RUHR
ECONOMIC PAPERS

Does It Add Up? Reconciling Forecasts and Impulse Responses for Hierarchical Macroeconomic Data

Boris Blagov and Clara Krause

RUHR UNIVERSITÄT BOCHUM RUB rwi

#1224

Imprint

Ruhr Economic Papers

Ruhr Economic Paper #1224 “Does It Add Up? Reconciling Forecasts and Impulse Responses for Hierarchical Macroeconomic Data”

Responsible Editor: Almut Balleer

Jointly published by

RWI – Leibniz-Institut für Wirtschaftsforschung e.V.
Hohenzollernstr. 1-3, 45128 Essen, Germany

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Bergische Universität Wuppertal, Schumpeter School of Business and Economics
Gaußstraße 20, 42119 Wuppertal, Germany

Series Coordination:
RWI Press Office, presse@rwi-essen.de, phone +49 (0) 201 8149-213
RWI – Leibniz Institute for Economic Research, Hohenzollernstr. 1–3, 45128 Essen, Germany
www.rwi-essen.de

RWI is funded by the Federal Government and the federal state of North Rhine-Westphalia.
The Institute has the legal form of a registered association; Vereinsregister, Amtsgericht Essen VR 1784

ISSN 1864-4872 (online)
ISBN 978-3-96973-409-4
DOI https://dx.doi.org/10.4419/96973409

Does it add up? Reconciling forecasts and impulse responses for hierarchical macroeconomic data*

Boris Blagov[1] and Clara Krause[1,2]

[1]RWI – Leibniz Institute for Economic Research
[2]Ruhr University Bochum

August 2026

Abstract

This paper introduces a mixed-frequency Gaussian state-space framework that embeds forecast reconciliation into Bayesian VAR modeling for hierarchical macroeconomic data. Using precision-based sampling to generate high-frequency latent estimates, we construct a consistent proxy for the forecast-error covariance matrix, enabling optimal reconciliation with short datasets. We prove the asymptotic convergence of this estimator and show that forecast reconciliation can be formally derived as a special case of conditional forecasting, allowing straightforward implementation with standard state-space algorithms. We derive reconciled impulse response functions (rIRFs) that ensure bottom-level structural responses aggregate exactly to the top-level impulse response. Applying the framework to UK and German regional economic data, we demonstrate improvements in the forecast accuracy alongside structurally consistent impulse responses.

Keywords: forecast reconciliation, regional forecasts, mixed-frequency, hierarchical impulse responses
JEL classification: C32, C53, R11

*This paper circulated previously as "Forecasting Regional Economic Activity". We would like to thank participants at the 15th RCEA Bayesian Econometrics Workshop, the International Association of Applied Econometrics 2025, and the 3^{rd} Vienna Economic Forecasting Workshop, Barbara Rossi, James Mitchell, Almut Balleer, Marvin Nöller, Torsten Schmidt, and Niklas Benner for helpful comments and suggestions, as well as Katja Heinisch for help with the dataset.

1 Introduction

In many applied settings, data exhibits a hierarchical structure in which certain series are exact linear combinations of others. A typical example of a hierarchical system is firm revenue, where the total reported by headquarters must match the sales from all individual entities. In macroeconomics, aggregate GDP and its components constitute one notable example. Likewise, regional economic accounts necessarily need to be in line with national developments. This property complicates forecasting as standard methods without additional restrictions cannot ensure internal coherence of the forecasts. Failure to capture the hierarchy also extends to policy applications such as conditional forecasting, and, more importantly, to structural analysis. Calculating impulse response functions (IRFs) with hierarchical macroeconomic data generate IRFs that suffer from an aggregation bias, as there is no mechanism to ensure that the bottom-level responses add up to the IRF of the upper-level aggregate or vice versa.

In this paper, we propose a solution that accommodates these inherent challenges. More specifically, we introduce forecast reconciliation in the mixed-frequency Gaussian linear state-space setting, making forecast reconciliation possible by deriving an estimator for the forecast error covariance, a key quantity needed for reconciliation. Crucially, our approach thereby overcomes the limitation that reconciliation methods are unsuited for wide data (short T, large N). Furthermore, we show how forecast reconciliation can be framed as a special case of a conditional forecast. This enables the adoption of off-the-shelf conditional forecasting tools to implement it (e.g. Antolín-Díaz et al., 2021; Chan et al., 2025) and allows reconciliation and conditional forecasting to be used simultaneously. Based on this framework, we show how the method can be used for conducting structural analysis with macroeconomic hierarchical data. We propose the notion of reconciled impulse responses (rIRF), which adhere to the hierarchical structure in the sense that aggregated lower-level impulse responses match the impulse response of the corresponding aggregate. To derive the rIRFs we rely on a property of impulse responses in linear VAR models that IRFs may be computed as the difference between a conditional forecast with a specific shock versus a forecast in the absence of any disturbances (Koop et al., 1996; Pesaran and Shin, 1998). Reconciling the forecast before calculating the difference yields the rIRFs, which ensures that the inherent hierarchical structure is maintained.

We illustrate the usefulness of our approach in forecasting exercises for the United Kingdom and Germany. In an application with regional and national UK data we show that reconciling the forecasts does not only make them coherent and useful for policy discussions but it also improves both the point and density forecasts, where the gains for the density are substantial. In an exercise for the German states, we show that state specific VARs with additional regional indicators can improve upon the forecast performance of a large VAR but simultaneously heighten the need for reconciliation. We also highlight the IRF aggregation bias in a proxy MF-VAR setting for the UK regions with an economic activity shock and show how reconciled impulse responses can resolve this issue by ensuring that the hierarchy is respected.

Our methodology builds upon the established framework of forecast reconciliation (Athanasopoulos et al., 2024), a model-free approach based on subspace projections which transforms a set of *base* forecasts into a group of *reconciled* forecasts that respect the hierarchical structure. Under very general conditions, reconciled forecasts are always at least as good and often better in terms of forecast accuracy compared to the base counterpart (Wickramasuriya et al., 2019). However, forecast reconciliation requires sufficient out-of-sample observations and works best in datasets with large T and small N, often referred to as tall data. In contrast to many business applications, macroeconomic hierarchical data is often available only for relatively short horizons and in low frequency, i.e. as wide data. These features make forecast reconciliation unsuitable in a macroeconomic context with short datasets.[1]

Recent methodological advancements have largely addressed the problem of short datasets in the context of macroeconomic hierarchical data. For example, a large body of the literature has dealt with regional economic accounts, specifically gross domestic product (GDP) and gross value added (GVA), which have a hierarchical structure when modeled together with national data. Most countries produce these account data for the regions annually, while they have quarterly data for the country as a whole. The most common approach is to frame this problem as an unobserved-variable problem - regional data is only observed at the low frequency, but is in fact generated at a higher frequency, e.g. quarterly. The literature then imposes a cross-sectional constraint that enforces the hierarchical structure within the sample. This formulation allows the use of latent-variable approaches such as the Kalman filter, simulation smoother, or precision-based sampler to generate unobserved high-frequency data while preserving the hierarchy. These techniques yield high-frequency estimates that are the best approximation of the true data (optimal in a MSE sense). Additionally, this solves the short T problem, by quadrupling the observations and improving parameter inference. Recent examples include Mixed-Frequency Bayesian Vector Autoregression models (MF-VARs) (Koop et al., 2020b, 2024b,a,c; Lehmann, 2024) or Mixed-frequency Dynamic Factor Models (Barbaglia et al., 2024). Yet, these state-of-the-art mixed-frequency models incorporate the hierarchical structure only in-sample. For out-of-sample applications such

[1] Nevertheless, it has been used in macroeconomic context with datasets that have sufficiently long observations, often in higher frequency. The classical example is monthly regional Australian tourism (Athanasopoulos et al., 2009; Hyndman et al., 2011; Wickramasuriya et al., 2019; Athanasopoulos et al., 2024). Eckert et al. (2021) use it to forecast monthly Swiss exports. It has also been employed to reconcile the quarterly GDP from the income and expenditure side of GDP (Athanasopoulos et al., 2020; Bisaglia et al., 2020; Di Fonzo and Girolimetto, 2023).

as forecasting, structural analysis, or scenario analysis, the cross-sectional constraint has not been utilized. This is partly because the restriction depends on the availability of the corresponding high-frequency national observation, which itself is released with delay. Consequently, the failure to capture the hierarchy out-of-sample creates a persistent gap for forward-looking policy evaluations and structural inference. Our method bridges this gap by integrating reconciliation into the state-space estimation, thereby enforcing cross-coherence even beyond the estimation period.

The remainder of the paper is organized as follows. In the methodology, Section 2, we begin by briefly sketching the basic MF-VAR framework for missing data and forecasting. We then introduce the concept of forecast reconciliation and show how it can be incorporated in our framework with wide data and tied to the mixed-frequency aspect. In the final subsection of the methodology, we demonstrate how forecast reconciliation can be cast as a conditional forecast. This setup then enables us to introduce the notion of reconciled IRFs. We illustrate this framework empirically in Section 3. First, Sections 3.1 and 3.2 focus on a forecasting exercise of UK and German regional data to assess the forecast accuracy of the reconciliation method versus the regular forecasts. We then consider the issue of aggregation bias in impulse response functions utilizing the UK regions in Section 3.3 and Section 4 concludes.

