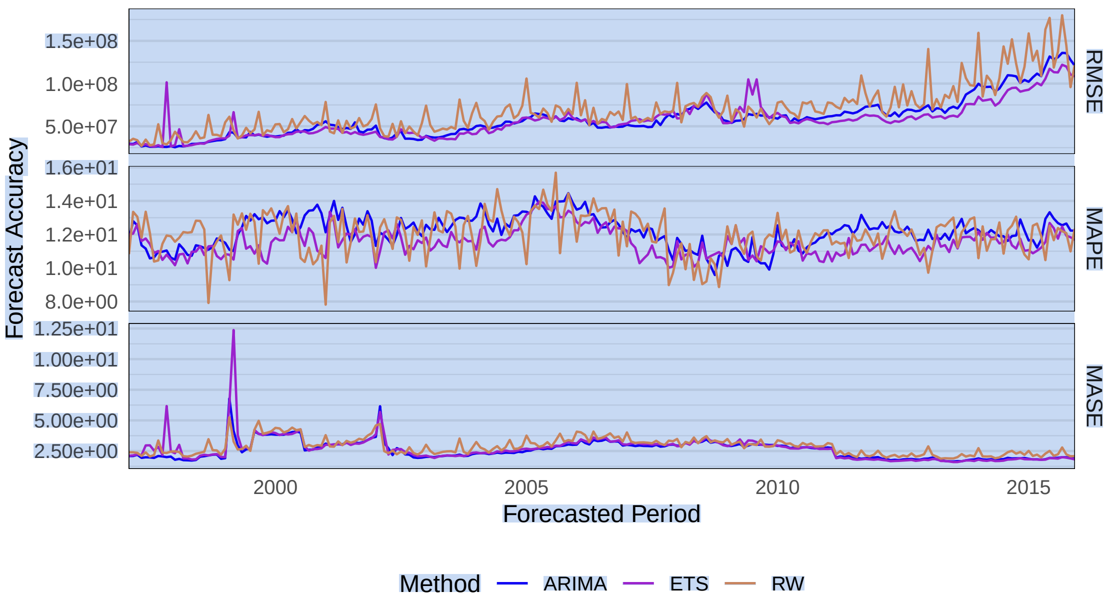

Figure 14: Accuracy of Forecasting Methods at the Bottom Level. Figure provides several measures of forecast accuracy for the bottom-level series from 1995 to 2015. Average of all forecast horizons. Prediction errors are weighted using their corresponding export shares.

A.3 Data

Monthly data on Swiss merchandise trade is compiled by the Swiss Federal Customs Administration.[3] Trade in precious metals, precious and semi-precious stones, works of art and antiques is generally omitted in business cycle research due to volatility and structural breaks. Figure 15 shows the changes in hierarchical composition of Swiss merchandise exports between 1988 and 2018.

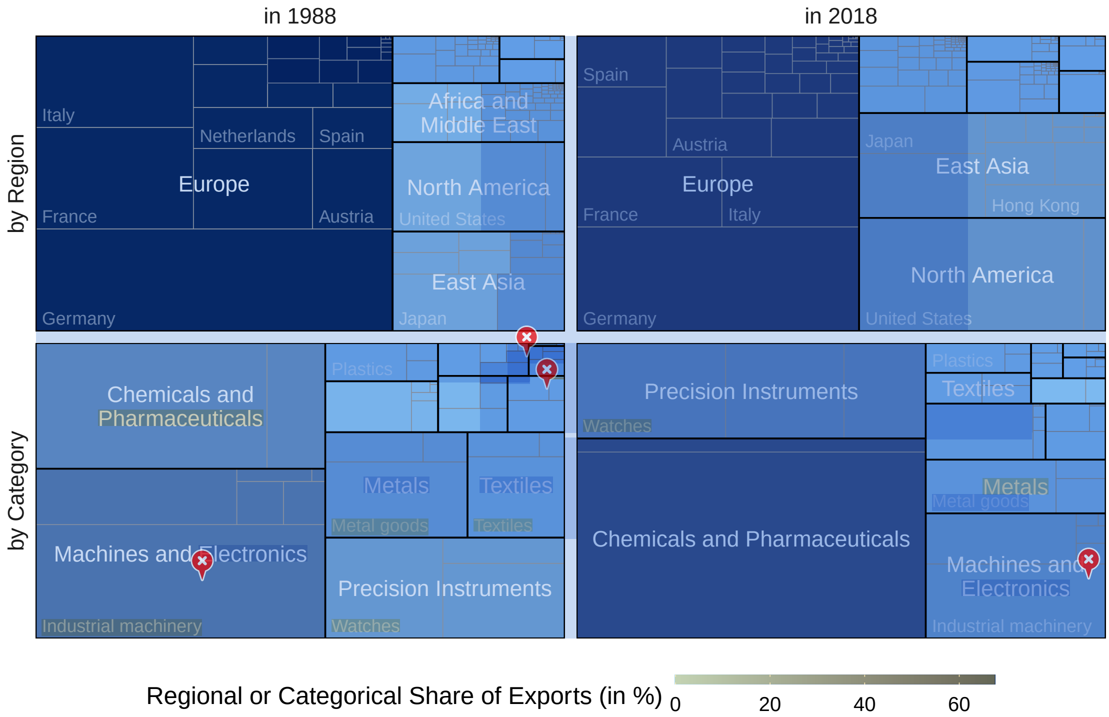

Figure 15: Composition of Swiss Goods Exports. Treemap shows hierarchical composition of Swiss exports by category and destination in 1988 and 2018.

Tables 4 and 5 provide a detailed overview of the regional and categorical dimensions of the dataset on Swiss merchandise trade.

Table 4: Description of Categorical Hierarchy

No.	Description
01	Forestry and agricultural products, fisheries
01.1	Food, beverages and tobacco
01.2	Feeding stuffs for animals
01.3	Live animals
01.4	Horticultural products
01.5	Forestry products (not firewood)
01.6	Products for commercial/industrial further processing such as oils, fats, starches, plants and vegetable parts, etc.
02	Energy source
02.1	Solid combustibles
02.2	Petroleum and distillates

continued ...

[3] https://www.ezv.admin.ch/ezv/en/home/topics/swiss-foreign-trade-statistics.html

...continued

02.3	Gas
02.4	Electrical energy
03	Textiles, clothing, shoes
03.1	Textiles
03.2	Articles of apparel and clothing
03.3	Shoes, parts and accessories
04	Paper, articles of paper and and products of the printing industry
04.1	Basic materials for paper production, such as cellulose and cellulose fibre and paper and carton waste
04.2	Paper and carton in rolls, strips or sheets
04.3	Goods from paper or carton
04.4	Products of the printing industry
05	Leather, rubber, plastics
05.1	Leather
05.2	Rubber
05.3	Plastics
06	Products of the chemical and pharmaceutical industry
06.1	Chemical raw materials, basic materials and unformed plastics
06.2	Chemical end products, vitamins, diagnostic products, including active substances
07	Stones and earth
07.1	Mineral raw materials and basic products
07.2	Goods from stone and cement
07.3	Ceramic wares
07.4	Glass
08	Metals
08.1	Iron and steel
08.2	Non-ferrous metals
08.3	Metal goods
09	Machines, appliances, electronics
09.1	Industrial machinery
09.2	Agricultural machines
09.3	Household appliances
09.4	Office machines
09.5	Electrical and electronic industry appliances and devices
09.6	Military equipment
10	Vehicles
10.1	Road vehicles
10.2	Railed vehicles
10.3	Air- and spacecraft
10.4	Watercraft
11	Precision instruments, clocks and watches and jewellery
11.1	Precision instruments and equipment
11.2	Watches
11.3	Jewellery and household goods made from precious metals
12	Various goods such as music instruments, home furnishings, toys, sports equipment, etc.
12.1	Exposed film
12.2	Music instruments
12.3	Home furnishings
12.4	Toys and sports equipment
12.5	Stationery goods
12.6	Various goods such as umbrellas, neon signs, festive articles, brushes, lighters, pipes, etc.

continued ...