2 Mixed-frequency VAR with a precision-based Sampler

To address the limitations of standard reconciliation methods in short datasets, we embed our approach within a mixed-frequency Gaussian linear state-space framework. This setup allows us to overcome the limitations of wide hierachical data while maintaining internal consistency. First, we sketch the mixed-frequency framework that we later combine with forecast reconciliation. Our chosen latent-variable model is precision-based sampling (Chan et al., 2023; Hauber and Schumacher, 2021; Mertens, 2023). The existing literature on hierarchical macroeconomic data with a focus on regional economic activity predominantly uses a Kalman filter-based approach (e.g. Koop et al., 2020b). Our results do not rely on this choice, as our proof of convergence needed for the implementation of forecast reconciliation is analogous in both methods and therefore it is ultimately a matter of preference. We derive forecast reconciliation as a form of conditional forecast in the framework of Chan et al. (2025) and therefore we find the use of the precision-based sampler more convenient. Our model of interest is represented by the following VAR specification

$$\mathbf{y}_t = \mathbf{b}_0 + \mathbf{B}_1\mathbf{y}_{t-1} + \cdots + \mathbf{B}_1\mathbf{y}_{t-p} + \varepsilon_t \qquad \varepsilon_t \sim \mathcal{N}(\mathbf{0}_n, \sigma) \tag{1}$$

with n variables of the same frequency. Here, $\mathbf{y}_t$ has size $n \times 1$ and $\mathbf{0}_n$ denotes an $n \times 1$ vector of zeros and p is the number of lags. σ is the time-invariant covariance matrix.[2] Stacking over $t = 1, \ldots, T$ we can rewrite the VAR in matrix notation

$$\mathbf{H_B y} = \mathbf{c_B} + \varepsilon \qquad \varepsilon \sim \mathcal{N}(\mathbf{0}_{Tn}, \mathbf{\Sigma}) \tag{2}$$

where $\mathbf{y}$, $\mathbf{c_B}$ and ε each are vectors of size $Tn \times 1$ and $\mathbf{H_B}$ has size $Tn \times Tn$. The matrix $\mathbf{H_B}$ and the vector $\mathbf{c_B}$ have the following structure

$$\mathbf{c_B} = \begin{bmatrix} \mathbf{b}_0 + \sum_{j=1}^{p} \mathbf{B}_j \mathbf{y}_{1-j} \\ \mathbf{b}_0 + \sum_{j=2}^{p} \mathbf{B}_j \mathbf{y}_{2-j} \\ \vdots \\ \mathbf{b}_0 + \mathbf{B}_p \mathbf{y}_0 \\ \mathbf{b}_0 \\ \vdots \\ \mathbf{b}_0 \end{bmatrix} \mathbf{H_B} = \begin{bmatrix} \mathbf{I}_n & \mathbf{0}_{n\times n} & \cdots & \cdots & \cdots & \cdots & \cdots & \mathbf{0}_{n\times n} \\ -\mathbf{B}_1 & \mathbf{I}_n & \mathbf{0}_{n\times n} & \cdots & \cdots & \cdots & \cdots & \mathbf{0}_{n\times n} \\ -\mathbf{B}_2 & -\mathbf{B}_1 & \mathbf{I}_n & \mathbf{0}_{n\times n} & \cdots & & & \mathbf{0}_{n\times n} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & & \vdots \\ -\mathbf{B}_p & \cdots & & -\mathbf{B}_1 & \mathbf{I}_n & \mathbf{0}_{n\times n} & & \vdots \\ \mathbf{0}_{n\times n} & & & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0}_{n\times n} & \cdots & \mathbf{0}_{n\times n} & -\mathbf{B}_p & \cdots & -\mathbf{B}_2 & -\mathbf{B}_1 & \mathbf{I}_n \end{bmatrix}$$

Suppose that some variables are fully or partially missing. Chan et al. (2023) show that by partitioning the data in observed and missing values we can derive the distribution of the missing data. Let $\mathbf{S^o}$ and $\mathbf{S^m}$ be selection matrices indicating which elements are observed and which are missing, respectively. We then rewrite eq. (2) as

$$\mathbf{H_B S^o y^o} + \mathbf{H_B S^m y^m} = \mathbf{c_B} + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(\mathbf{0}_{Tn}, \mathbf{\Sigma}), \tag{3}$$

where $\mathbf{y^o}$ has size $n^o \times 1$ and $\mathbf{y^m}$ has size $n^m \times 1$ with n^o and n^m giving the total number of observed and missing values respectively. Then, the conditional distribution of the missing data given the model, observed data, and all parameters θ is

$$(\mathbf{y^m} | \mathbf{y^o}, \theta) \sim \mathcal{N}(\mu_{\mathbf{y^m}}, \mathbf{K}_{\mathbf{y^m}}^{-1}) \tag{4}$$

$$\mathbf{K}_{\mathbf{y^m}} = (\mathbf{H_B S^m})' \mathbf{\Sigma}^{-1} (\mathbf{H_B S^m}) \qquad \mu_{\mathbf{y^m}} = \mathbf{K}_{\mathbf{y^m}}^{-1} (\mathbf{H_B S^m})' \mathbf{\Sigma}^{-1} (\mathbf{c_B} - \mathbf{H_B S^o y^o})$$

This efficient sampling approach allows for handling high-dimensional settings with complex patterns of missing data such as mixed-frequency, time series of different lengths and ragged edge.

Drawing from the above normal distribution allows us to construct a balanced data set by generating estimates for the missing observations. In the context of mixed-

[2]Note that homoskedasticity is not a requirement for using forecast reconciliation, in fact the method is completely model free. We choose this version as stochastic volatility is not central to the application.

frequency and hierarchical data, however, these imputations cannot be treated as unconstrained independent draws. Instead, they must respect the intertemporal and cross-sectional structural dependencies inherent to the data. Specifically, the temporal relationship between low and high-frequency observations as well as the hierarchical structure must be sustained. This additional information can be incorporated in the model.

First, the inter-temporal constraint ensures that the constructed high-frequency series remain consistent with the available low-frequency observations. We employ the widely used approximation proposed by Mariano and Murasawa (2003) which establishes a weighted relationship between the low-frequency log-growth rate and the underlying high-frequency growth rates. In the following applications, we focus primarily on a quarterly–annual frequency configuration although the framework easily accommodates any mixed-frequency setup. Therefore, we rely on the following inter-temporal constraint linking annual and quarterly log-growth rates.

$$y^a_{i,\tau} = \frac{1}{4}y^q_{i,t} + \frac{2}{4}y^q_{i,t-1} + \frac{3}{4}y^q_{i,t-2} + y^q_{i,t-3} + \frac{3}{4}y^q_{i,t-4} + \frac{2}{4}y^q_{i,t-4} + \frac{1}{4}y^q_{i,t-5} + o^{int}_{i,t} \tag{5}$$

The equation approximates the annual growth rate of variable i in year τ as a weighted sum of the growth rate of the four quarters in τ and the last three from $\tau - 1$. We impose this as a soft constraint, given its approximative nature, which is captured by the term $o^{int}_{i,t}$.

Second, the cross-sectional constraint enforces that the high-frequency estimates adhere to the hierarchical system. This is achieved by modeling the upper-level variable as an approximate aggregation of the lower-level components. Any form of single or multiple layer hierarchy can be modeled using such constraints. The specification of the aggregation scheme typically depends on the problem at hand and the definition of the variable. For instance, one might utilize a simple arithmetic sum for absolute levels or a weighted average for growth rates. A representative cross-sectional constraint takes the following form.

$$y^q_t = \sum_{i=1}^{nR} \omega_{t,i} y^q_{i,t} + o^{cs}_t \tag{6}$$

The above equation implements the upper level as a weighted average of the lower-level data. The weights, $\omega_{t,i}$, can be chosen as constant, assigning equal weight across all lower-level series such that $\omega_{t,i} = \frac{1}{nR} \quad \forall t, i$ (Koop et al., 2020a; Lehmann and Wikman, 2025). Alternatively, the framework can accommodate more complex structures where the aggregation reflects dynamic economic shifts. This is achieved by employing a cross-sectional constraint using stochastic hierarchical aggregation, as proposed by Koop et al. (2024a). Here, the weights $\omega_{t,i}$ are modeled as component-specific and time-varying parameters rather than fixed constants.

To incorporate the constraints in the model we first collect observed data relevant for the constraints in a large vector $\mathbf{z}$. Then, we map $\mathbf{z}$ to the respective missing observations with the help of the matrix $\mathbf{M}$ which is populated with both the inter-temporal and the cross-sectional constraints

$$\mathbf{z} = \mathbf{M}\mathbf{y}^{\mathbf{m}} + \varepsilon^{z}, \varepsilon^{z} \sim \mathcal{N}(\mathbf{0}_M, \mathbf{O}). \tag{7}$$

The error-terms are collected in the vector ε^{z}, reflecting the soft-constraint. As Chan et al. (2023) show the restricted missing values can be drawn from the following conditional normal distribution.

$$(\mathbf{y}^{\mathbf{m}}|\mathbf{y}^{\mathbf{o}}, \theta, \mathbf{z}) \sim \mathcal{N}(\hat{\mu}_{\mathbf{y}^{\mathbf{m}}}, \bar{\mathbf{K}}_{\mathbf{y}^{\mathbf{m}}}^{-1}) \tag{8}$$

$$\bar{\mathbf{K}}_{\mathbf{y}^{\mathbf{m}}} = \mathbf{M}'\mathbf{O}^{-1}\mathbf{M} + \mathbf{K}_{\mathbf{y}^{\mathbf{m}}} \qquad \hat{\mu}_{\mathbf{y}^{\mathbf{m}}} = \bar{\mathbf{K}}_{\mathbf{y}^{\mathbf{m}}}^{-1}(\mathbf{M}'\mathbf{O}^{-1}\mathbf{z} + \mathbf{K}_{\mathbf{y}^{\mathbf{m}}}\mu_{\mathbf{y}^{\mathbf{m}}})$$

2.1 Forecasting

Forecasting with this framework in both conditional and unconditional settings is straightforward. One can simply treat future observations as missing and thus draw them directly from (8) along with all missing variables in one step (e.g. the unobserved values in the mixed-frequency setting).