. . . continued

13	Precious metals, precious and semi-precious stones
13.1	Precious and semi-precious stones
13.2	Precious metals (including gold and silver bars from 1.1.2012)
14	Works of art and antiques
14.1	Works of art
14.2	Antiques and collectors' items

Table 5: Description of Geographical Hierarchy

Country		valid from	valid to
Western Europe			
DE	Germany	01/1988	-
FR	France	01/1988	-
IT	Italy	01/1988	-
NL	Netherlands	01/1988	-
BE	Belgium-Luxembourg	01/1988	12/1998
BE	Belgium	01/1999	-
LU	Luxembourg	01/1999	-
AT	Austria	01/1988	-
GB	United Kingdom	01/1988	-
DK	Denmark	01/1988	-
NO	Norway	01/1988	-
SE	Sweden	01/1988	-
PT	Portugal	01/1988	-
FI	Finland	01/1988	-
HR	Croatia, Republic of	02/1992	-
SI	Slovenia	02/1992	-
BA	Bosnia and Herzegovina	05/1992	-
MK	North Macedonia	05/1992	-
ME	Montenegro	05/1992	12/1996
ME	Montenegro	01/2007	-
XM	Montenegro	01/2006	12/2006
SQ	Serbia	05/1992	12/1996
RS	Serbia	01/2007	-
XS	Serbia	01/2006	12/2006
YU	Federal Republic of Yugoslavia	01/1997	12/2003
CS	Serbia and Montenegro	01/2004	12/2005
XK	Kosovo	01/2006	-
IS	Iceland	01/1988	-
IE	Ireland	01/1988	-
ES	Spain	01/1988	-
GR	Greece	01/1988	-
TR	Turkey	01/1988	-
DD	GDR	01/1988	10/1990
PL	Poland	01/1988	-
CZ	Czech Republic	01/1993	-
CS	Czechoslovakia	01/1988	02/1992
SK	Slovakia	01/1993	-
HU	Hungary	01/1988	-
AL	Albania	01/1988	-
BG	Bulgaria, Republic of	01/1988	-
RO	Romania	01/1988	-
SU	USSR	01/1988	12/1991
YU	Yugoslavia	01/1988	04/1992
CY	Cyprus	01/1988	-

continued . . .

. . . continued

SJ	Svalbard and Jan Mayen Island	01/1999	-
MT	Malta	01/1988	-
GI	Gibraltar	01/1988	-
FO	Faeroe Islands	01/1988	-
SM	San Marino	01/1999	-
VA	Holy See	01/1999	-
AD	Andorra	01/1988	-
EE	Estonia	01/1992	-
LV	Latvia	01/1992	-
LT	Lithuania	01/1992	-
QX	Countries not specified	01/2002	-
Eastern Europe and Central Asia			
RU	Russian Federation	01/1992	-
AM	Armenia	01/1992	-
AZ	Azerbaijan	01/1992	-
BY	Belarus	01/1992	-
GE	Georgia	01/1992	-
KZ	Kazakhstan	01/1992	-
KG	Kyrgyz, Republic	01/1992	-
MD	Moldova, Republic of	01/1992	-
TJ	Tajikistan	01/1992	-
TM	Turkmenistan	01/1992	-
UA	Ukraine	01/1992	-
UZ	Uzbekistan	01/1992	-
Africa and Middle East			
EG	Egypt	01/1988	-
SD	Sudan	01/1988	-
SS	South Sudan, Republic of	09/2011	-
LY	Libya	01/1988	-
TN	Tunisia	01/1988	-
DZ	Algeria	01/1988	-
XA	Canary Islands	01/1988	-
MA	Morocco	01/1988	-
EH	Western Sahara	01/1999	-
XB	Ceuta and Melilla	01/1988	12/2010
GQ	Equatorial Guinea	01/1988	-
XC	Ceuta	01/2001	-
XL	Melilla	01/2001	-
TG	Togo	01/1988	-
SN	Senegal	01/1988	-
ML	Mali	01/1988	-
MR	Mauritania	01/1988	-
CI	Côte d'Ivoire	01/1988	-
BF	Burkina Faso	01/1988	-
BJ	Benin	01/1988	-
NE	Niger	01/1988	-
GN	Guinea	01/1988	-
GM	Gambia	01/1988	-
SL	Sierra Leone	01/1988	-
LR	Liberia	01/1988	-
GH	Ghana	01/1988	-
NG	Nigeria, Federal Republic of	01/1988	-
CM	Cameroon	01/1988	-
GA	Gabon	01/1988	-
CG	Congo, Republic of the	01/1988	-

continued . . .