Performing unconditional forecasting means extending the dataset by projecting 'missing' values for h forecast periods. Conditional forecasting, on the other hand, conditions on a specific path of observations. We distinguish two main contexts in which this approach applies. First, we address the ragged edge problem where varying publication lags imply that some data points are available while others remain missing. And second, we consider the possibility of postulating specific future trajectories for individual variables, often within a scenario framework. For both cases missing observations can be drawn from eq. (8) given the (hypothetical) realizations of the other indicators and the full history.

There is, however, one drawback when dealing with a cross-sectional constraint such as needed with hierarchical data. The restriction can only be imposed if $\mathbf{z}$ contains the necessary information, i.e. the respective upper-level aggregate is observed. Consequently, coherent forecasts cannot be generated for either unconditional or conditional settings unless the path of the upper-level data is known a priori. Thus, coherent forecasts cannot be generated as this crucial information is missing and is only achievable once the upper-level series has been published. As this is associated with a publication lag in most cases, also coherent nowcasts are typically infeasible.

2.2 Forecast reconciliation

Suppose we have a hierarchical structure within our data, i.e. one that adheres to known linear constraints. A forecast can then only be suitable if it is compliant with the underlying hierarchy. One typical example in the literature are sales forecasts from different branches within a company, where the sum over all branches should result in the overall company sales. Or in our later applications, the national GDP growth needs to be consistent with the weighted sum over all regions.

Given the $n \times 1$ vector $\mathbf{y}_t$ from eq. (1), suppose that a subset $\mathbf{y}_t^R$ is an $r \times 1$ collection of hierarchical time series, such that

$$\mathbf{y}_t^R = \mathbf{R}\mathbf{y}_t \tag{9}$$

where matrix $\mathbf{R}$ selects the relevant rows from $\mathbf{y}_t$, i.e. it has full row rank. The time series can have multiple layers of hierarchy (e.g. national, regional, and sub-regional level). Without loss of generality, assume only two levels of hierarchy, a bottom level with n^b series and an upper level with the aggregate. The linear relationship may be written $\mathbf{u}_t = \mathbf{A}\mathbf{b}_t$, such that $\mathbf{y}_t^R = [\mathbf{u}_t, \mathbf{b}_t]'$ and $\mathbf{A}$ describes how the bottom series relate to the upper level.

Consider a set of forecasts that do *not* fulfill the hierarchical structure, referred to as *base* forecasts and denoted by $\hat{\mathbf{y}}_{T+h}$. They are produced with information up to time period T for forecast horizon h. As shown by Hyndman et al. (2011), a set of *reconciled* forecasts, $\tilde{\mathbf{y}}_{T+h}^R$, which fulfills the constraints implied by the hierarchical structure, can be written as a linear combination of the *base* forecasts and two matrices $\mathbf{S}$ and $\mathbf{P}$

$$\tilde{\mathbf{y}}_{T+h}^R = \mathbf{S}\mathbf{P}\hat{\mathbf{y}}_{T+h}^R \tag{10}$$

where $\mathbf{S} = (\mathbf{A}', \mathbf{I}_{n^b})'$ is an $r \times n^b$ summing matrix that contains the arithmetic relationship between the bottom and top levels. $\mathbf{P}$ is an $n^b \times r$ appropriately selected weighting matrix that maps the base to the reconciled forecasts.

The matrix $\mathbf{P}$ can be specified in various forms by the researchers. For instance, a simple bottom up approach takes the sum of all bottom forecasts and discards the upper level forecast. Or alternatively, one may consider a top down approach distributing the top level forecast and disregarding the bottom ones.

The forecast reconciliation literature has focused extensively on what is an optimal choice of $\mathbf{P}$ (see Athanasopoulos et al. (2024) for a review of the forecast reconciliation literature and Wickramasuriya (2021) for a review of various choices of $\mathbf{P}$). One of the central results in the literature is the trace minimization (MinT) approach of Wickramasuriya et al. (2019). There, $\mathbf{P}$ is formulated such that it minimizes the sum of forecast error variances between the *reconciled* forecasts and the *true* data. The solution to this problem is

$$\mathbf{P}_h = (\mathbf{S}'\mathbf{W_h^{-1}}\mathbf{S})^{-1}\mathbf{S}'\mathbf{W_h^{-1}}, \tag{11}$$

with $\mathbf{W}_h = \text{Var}(\hat{\mathbf{y}}_{t+h}^{R} - \mathbf{y}_{t+h}^{true}|\mathcal{I}_T)$ being the error variance-covariance matrix of the base forecasts for forecast horizon h conditional on the information set up to time T, $\mathcal{I}_T$.

This result, based on an optimization problem, has been the solution to various approaches aimed at finding a good choice for $\mathbf{P}_h$. For example, Van Erven and Cugliari (2015) show that reconciled forecasts using a loss-function approach are always strictly at least as good if not better than the base forecasts, while Panagiotelis et al. (2021) find that eq. (11) can be derived as the solution to the optimal loss function, generalizing the two approaches. Furthermore, it can also be considered as a special case of a probabilistic forecast reconciliation approach, as shown by Wickramasuriya (2024).

With the substitution of eq. (11) into eq. (10) completing a conventional reconciliation framework, we now turn to the general properties of the method. The approach is model-free in the sense that it is derived without any assumptions of how the base forecasts are generated. However, any (undesired) properties of the base forecasts propagate to the reconciled forecasts, e.g. biasedness (Eckert et al., 2021; Athanasopoulos et al., 2024).

The important feature that, in terms of forecast accuracy, reconciled forecasts will always be at least as good as base forecasts (and can be better), relies on the assumption that $\mathbf{W}_h$ is "well" estimated. This can be especially problematic in short T large N systems, since the method requires a sufficient number of out-of sample forecasts to estimate $\mathbf{W}_h$. In general, Wickramasuriya et al. (2019) highlight that the number of forecast errors must be at least equal to r, i.e. the number of hierarchical series. While this figure serves as a theoretical lower bound, increasing the number of out-of sample forecasts – unless driven by extreme events – typically enhances the estimation of $\mathbf{W}_h$ and in some cases has proven necessary to have a well defined matrix. This presents a significant challenge with low-frequency macroeconomic data given the limited number of observations, particularly in the case of large cross-sectional dimensions.

We propose to introduce forecast reconciliation in a mixed-frequency framework at the higher frequency. Namely, not reconciling the official low-frequency data, but reconciling the unobserved high-frequency data. This not only yields a sufficient number of forecasts but also opens the possibility to generate high-frequency reconciled forecasts for the low-frequency variables. This idea draws on the ability of mixed-frequency models to generate an accurate assessment of unobserved variables.

Formally, we propose to estimate $\mathbf{W_h}$ treating the disaggregated high-frequency time series as the true data, namely define

$$\mathbf{W_h^m} = \text{Var}\left(\hat{\mathbf{y}}_{T+h}^{R} - \mathbf{y^m}_{T+h}^{,R}\right) \tag{12}$$

and use $\mathbf{W_h^m}$ in equation (11) instead of $\mathbf{W_h}$. Here, $\mathbf{y}^{\mathbf{m},R}$ denotes the relevant entries from $\mathbf{y^m}$ in eq. (4) that correspond to the hierarchical series only. For this approximation to be valid we assume that $\mathbf{y}^{\mathbf{m},R}$ is a good approximation of the true data $\mathbf{y^{true}}$ which implies the following proposition.

Proposition 1. *The forecast error variance-covariance matrix estimated on the unobserved high-frequency values converges in distribution to the true forecast error variance-covariance matrix*

$$\mathbf{W_h^m} \xrightarrow{d} \mathbf{W_h}.$$

Proof. See Appendix A □

The idea is that the latent variable methods such as the Kalman filter and the precision-based sampler deliver accurate estimates of the unobserved data, hence we can use it in place of the true data.

2.3 Forecast reconciliation and conditional forecasts

As described in Section 2.1, in this framework forecasts are essentially missing values and can be directly incorporated at the estimation step – drawing from eq. (8). Nevertheless, to formalize forecast reconciliation it is easier to consider only the forecast distribution. Consider eq. (2) where we drop the first $T-p$ observations, saving only the last p entries and add additional h rows of missing data. Given a set of coefficient matrices the forecast distribution may be written as

$$\mathbf{y}_{T+1:T+h} = \mathbf{H}^{-1}\mathbf{c} + \mathbf{H}^{-1}\varepsilon_{T+1:T+h} \sim \mathcal{N}(\mathbf{H}^{-1}\mathbf{c}, \mathbf{H}^{-1}\mathbf{\Xi}\mathbf{H}^{-1\prime}), \tag{13}$$

where $\mathbf{H}$ is comprised of the corresponding subsection of $\mathbf{H_B}$, and respectively $\mathbf{c}$ is a subsection of $\mathbf{c_b}$ and $\mathbf{\Xi}$ is a subsection of $\mathbf{\Sigma}$.[3]

In the absence of observations past $t = T$, the forecast is centered at $\mathbf{H}^{-1}\mathbf{c}$. This is essentially an unconditional forecast. If there is a 'ragged edge', i.e. some values past $t = T$ are available, and/or an assumption of a future path to condition on, we follow eqs. (3) through (8) to derive the conditional *base* forecast distribution.

$$(\hat{\mathbf{y}}_{T+1:T+h}|\mathbf{y^o}, \theta, \mathbf{z}) \sim \mathcal{N}(\hat{\mu}_{T+1:T+h}, \bar{\mathbf{K}}^{\mathbf{-1}}), \tag{14}$$

where $\hat{\mu}_{T+1:T+h}$ are the relevant periods of $\hat{\mu}_{\mathbf{y^m}}$ from eq. (8) and $\bar{\mathbf{K}}^{-1}$ is specified by $\mathbf{H}$ and $\mathbf{c}$.

[3]Even though we build on Chan et al. (2025), in this paper we work only in the reduced form of the model, meaning that $\mathbf{H_B}$ in eq. (2) and $\mathbf{H}$ in eq. (13) contain only reduced form coefficients. Any structural identification is formulated as $\varepsilon = \mathbf{B_0}\epsilon$. In the literature, this is often referred to as the *B-model*. Therefore, the formulation of eq. (13) differs from Chan et al. (2025), who cast it in a structural form and have $\mathbf{\Xi} = \mathbf{I}$ while $\mathbf{H}$ contains structural coefficients and is not a subset of $\mathbf{H_B}$ from eq. (2). This difference is illustrated in Appendix C.

Chan et al. (2025) introduce a novel precision-based sampler for generating conditional forecasts, which easily handles large systems and significantly reduces computational time. They propose to introduce conditions on steps (13) and (14) in the form of an additional distribution. The conditions, ρ, are a set of linear restrictions on the future path of observables implemented in the general form $\mathcal{N}(\rho, \mathbf{\Omega})$. If $\mathbf{\Omega}$ is a zero matrix, the path is enforced as a hard constraint. A soft constraint on the other hand, emulates tilting.[4] In the case of hard constraints, the conditions are essentially treated as observed data, $\mathbf{y^o}$, and eq. (14) draws the conditional base forecast.

Forecast reconciliation can then be implemented as another specific set of linear constraints in the above setup. First, consider the base forecasts $\hat{\mathbf{y}}_{T+h}$ for period $T+h$. The reconciled forecast for the hierarchical subset is then $\tilde{\mathbf{y}}^R_{T+h}$:

$$\tilde{\mathbf{y}}^R_{T+h} = \mathbf{S}\mathbf{P_h}\mathbf{R}\hat{\mathbf{y}}_{T+h}. \tag{15}$$

Second, stack the forecasts for $T+1, \ldots, T+h$ to get

$$\tilde{\mathbf{y}}^R_{T+1:T+h} = \mathbf{\Pi}\hat{\mathbf{y}}_{T+1:T+h}, \tag{16}$$

with

$$\mathbf{\Pi} = \begin{pmatrix} \mathbf{S}\mathbf{P_1}\mathbf{R} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{S}\mathbf{P_h}\mathbf{R} \end{pmatrix}.$$

The reconciled forecast is equivalent to a draw from a normal distribution with the mean of the base forecasts augmented with the matrix $\mathbf{\Pi}$ and a variance-covariance containing zeros. In a Bayesian framework utilizing a Gibbs sampler, reconciliation is performed for each draw.

$$\tilde{\mathbf{y}}^R_{T+1:T+h} \sim \mathcal{N}(\mathbf{\Pi}\hat{\mu}_{T+1:T+h}, \mathbf{0}_{r\times r}). \tag{17}$$

The zero variance-covariance matrix implies that we are introducing reconciliation as a hard constraint. This may be easily relaxed if the hierarchical constraints do not have to match up exactly.

Lastly, we need to consider the remaining non-hierarchical series $\tilde{\mathbf{y}}^{(-R)}_{T+1:T+h}$. Incorporating the reconciled draw from (17) as observed data in the framework of (14), leads to the following forecast distribution conditional on the forecast reconciliation

$$(\tilde{\mathbf{y}}^{(-R)}_{T+1:T+h} | \tilde{\mathbf{y}}^R_{T+1:T+h}, \mathbf{y^o}, \theta) \sim \mathcal{N}(\tilde{\mu}_{\mathbf{y}^R}, \bar{\mathbf{K}}^{-1}_{\mathbf{y^m}}) \tag{18}$$

[4]Their framework is even more general and can accommodate also inequality constraints, which are not the purpose here.

with $\tilde{\mu}_{\mathbf{y}^R} = \bar{\mathbf{K}}^{-1}(\mathbf{HS^m})'\Xi^{-1}(\mathbf{c} - \mathbf{\Pi}\hat{\mu}_{y^m,T+1:T+h})$. Note that $\bar{\mathbf{K}}^{-1}$ is unchanged since we impose forecast reconciliation as a hard constraint.

The full set of reconciled forecasts is then

$$\tilde{\mathbf{y}}_{T+1:T+h} = \mathbf{R}^{-}\tilde{\mathbf{y}}^{(-R)}_{T+1:T+h} + \mathbf{R}\tilde{\mathbf{y}}^{R}_{T+1:T+h}$$

where $\mathbf{R}^{-}$ indicates the observations associated with the non-hierarchical time series and is therefore the full row rank counterpart of $\mathbf{R}$.

To summarize, introducing forecast reconciliation in a dataset of size $t = 1, \ldots, T$ can be implemented in the following manner:

1. On a subset with $t = 1, \ldots, T_0$ estimate the model and create base forecasts for $T_0 + 1, \ldots T_0 + h$ where $T_0 \geq r$ (number of hierarchical series). Save the base forecasts.
2. For $t = T_0 + 1$ repeat 1).
3. Use true data, or in case of mixed-frequency $\mathbf{y}^{\mathbf{m},R}$, to calculate all forecast errors and the variance, $\mathbf{W}^{\mathbf{m}}_{\mathbf{h}}$, for use in eq. (11).
4. Perform forecast reconciliation for $T_0 + 2, \ldots, T_0 + 1 + h$.
   (a) Given the base forecast from 2), use eq. (17) to draw the reconciled hierarchical series.
   (b) Treating the hierarchical series as observed, draw a congruent estimate for the non-hierarchical series with the help of eq. (18).
5. Repeat 2) through 4) for $t = T_0 + 2, \ldots, T$.

Finally, we would like to raise several points regarding the derived distribution in (17). First, eq. (17) is general and does not depend on the choice of $\mathbf{P_h}$, allowing for all types of forecast reconciliation to be employed.

Introducing reconciliation as a form of conditional forecast is conceptually different than introducing it at the forecasting step, as for example the approach in Ashouri et al. (2022). This could, in principle, be achieved here by pre-multiplying $\mathbf{SP}$ along the main diagonal of $\mathbf{H}$ in eq. (13). However, as evident from that equation, we need to invert $\mathbf{H}$ and nothing ensures that $\mathbf{SP}$ is invertible and therefore $det(\mathbf{H}) \neq 0$. In fact, in most applications $\mathbf{SP}$ does not have full-rank. Even in the simplest bottom-up hierarchical forecast reconciliation the top-level forecast is a row of zeroes. Inverting $\mathbf{H}$ is possible using methods such as the Moore–Penrose inverse, however this pevents the use of the efficient algorithm of Chan and Jeliazkov (2009) that employs forward and backward substitution via the Cholesky decomposition and thus avoids calculating the large inverse of $\bar{\mathbf{K}}^{-1}$.

Second, the formulation of $\mathbf{\Pi}$ implies that the distribution gives the forecast path conditional on consecutively reconciled one-step ahead forecasts. To see this point consider the equivalent forecast reconciliation in the traditional way by iterating on $\mathbf{y}_t = \mathbf{b} + \mathbf{B}_1\mathbf{y}_{t-1} + \cdots + \mathbf{B}_1\mathbf{y}_{t-p}$. First generate base $\hat{\mathbf{y}}_{t+1}$, then reconcile it via $\tilde{\mathbf{y}}_{t+1} = \mathbf{SP}\hat{\mathbf{y}}_{t+1}$, and then generate $\hat{\mathbf{y}}_{t+2}$ using *the reconciled values* $\tilde{\mathbf{y}}_{t+1}$ on the right hand side. This is conceptually different than generating a series of base forecasts for $\hat{\mathbf{y}}_{T+1:T+h} = \hat{\mathbf{y}}_{t+1}, \hat{\mathbf{y}}_{t+2}, \ldots$ and then reconciling them following Wickramasuriya et al. (2019)'s equation using an appropriate $\mathbf{W}_h$ matrices for $h \in [1, 2, 3, \ldots]$. We argue from a policy perspective the path of consecutively reconciled forecasts is more appropriate, i.e. once the economy arrives in $\tilde{\mathbf{y}}_{t+1}$ it should continue from there, whereas in the alternative scenario the two-step ahead forecast $\tilde{\mathbf{y}}_{t+2}$ is a reconciliation of $\hat{\mathbf{y}}_{t+2}$ that was generated originating from $\hat{\mathbf{y}}_{t+1}$.

2.4 Impulse response function reconciliation

To extend the framework to structural analysis, we rely on the property that IRFs can be computed as the difference between a forecast subject to a specific shock and a forecast in the absence of such disturbances. Given this definition, structural analysis faces the same challenge as forecasting. Without explicit constraints, there is no mechanism that ensures that the IRFs of the lower level of the hierarchy will aggregate to the IRF of the upper level. Any incoherence of the underlying forecasts propagates directly to the structural responses.

To address this, we propose deriving reconciled impulse response functions. By applying the forecast reconciliation procedure described above *prior* to computing the difference between the shocked and un-shocked paths, we ensure that the resulting IRFs inherently satisfy the cross-sectional constraints of the hierarchy.

Base impulse responses

Following Koop et al. (1996) and Pesaran and Shin (1998), the h-step ahead impulse response is defined as the difference between the respective forecast path subject to a structural shock and the forecast path in its absence.

Consider a VAR of the form of eq. (1) with a structural decomposition of the residuals

$$\mathbf{y}_t = \mathbf{b}_0 + \mathbf{B}_1\mathbf{y}_{t-1} + \cdots + \mathbf{B}_1\mathbf{y}_{t-p} + \mathbf{B}_0\epsilon_t \qquad \epsilon_t \sim \mathcal{N}(\mathbf{0}_n, \mathbf{\Omega}). \tag{19}$$

Here we suppose that $\varepsilon_t = \mathbf{B}_0\epsilon_t$ is a chosen identification strategy that disentangles the residuals ε_t into a vector of structural shocks ϵ, i.e. $\mathbf{B}_0\mathbf{\Omega}\mathbf{B}_0' = \sigma$ with σ defined as in eq. (1).