. . . continued

CF	Central African Republic	01/1988	-
TD	Chad	01/1988	-
CD	Congo, Democratic Republic of the	06/1997	-
ZR	Zaire	01/1988	05/1997
AO	Angola	01/1988	-
GW	Guinea-Bissau	01/1988	-
BW	Botswana	01/1988	-
CV	Cabo Verde, Republic of	01/1988	-
LS	Lesotho	01/1988	-
ST	Sao Tomé and Principe	01/1988	-
NA	Namibia	01/1988	-
ZA	South Africa	01/1988	-
SZ	Swaziland	01/1988	-
ZM	Zambia	01/1988	-
ZW	Zimbabwe	01/1988	-
MW	Malawi	01/1988	-
MZ	Mozambique	01/1988	-
MG	Madagascar, Republic of	01/1988	-
RE	Rèunion	01/1988	-
SH	St Helena, Ascen. and Tristan da Cunha	01/1988	-
KM	Comoros, Union of	01/1988	-
AQ	Antarctica	01/1988	-
MU	Mauritius	01/1988	-
IO	British Indian Ocean Territory	01/1988	-
TZ	Tanzania, United Republic of	01/1988	-
SC	Seychelles, Republic of	01/1988	-
RW	Rwanda	01/1988	-
BV	Bouvet Island	01/1999	-
BI	Burundi	01/1988	-
YT	Mayotte	01/1999	-
SO	Somalia, Federal Republic of	01/1988	-
TF	French Southern Territories	01/1999	-
DJ	Djibouti	01/1988	-
ER	Eritrea	01/1994	-
ET	Ethiopia, Fed. Democratic Republic of	01/1988	-
KE	Kenya	01/1988	-
UG	Uganda	01/1988	-
SY	Syrian Arab Republic	01/1988	-
LB	Lebanon	01/1988	-
IL	Israel	01/1988	-
PS	Palestine, the State of	01/1997	-
JO	Jordan	01/1988	-
SA	Saudi Arabia	01/1988	-
YE	Yemen (Nord)	01/1988	12/1990
YE	Yemen	01/1991	-
YD	Yemen (Sud)	01/1988	12/1990
QA	Qatar	01/1988	-
BH	Bahrain	01/1988	-
AE	United Arab Emirates	01/1988	-
OM	Oman	01/1988	-
KW	Kuwait	01/1988	-
IQ	Iraq	01/1988	-
IR	Iran, Islamic Republic of	01/1988	-
South Asia			
AF	Afghanistan	01/1988	-
PK	Pakistan	01/1988	-

continued . . .