Without loss of generality consider a VAR(1). The one-step ahead unconditional forecast in the absence of shocks is then given by

$$\hat{\mathbf{y}}_{T+1} = \mathbf{b}_0 + \mathbf{B}_1\hat{\mathbf{y}}_T. \tag{20}$$

Iteration of eq. (19) h-steps ahead from time period $t = T$ onward given information up to point T results in a set of h unconditional forecasts and can be collected in the following matrix

$$\hat{\mathbf{y}}_{T+1:T+h}|\mathbf{B_1}, \mathbf{b_0} = \begin{bmatrix}\hat{\mathbf{y}}_{T+1}, & \dots, & \hat{\mathbf{y}}_{T+h}\end{bmatrix}'. \tag{21}$$

Now, consider that a shock ϵ hits the economy in period $T+1$ and given a suitable identification scheme $\mathbf{B_0}$ yields a vector of contemporaneous impact $\mathbf{b}_{\epsilon_{T+1}} = \mathbf{B_0}\epsilon_{T+1}$.[5] The forecast for the first period with the shock is simply

$$\hat{\mathbf{y}}^{fs}_{T+1} = \mathbf{b}_0 + \mathbf{B}_1\hat{\mathbf{y}}_T + \mathbf{b}_{\epsilon_{\mathbf{T+1}}} \tag{22}$$

and an iteration on (19) for the subsequent $h-1$ periods yields a matrix of forecasts conditional on that shock

$$\hat{\mathbf{y}}^{fs}_{T+1:T+h}|\mathbf{B_1}, \mathbf{b_0}, \mathbf{b}_{\epsilon_{\mathbf{T+1}}} = \begin{bmatrix}\hat{\mathbf{y}}^{fs}_{T+1}, & \dots, & \hat{\mathbf{y}}^{fs}_{T+h}\end{bmatrix}' \tag{23}$$

Note that the IRFs of the system can then be computed by subtracting eq. (21) from (23), which is mathematically equivalent to IRF calculation using the companion form in linear VARs.

$$IRF(h, |\mathbf{B_1}, \mathbf{b_0}, \omega_{t-1}) = \hat{\mathbf{y}}^{fs}_{T+1:T+h} - \hat{\mathbf{y}}^{f}_{T+1:T+h} \tag{24}$$

Furthermore for linear systems this calculation does not depend on the origin point of the forecast (e.g. the choice $t = T$ above). Koop et al. (1996) generalize this to non-linear VAR systems, where the forecasts are done at every time point and thus derive the generalized impulse response functions (GIRF) as the difference of the expectations over forecasts distributions

$$GIRF(h, \omega_{t-1}) = E(\hat{\mathbf{y}}_{t+1:t+h}|\epsilon_t, \omega_{t-1}) - E(\hat{\mathbf{y}}_{t+1:t+h}|\omega_{t-1}). \tag{25}$$

Reconciled impulse responses

Notably, when working with hierarchical macroeconomic data, the forecasts in both $\hat{\mathbf{y}}^{s}_{T+1:T+h}$ and $\hat{\mathbf{y}}^{fs}_{T+1:T+h}$ are incoherent in the sense that the values do not respect the

[5] Note that we do not rely on any specific identification scheme, we only need a vector of contemporaneous impact of the shocks $\mathbf{b}_{\epsilon,t}$. Furthermore, note that $\mathbf{B}_1, \dots, \mathbf{B}_p$ are the reduced-form coefficients and only $\mathbf{B}_0$ requires a structural identification.

hierarchical structure without any additional constraints. Since impulse responses are functions of the forecasts they would also inherit any such inconsistency inherent to the base forecasts.

We propose to solve this problem by introducing *reconciled impulse response functions*, which can be calculated after we reconcile the forecasts in eqs. (20) and (22) as in eq. (10). Given a suitable matrix $\mathbf{\Pi}$, as defined in eq. (16) one can reconcile the forecasts with and without shock, $\hat{\mathbf{y}}^{fs}_{T+1}$ and $\hat{\mathbf{y}}^{f}_{T+1}$, respectively, e.g.

$$\tilde{\mathbf{y}}^{f}_{T+1:T+h} = \mathbf{\Pi}\hat{\mathbf{y}}^{f}_{T+1:T+h}, \quad \text{and} \quad \tilde{\mathbf{y}}^{fs}_{T+1:T+h} = \mathbf{\Pi}\hat{\mathbf{y}}^{fs}_{T+1:T+h} \tag{26}$$

The difference between the two matrices with reconciled forecasts yields an impulse response that respects the hierarchy

$$rIRF(h, \omega_{t-1}) = \tilde{\mathbf{y}}^{fs}_{T+1:T+h} - \tilde{\mathbf{y}}^{f}_{T+1:T+h}. \tag{27}$$

Note, that this method can also be used to reconcile generalized impulse response functions from a non-linear model in the same manner and does not rely in any shape or form on the linear specification or on a specific cross-sectional restriction.

$$rGIRF(h, \omega_{t-1}) = E(\tilde{\mathbf{y}}_{t+1:t+h}|\epsilon_t, \omega_{t-1}) - E(\tilde{\mathbf{y}}_{t+1:t+h}|\omega_{t-1}). \tag{28}$$

Finally, this method of calculating reconciled impulse responses can very easily be implemented by following the setup in Section 2.3. We can apply the above presented logic to generate forecasts conditional on two different paths of shock realizations and reconcile them subsequently. The forecast distribution without shocks is directly sampled with eq. (13), while to impose a shock at period $T+1$, the constant $\mathbf{c}$ is augmented with the contemporaneous vector $\hat{\mathbf{b}}_{\epsilon_\mathbf{t}}$ in the relevant row, i.e. $\mathbf{c}$ is replaced with

$$\mathbf{c}^{\mathrm{s}} = \begin{bmatrix} \mathbf{b} + \sum_{j=1}^{p} \mathbf{B}_j \mathbf{y}_{T+1-j} \\ \mathbf{b} + \sum_{j=2}^{p} \mathbf{B}_j \mathbf{y}_{T+2-j} \\ \vdots \\ \mathbf{b} + \mathbf{B}_p \mathbf{y}_T + \mathbf{b}_{\epsilon_{\mathbf{T+1}}} \\ \mathbf{b} \\ \vdots \\ \mathbf{b} \end{bmatrix}. \tag{29}$$

Next, we illustrate the practical utility of the proposed approach through several applications of forecast reconciliation using the mixed-frequency gaussian framework. First, we show how the method besides preserving the hierarchical structure also improves the forecasts of regional economic aggregates using datasets from the UK and Germany. Then, we address the issue of aggregation bias in structural analysis involving hierarchical macroeconomic data.

3 Applications

In the previous section, we established a theoretical framework that integrates forecast reconciliation into a mixed-frequency VAR setting. We showed how to derive reconciled impulse responses and ensure consistency across hierarchy levels. To demonstrate the usefulness of this framework, we now turn to empirical applications. We focus on two primary exercises: first, improving point and density forecasts of regional economic activity; second, illustrating how the method resolves aggregation bias in structural analysis. Macroeconomic hierarchical datasets present unique challenges, particularly in the realm of regional accounts such as state level GDP and GVA, which serve as a primary example in recent literature (Koop et al., 2020a, 2024a,b,c; Barbaglia et al., 2024; Lehmann, 2024). These datasets typically feature the wide data curse. While the cross-sectional dimension is large (e.g., 12 regions in the UK, 16 in Germany), the time series length is extremely short because regional account data is typically only provided annually. This limitation is compounded by the fact that many region-specific economic indicators have emerged only recently, that is if they are available at all. A critical consequence of these data structures is a severe publication lag. As annual data is typically released well into the subsequent year – ranging from three months in Germany to twelve months in the UK – generating timely insights for the ongoing as well as upcoming year is exceptionally difficult. This creates a disconnect for policymakers, who require current and forward-looking information before data is officially available. For instance, budget planning for the upcoming fiscal year typically occurs towards the end of the current year. At this point in time not even regional accounts for the ongoing calendar year have yet been released. Similarly, even given the recent advancements working with short hierarchical data, inference for the current quarter relies heavily on the availability of quarterly national GDP which is subject to its own publication lag.

In the following applications, we demonstrate how our framework addresses these real-world constraints to generate information beyond the latest available data point. First, we show that our method enables the generation of coherent nowcasts and forecasts, overcoming the reliance on delayed top-level aggregates. Second, the mixed-frequency setting allows us to produce quarterly forecasts even for variables traditionally observed only annually. Though forecast performance is evaluated at an annual frequency, given that the *true* data for regional accounts is only available on a yearly basis . Finally, we illustrate how our approach resolves structural aggregation bias, providing policymakers with impulse response functions that strictly respect the underlying economic hierarchy.

3.1 Forecasting regional economic activity in the UK

In their seminal paper, Koop et al. (2020b), show how the mixed-frequency framework may be used to track regional economic activity in the United Kingdom. They compile a dataset for the twelve UK regions shown in Table 1 (excluding the continental shelf) from 1967 up to 2017.

North East	NE	East Midlands	EM	East Anglia	EA
Yorks	YO	London	LO	South East	SE
South West	SW	West Midlands	WM	North West	NE
Wales	WA	Scotland	SC	Northern Ireland	NI

Table 1: Regions of the United Kingdom and their abbreviations

In the UK, regional accounts are published in a yearly frequency with a delay of 12 months. In December of the current year, say year τ, the regional annual growth rate data for year $\tau - 1$ is published.[6] This raises particular problems not only for assessing the current economic situation but also for the budgetary planning, which require a forecast for the full $\tau + 1$. For example, the Scottish government begins with the budget planning for $\tau + 1$ in December of year τ, while official data is only available for $\tau - 1$. Therefore, for any up-to-date information on regional economic activity the government would need to rely on a nowcast for the current and a forecast for the upcoming year, τ and $\tau + 1$ respectively. These should respect the hierarchy imposed by the geographical order just as ex-post published official data does, in the sense that the regions aggregate to the national, even if the national value is not yet observed.

To provide such forward looking estimates, we begin with the dataset of Koop et al. (2020b) containing annual regional and quarterly national gross value added and use a Bayesian mixed-frequency VAR to generate a balanced quarterly data set. To do so, we model the UK regions and the UK GVA jointly by imposing the cross-sectional constraint in eq. (6). In addition, following Koop et al. (2020b), we assume equal and time-invariant weights for all UK regions. The inter-temporal constraint in eq. (5) ensures that the annual regional growth rates align with the quarterly evolution. In contrast to Koop et al. (2020b) we use a classical Minnesota prior with homoscedastic variances, which is in most cases the most reliant setup for forecasting. Nevertheless, our goal here is not to find the best forecasting model for the UK but to highlight the value added from forecast reconciliation.

While the original dataset begins in 1967 and runs to 2017, we extend it with current data up to 2023.[7] We generate quarterly pseudo out-of sample forecasts starting from

[6] Office for National Statistics: `https://www.ons.gov.uk/economy/grossvalueaddedgva/methodologies/regionalgrossvalueaddedbalancedqmi`.

[7] In the main results we evaluate the performance only until 2019 to avoid the complications from the global pandemic.

1985Q1 until 2019Q4 for up to 8 quarters ahead thus allowing us to calculate the yearly forecasts for the current year τ and the upcoming year $\tau + 1$. We include the twelve regions and quarterly national GVA and the model is estimated using 10 000 draws, of which we save the last 2000.

Since the total number of time series to be reconciled is 13, we need at least 13 out-of-sample forecasts to be able to calculate the forecast error variance-covariance matrix for the one-step ahead quarterly forecast, $\mathbf{W_1}$ from eq. (11) and at least 21 forecasts for $\mathbf{W_8}$, which ensures that we can calculate the yearly forecast for the upcoming year $\tau+1$. This limitation is particularly binding in Q1 of year τ, where information on regional variables is strictly confined to historical observations ending at Q4 of the previous year. Therefore, we begin with the forecast reconciliation in 1990Q1.

This setup generates four annual forecasts for the current year and upcoming year each encompassing one more quarter of observed information. This results in a total of 116 yearly forecast for year τ and 112 for $\tau + 1$ over 29 years. Using these forecasts we calculate the Root Mean Squared Forecast Error (RMSFE) for point forecasts and the Continuous Rank Probability Score (CRPS) as a measure for forecast density. To compare the forecast performance of the base forecasts versus the reconciled forecasts we calculate the relative measures. That is we divide the RMSFE and CRPS of the reconciled and the base forecasts to create their relative counterparts.

$$rRMSFE = \frac{RSMFE_{rec}}{RSMFE_{base}}, \qquad rCRPS = \frac{CRPS_{rec}}{CRPS_{base}}, \tag{30}$$

A value below one suggests that the reconciled measure is better. We use a Diebold-Mariano test to gauge whether the differences are statistically significant. The results are presented in Table 2.

Table 2: Point and density forecast accuracy, coherent versus incoherent forecasts.

		NE	YO	EM	EA	LO	SE	SW	WM	NW	WA	SC	NI
τ	**rRMSFE**	1.01	1.01	1.02	1.00	0.98**	0.99	1.04	0.99	1.00	1.03	1.00	1.00
	rCRPS	1.01	0.96**	1.00	0.96**	0.97**	0.94***	1.00	0.92***	0.95***	0.93***	0.97**	0.99
		NE	**YO**	**EM**	**EA**	**LO**	**SE**	**SW**	**WM**	**NW**	**WA**	**SC**	**NI**
$\tau+1$	**rRMSFE**	1.40	1.20	1.02	0.98	0.94**	0.93*	1.05	0.96	0.97	0.97	1.00	1.01
	rCRPS	1.20	1.05	0.99	0.98	0.98	0.89**	1.07	0.85***	0.91*	0.79***	0.99	1.06

Notes: **rRMSFE** (lower is better) is relative Root-Mean Squared Forecast Error; **rCRPS** (lower is better): relative Continuous Rank Probability Score. Reconciled forecasts versus base forecasts for the official yearly data. Stars indicate whether the reconciled model is statistically better according to a Diebold-Mariano test where H_1 : alternative (reconciled) model is better vs H_0 : alternative (reconciled) model is not better. One star indicates a rejection of H_0 : at the 10% level, two stars for 5%, and three stars for 1%.

Reconciling the point forecasts on average appears to generate similar errors to the base forecasts. We find one significant improvement for the current year τ and two in the case of $\tau + 1$. Out of all combinations there are two point forecasts, for North

East and York in $\tau + 1$, that come out behind relative to the baseline MF-VAR. This result is overall in line with the forecast reconciliation literature, which has shown that reconciled forecasts are typically at least as good as the original forecasts. Furthermore, point forecasts from a large mixed-frequency VAR with a cross-sectional constraint may still retain some degree of coherence. This stems from the fact that the hierarchy is enforced at the final observed period and in general, forecasts produced by a linear model of the form $y^f_{T+1} = \mathcal{F} y_T$ are expected to be close to y_T for large sparse systems where many values in $\mathcal{F}$ are close to 0. Moreover, for horizons larger than the lag length, $h > p$, the forecasts of the linear system converge to the conditional means of the series which are expected to be coherent without any additional constraint.

In this MF setup, forecast reconciliation brings the largest gains when it comes to the forecast density. We find highly statistically significant relative CRPS scores for 8 out of the 12 regions for the current year and in 4 out of the 12 regions for the next year with gains up to 20%. The intuition here is that when using the simulation smoother, the individual draws of the forecasts are highly inconsistent. However, as we use a large number of draws to cover the whole forecasting distribution and average them, these inconsistencies disappear in the median, hence the similar point forecasts. However, by reconciling the individual draws, we achieve a much better forecast density.

Forecast reconciliation versus constraint forecast

If all hierarchical series are forecast in one large system, a simple trick could still render the implementation of the cross-sectional constraint in the forecast horizon feasible without relying on reconciliation. For example, even if the UK GDP is not observed beyond the last period, we may still introduce the following cross sectional constraint such that the UK GDP is a weighted average of the regions

$$0 = -y^{UK}_{T+h} + \sum_i \frac{1}{12} y^i_{T+h}, \tag{31}$$

where y^i_{T+h} denotes the quarterly growth rate of UK region i with $i = 1, \cdots, 12$ for each of the h-step ahead forecasts. This restriction can be cast in the form of eq. (7), because beyond $t = T$, the growth rate of GDP, y^{UK}_{T+h}, appears in the vector of missings $\mathbf{y^m}$ and we can set the relevant entries in $\mathbf{z}$ to 0 on the left hand side.

Next, we consider a model, where the forecasts are constrained using the above cross-sectional restriction beyond the last observed quarterly period T. For a direct comparison with forecast reconciliation, we contrast the relative point and density forecast performance of the two approaches. Therefore, values smaller than one in Table 3 indicate that forecast reconciliation outperforms the direct implementation of a cross-sectional constraint. Overall, we find the same gains of the forecast reconciliation compared to the unconstrained case and thus conclude that it could be a very useful tool in the arsenal of the professional forecaster.

Table 3: Point and density forecast accuracy, reconciled versus constrained forecasts.

		NE	YO	EM	EA	LO	SE	SW	WM	NW	WA	SC	NI
τ	**rRMSFE**	1.00	1.00	1.01	0.99	0.98**	0.99	1.03	0.98**	1.00	1.02	1.01	1
	rCRPS	1.00	0.96***	0.99	0.96**	0.97**	0.95***	0.99	0.92***	0.96***	0.93***	0.98	0.99
		NE	**YO**	**EM**	**EA**	**LO**	**SE**	**SW**	**WM**	**NW**	**WA**	**SC**	**NI**
$\tau+1$	**rRMSFE**	1.03	1.00	0.99	1.00	0.96**	0.96*	1.03	0.97*	0.98	0.99	1.01	1.02
	rCRPS	1.09	0.97	0.97	1.00	1.01	0.91**	1.05	0.87***	0.94	0.81***	1.01	1.07

Notes: **rRMSFE** (lower is better) is relative Root-Mean Squared Forecast Error; **rCRPS** (lower is better): relative Continuous Rank Probability Score. Reconciled forecasts versus constrained base forecasts for the official yearly data. Stars indicate whether the reconciled model is statistically better according to a Diebold-Mariano test where H_1 : alternative (reconciled) model is better vs H_0 : alternative (reconciled) model is not better. One star indicates a rejection of H_0 : at the 10% level, two stars for 5%, and three stars for 1%.

3.2 Forecasting with regional indicators

Next, we consider forecasting regional GDP using separate state-specific models, which would presumably suffer from larger discrepancies between the national and aggregated lower-level forecast. This is a classic case for forecast reconciliation, since the regional forecasts are generated without additional information from the other regions. Moreover, implementing a cross-sectional constraint that encompasses the forecast horizon as depicted in the previous section is no longer viable.

The literature on tracking regional economic activity has explored the use of additional regional economic data. Both Koop et al. (2024a) and Koop et al. (2024b) find that regional indicators improve the quarterly estimates only as long as the national data is not already available. In both cases, however, the additional data is included in the joint MF-VAR framework of Koop et al. (2020b) with the cross-sectional constraint to respect the hierarchy. This setup severely increases the VAR size as even two additional regressors per region will triple the number of variables in an already sizable system. This in turn requires much more aggressive shrinkage that could dilute the benefits of additional indicators even with flexible prior specifications.

Given these dimensionality constraints, we explore an alternative modeling strategy aimed at capturing local dynamics while mitigating the need for heavy shrinkage. In contexts where policymakers have access to informative regional data, it may be feasible to build specialized models for individual, large-scale regions. For example, the state government of the German region of North-Rhine-Westphalia, the largest German region in terms of both GDP and population, uses a MF-VAR to generate forecasts that are used in federal fiscal planning (Blagov and Schmidt, 2022).