. . . continued

BD	Bangladesh	01/1988	-
IN	India	01/1988	-
LK	Sri Lanka	01/1988	-
MV	Maldives	01/1988	-
NP	Nepal, Federal Democratic Rep.	01/1988	-
BT	Bhutan	01/1988	-
East Asia			
MM	Myanmar, Union of	01/1988	-
TH	Thailand	01/1988	-
MY	Malaysia	01/1988	-
BN	Brunei Darussalam	01/1988	-
SG	Singapore	01/1988	-
KH	Cambodia	01/1988	-
LA	Lao, People's Democratic Republic	01/1988	-
VN	Viet Nam, Socialist Republic of	01/1988	-
MN	Mongolia	01/1988	-
CN	China, People's Republic of	01/1988	-
HK	Hong Kong	01/1988	-
TW	Taiwan	01/1988	-
MO	Macau	01/1988	-
KP	Korea, People's Democratic Republic of	01/1988	-
KR	Korea, Republic of	01/1988	-
JP	Japan	01/1988	-
PH	Philippines	01/1988	-
ID	Indonesia	01/1988	-
TL	East Timor	01/2004	-
TP	East Timor	01/1999	12/2003
North America			
CA	Canada	01/1988	-
PM	St Pierre and Miquelon	01/1988	-
US	United States	01/1988	-
GL	Greenland	01/1988	-
Latin America and the Caribbean			
MX	Mexico	01/1988	-
BZ	Belize	01/1988	-
GT	Guatemala	01/1988	-
HN	Honduras	01/1988	-
SV	El Salvador	01/1988	-
NI	Nicaragua	01/1988	-
CR	Costa Rica	01/1988	-
PA	Panama	01/1988	-
KY	Cayman Islands	01/1988	-
TC	Turks and Caicos Islands	01/1988	-
BS	Bahamas	01/1988	-
BM	Bermuda	01/1988	-
JM	Jamaica	01/1988	-
CU	Cuba	01/1988	-
HT	Haiti	01/1988	-
DO	Dominican Republic	01/1988	-
VI	American Virgin Islands	01/1988	-
PR	Puerto Rico	01/1988	12/2005
DM	Dominica	01/1988	-
VC	St Vincent and the Grenadines	01/1988	-
LC	St Lucia	01/1988	-
MS	Montserrat	01/1988	-

continued . . .

. . . continued

AG	Antigua and Barbuda	01/1988	-
BB	Barbados	01/1988	-
GD	Grenada	01/1988	-
KN	St Kitts and Nevis	01/1988	-
AI	Anguilla	01/1988	-
GP	Guadeloupe	01/1988	-
VG	British Virgin Islands	01/1999	-
MQ	Martinique	01/1988	-
TT	Trinidad and Tobago	01/1988	-
BL	Saint BarthÈlemy	01/2013	-
AN	Netherlands Antilles	01/1988	12/2012
AW	Aruba	01/1999	-
BQ	Bonaire, Sint Eustatius and Saba	01/2013	-
CW	Curacao	01/2013	-
SX	Sint Maarten (NL)	01/2013	-
CO	Colombia	01/1988	-
VE	Venezuela, the Bolivarian Republic of	01/1988	-
GY	Guyana	01/1988	-
SR	Suriname	01/1988	-
GF	French Guiana	01/1988	-
BR	Brazil	01/1988	-
PY	Paraguay	01/1988	-
UY	Uruguay	01/1988	-
AR	Argentina	01/1988	-
FK	Falkland Islands	01/1988	-
GS	South Georgia and South Sandwich Islands	01/1999	-
CL	Chile	01/1988	-
BO	Bolivia, the Plurinational State of	01/1988	-
PE	Peru	01/1988	-
EC	Ecuador	01/1988	-
Australia and Oceania			
AU	Australia	01/1988	-
PG	Papua New Guinea	01/1988	-
CC	Cocos (Keeling) Islands	01/1999	-
HM	Heard and McDonald Islands	01/1999	-
NF	Norfolk Island	01/1999	-
CX	Christmas Island	01/1999	-
NZ	New Zealand	01/1988	-
CK	Cook Islands	01/1998	-
WS	Samoa	01/1988	-
NU	Niue Island	01/1999	-
KI	Kiribati, the Republic of	01/1988	-
TK	Tokelau Islands	01/1999	-
TV	Tuvalu	01/1988	-
PN	Pitcairn Islands	01/1988	-
SB	Solomon Islands	01/1988	-
PF	French Polynesia	01/1988	-
NC	New Caledonia	01/1999	-
WF	Wallis and Futuna	01/1999	-
PU	American Oceania	01/1988	12/1996
UM	American Oceania	01/2006	-
UM	American Oceania	01/1997	12/2005
MP	Northern Mariana, Islands	01/1997	-
MH	Marshall Islands	01/1997	-
FM	Micronesia, Federated States of	01/1997	-
PW	Palau	01/1997	-

continued . . .

. . . continued

FJ	Fiji, Republic of	01/1988	-
AS	American Samoa	01/2006	-
GU	Guam	01/2006	-
VU	Vanuatu	01/1988	-
NR	Nauru	01/1988	-
TO	Tonga	01/1988	-