We explore the benefits of forecast reconciliation using a regional dataset for Germany, comprising 16 states. Three of them, North-Rhine-Westphalia (NW), Bavaria (BY), and Baden-Württemberg (BW) account for well over half of German GDP and have large statistical offices that collect an array of indicators. For example, all three are large industrial states and collect regional industrial production indices, in contrast to many of the economically smaller states.

BW	Baden-Württemberg	BY	Bavaria
BE	Berlin	BB	Brandenburg
HB	Bremen	HH	Hamburg
HE	Hesse	MV	Mecklenburg-Vorpommern
NI	Lower Saxony	NW	North Rhine-Westphalia
RP	Rhineland-Palatinate	SL	Saarland
SN	Saxony	ST	Saxony-Anhalt
SH	Schleswig-Holstein	TH	Thuringia

Table 4: Regions of Germany and their abbreviations

Therefore, we first estimate a set of five models. One for Germany (DE) as a whole, three separate ones for NW, BY, and BW, and lastly one joint model for the remaining smaller states. In the models for the three big states we include the following quarterly *regional* indicators: Industrial production; New orders in Industry; Unemployment rate; Employment (measured by the persons covered by social security); Consumer price index; and sentiment based economic expectations (except for Bayern). We add quarterly German GDP and the truck-toll index as country-wide indicators, as well as our main variable of interest - annual state GDP. Thus, we have 9 variables per model (8 for BY). For the DE model we add the same variables in their national variation. The remaining states we gather in one model along with the corresponding national quarterly indicators.

As a benchmark, we estimate also the state-of-the art large MF-VAR model with a cross-sectional constraint with all German regions as well as quarterly national GDP, industrial production, new orders in industry, expectations, and the consumer price index. All specifications are estimated using the Minnesota prior and independent-normal inverse Wishart prior on the variance-covariance matrix. We sample 10 000 draws from the conditional posterior distributions and save 2000.

The data span is quite heterogeneous. Annual state-level GDP begins in 1991 and runs until 2024 and most national indicators are available for the same time-period, so the large MF-VAR with only national data has (mostly) the full information content. The regional indicators have varying degree of observations. For example, consumer price index and unemployment is available for almost the full sample for all regions, while the regional industrial indicators for BW along with all employees subject to social security contributions are the shortest starting in 2008. The data for the regional indicators is also the final revision, hence the full exercise is done in pseudo-out-of-sample for all comparisons. More information on the data set can be found in Appendix B.

We have 16+1 hierarchical series to reconcile and we need at least 17+8=25 observations to calculate the forecast-error variance-covariance matrices up to $\mathbf{W_{T+8}}$. We add additional 4 to have it somewhat more precisely estimated. Thus, we start forecast evaluation from 2004Q1 and start the reconciliation 29 quarters later in 2011Q1. In order to deal with the ragged edge at the beginning of the sample, we don't include the indicators that are fully missing, i.e. for the BW model a forecast in 2004 does not yet include industrial production and it is only included after 2011 once we have sufficient observations.

In this application, we employ a stochastic hierarchical aggregation scheme. This involves using state-specific and time-varying weights not only for the cross-sectional constraints in the joint VAR but also for the aggregation weights in the reconciliation matrix $\mathbf{A}$. This captures the significant heterogeneity in the economic contributions of the German states.

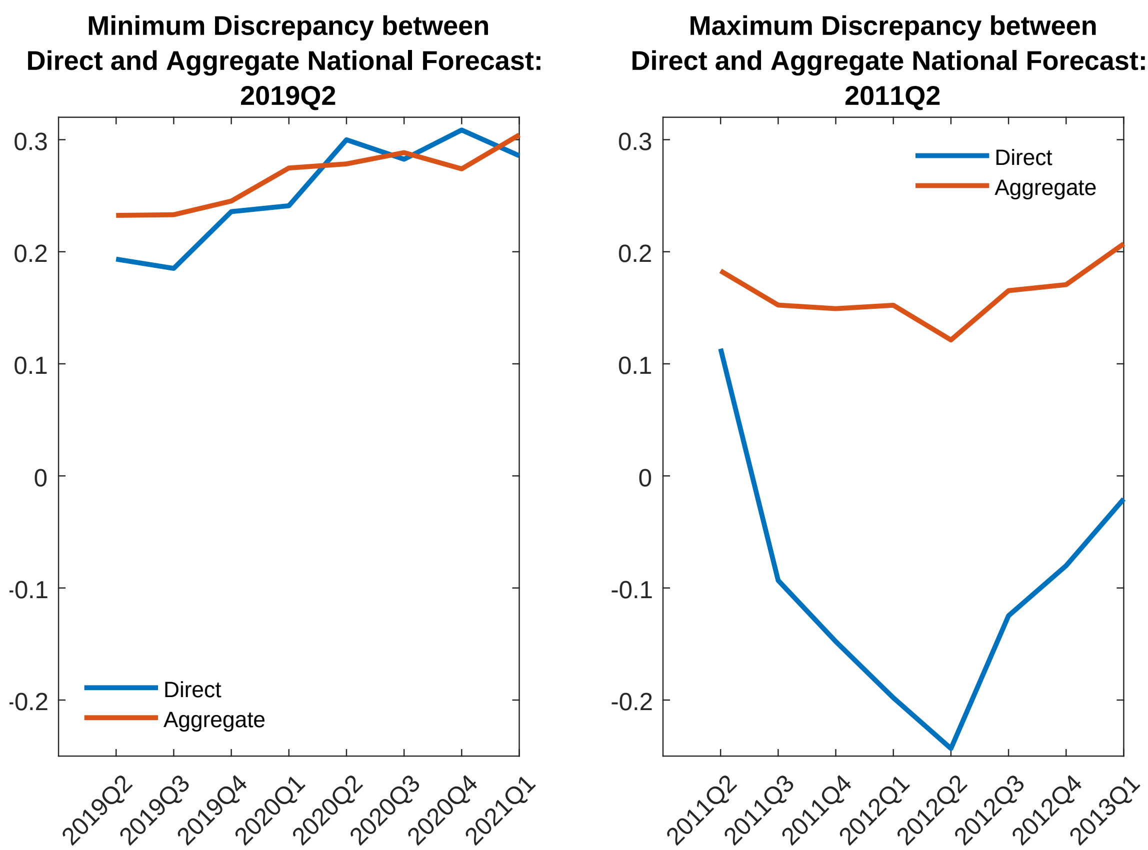

Figure 1: The minimal and maximal discrepancy between the national GDP forecast and the aggregated forecast over the regions. Forecasts are generated by separate models for Germany, Baden-Württemberg, Bavaria, North-Rhine-Westphalia and all remaining states for the time frame spanning 2011Q1 to 2019Q4.

First, we examine the magnitude of the aggregation bias stemming from the base forecasts. To illustrate this, we select the two forecast horizons with the minimal and maximal discrepancy based on the Euclidean norm. Figure 1 depicts these two scenarios. We observe that the maximal discrepancy between the GDP forecast and the implied aggregate (computed as the weighted average of all 16 states) exceeds 0.3 percentage points. Although the Euclidean norm is relatively small for the horizon with minimal discrepancy, the forecast trajectories still show noticeable deviations. This demonstrates that the aggregation bias can reach substantial magnitudes.

We report the forecast comparison of the two reconciled models, namely the separate and joint specifications, in Table 5. We compare the single large MF-VAR model forecast errors with the separate models and report again the RMSFE and the CRPS for the current year (τ) and the next year ($\tau + 1$). Overall using smaller VARs leads to massive improvements for forecasts of the current year, while for the next year the large MF-VAR model has a slight edge when it comes to point forecasting, while it loses w.r.t density forecasts. Thus, for longer forecast horizons the additional information mainly helps with density forecasts, while the point estimates appear similar.

We want to point out that the large increases in the forecast accuracy for the seperate VAR models comes mainly from lower *base* forecast errors, not from the reconciliation itself. That is, using a specific model for BW, BY, and NW leads to significant improvement of the forecast error in τ. What the reconciliation does is that it makes them coherent, solving the main drawback of using individual models.

Table 5: Reconciled forecast errors from a large MF-VAR versus individual models with regional indicators for the largest German states.

	BW	BY	BE	BB	HB	HH	HE	MV	NI	NW	RP	SL	SN	ST	SH	TH
RMSFE τ																
Single MF-VAR	1,08	1,05	1,67	1,64	1,66	1,84	0,79	2,13	2,13	1,06	0,98	1,64	0,84	2,11	1,10	1,08
Multiple MF-VARs	0,57	1,13	1,54	1,55	1,30	1,32	0,76	1,79	1,92	0,86	0,79	1,40	0,73	1,83	0,98	1,01
Ratio	0,53	1,08	0,92	0,94	0,78	0,71	0,96	0,84	0,90	0,81	0,80	0,85	0,87	0,87	0,89	0,94
RMSFE $\tau + 1$																
Single MF-VAR	1,33	1,06	2,39	1,56	1,28	1,35	1,33	2,02	2,10	1,46	0,89	2,58	1,24	1,23	1,33	1,61
Multiple MF-VARs	1,48	1.00	2,04	1,54	1,55	1,31	1,37	2,13	2,17	1,68	0,88	2,74	1,22	1,41	1,29	1,71
Ratio	1,11	0,94	0,85	0,99	1,21	0,97	1,03	1,05	1,03	1,15	0,99	1,06	0,98	1,15	0,97	1,06
CRPS τ																
Single MF-VAR	0,72	0,62	1.00	0,97	0,96	1,07	0,57	1,26	1,22	0,65	0,64	1,02	0,64	1,18	0,68	0,71
Multiple MF-VARs	0,54	0,58	1,12	1,01	1,06	1,12	0,83	0,95	1,56	0,93	0,45	0,75	0,44	0,89	0,50	0,54
Ratio	0,75	0,94	1,12	1,04	1,10	1,05	1,45	0,75	1,28	1,43	0,70	0,73	0,68	0,75	0,73	0,76
CRPS $\tau + 1$																
Single MF-VAR	1,75	1,12	1,67	1,22	1,64	1,02	1,13	1,26	1,57	1,22	1,14	1,99	1,00	1,47	1,08	1,20
Multiple MF-VARs	1,02	0,7	1,79	0,76	1,31	0,84	0,62	1,30	1,25	0,94	0,55	1,54	0,67	0,80	0,75	0,95
Ratio	0,58	0,62	1,07	0,62	0,80	0,82	0,55	1,03	0,79	0,77	0,48	0,77	0,67	0,55	0,69	0,79

Notes: Absolute reconciled root-mean-squared forecast errors (RMSFE) of the yearly forecasts in percentage points and the ratio between them for a single large MF-VAR model versus 5 models, one for BW, BY, NW, DE, and the rest. Evaluation sample 2011Q1 to 2019Q4.

3.3 Impulse response function reconciliation

For our last application, we examine how structural analysis can benefit from forecast reconciliation techniques. Section 2.4 discusses the challenges of calculating impulse response functions with macroeconomic hierarchical data, namely that there is no mechanism that ensures that the response to an economic shock of individual regions would aggregate to the response of the national variable. The severity of this problem could vary across datasets - some could aggregate closer to the direct response of the GDP series than others and estimation uncertainty could also mask the issue.[8] However, it could be enough for one region to have an improper response to skew the national aggregate.

In this section, we analyze the regional effects of an economic activity shock in the UK. We rely again on the dataset from Koop et al. (2020b) for the twelve UK regions (excluding the continental shelf) and UK GVA, which expands from 1976 to 2019 and combines annual and quarterly data.

We estimate a proxy mixed-frequency VAR (MF-VAR IV). The model setup is identical to that in Section 3.1 - a 13 variable MF-VAR model with the usual inter-temporal and cross-sectional restrictions, estimated with a homoskedastic Minnesota prior with 10 000 draws of which 2000 are saved.

As a proxy for the shock we take the series for world economic activity shocks estimated by Baumeister and Hamilton (2019), which we aggregate to quarterly frequency. We follow Mertens and Ravn (2013) and use the residuals from the first stage to orthogonalize the proxy which gives the impact vector $\mathbf{b}_{\epsilon_t}$. Since the vector is identified up to a sign and scale, we assume that a positive economic activity raises the national and regional GDPs and we normalize the scale to 1. We then implement the methodology from sections 2.3 and 2.4 to calculate the forecasts with and without a one standard deviation shock in period $T+1$ and reconcile them.

A key distinction relative to the forecasting application in Section 3.1 focused on the UK concerns the specification of the forecast reconciliation matrix $\mathbf{\Pi}$. The matrix consists of the blocks containing $\mathbf{P_1}$ to $\mathbf{P_h}$ that in turn are ultimately determined by $\mathbf{W_1}$ to $\mathbf{W_h}$. Wickramasuriya et al. (2019) argue that the out-of-sample forecast error variance covariance matrix is the optimal choice for $\mathbf{W}$, as specified in eq. (12). However, when calculating impulse responses that approach might be impractical for several reasons.

First, the impulse response horizon that is of interest is often much longer than the typical forecast horizon. With quarterly data it is usual to look at impulse responses horizons from 8 to 12 quarters, while for monthly data 12 to 24 months are often

[8] As with any econometric technique, in an infinite dataset where the true parameters are recovered this problem should not exist. A shock to all regions would correctly propagate through the regions and thus forecasts and responses should aggregate correctly. This is not necessarily the case in finite samples, and especially in large systems where shrinkage is imposed to make the model tractable.

depicted. This would require to generate forecasts over very long forecast horizons to estimate $\mathbf{W}$ - up to $\mathbf{W_{12}}$ and $\mathbf{W_{24}}$ respectively. However, it is well known, that the predictive content of forecasts deteriorates strongly with forecast horizon (Breitung and Knüppel, 2021). In a VAR framework, the model reverts to the conditional mean for periods $h > p$. Second, these matrices still need to be somewhat precisely estimated, which further reduces the time-span. For 13 series of quarterly data to be reconciled one needs 14 out-of sample forecasts to generate the first $\mathbf{W_1}$ and 26 out-of-sample forecasts for $\mathbf{W_{12}}$. Finally, from a practicality perspective this would require the researcher to do an out-of-sample forecasting exercise for every set of impulse responses, which on a large model can quickly become infeasible.

An alternative estimator for $\mathbf{W}$ discussed in the forecast reconciliation literature is the in-sample variance-covariance matrix (Wickramasuriya et al., 2019; Athanasopoulos et al., 2024). This is advantageous, because it is generated when the model is estimated and thus can be easily incorporated. Therefore, for the IRF analysis we set $\mathbf{W} = \hat{\sigma}$, the estimated counterpart to σ from eq. (1) for all periods. Nevertheless, since we do have the out-of-sample forecast-error matrix $\hat{\mathbf{W}}$ from the forecasting analysis we also estimate the rIRFs using this matrix as a robustness check. In Appendix E we show that for the purpose of structural analysis, both the in-sample and out-of-sample $\hat{\sigma}$ and $\hat{\mathbf{W}}$ matrices deliver similar impulse responses.

To highlight the inherent limitations of structural analysis with hierarchical data , we construct the implied *aggregated* GDP response. We compute this series by weighting the regional responses using the cross-sectional constraint from the MF-VAR, repeating the aggregation procedure for each individual simulation draw. The *base* IRFs, are depicted on the top panel of Figure 2. The national GDP response diverges markedly from that of the aggregated regions. This discrepancy constitutes a clear violation of the aggregation identity even though the MF-VAR model explicitly enforces the cross-sectional restriction in-sample.

On the bottom panel we plot the reconciled impulse responses. Notably, the national response is now fully consistent with the aggregated response across all regions. To obtain the reconciled responses, adjustments to the base IRFs were required. The initial response on impact is weaker than that of the base IRF of the national GDP and stronger than the aggregated GDP series. The individual regions are underestimating the true response, while the national GDP is overestimating it. To facilitate comparison, Figure 3 displays the median responses for the national and aggregate base IRFs, together with the reconciled response (for the reconciled, national GDP and aggregate GDP overlap).

The detailed regional impulse responses are presented in Appendix D. The underlying logic of forecast reconciliation also governs how the IRFs are adjusted. Regions, with historically more accurate base forecasts undergo smaller corrections. In this application, the region of Northern Ireland (NI) is adjusted less than the others and therefore

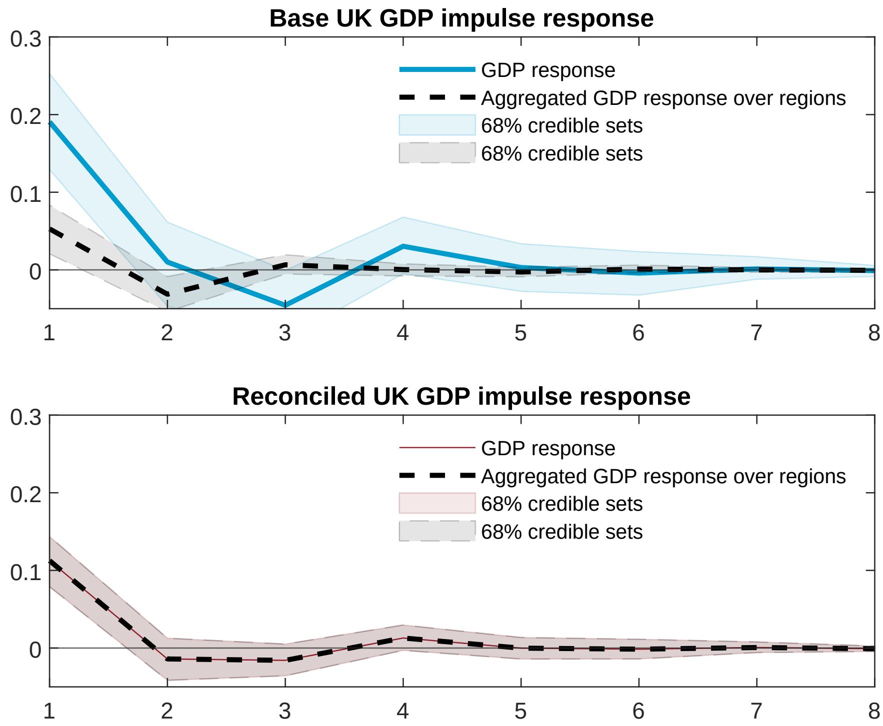

Figure 2: Top panel: Base IRFs of GDP and an aggregated GDP as a weighted average of the regional IRFs, along with 68% probability intervals. Bottom panel: reconciled GDP response and reconciled aggregated GDP response.

the base and reconciled IRFs are much more aligned than for other regions, providing a clear illustration of the core forecast error variance-weighting mechanism of the reconciliation technique.

4 Conclusion

In this paper we introduce a novel approach to use forecast reconciliation within Gaussian linear state-space models for macroeconomic hierarchical data. Since in macroeconomic applications the dataset is often very short, we propose to use a mixed-frequency framework to generate high-frequency data of the unobserved low-frequency variables and treat the estimated missing observations as true data for the implementation of forecast reconciliation. We prove the consistency of this approach and show how it can be used for forecasting macroeconomic hierarchical data. Furthermore, we show how the technique can be framed as a conditional forecast and thus can be easily implemented using popular conditional forecast approaches. This allows for scenario analysis and counterfactual simulation, where the conditional path of the lower-level hierarchy must match the path of the upper-level.

By implementing forecast reconciliation in a macroeconomic hierarchical setup we

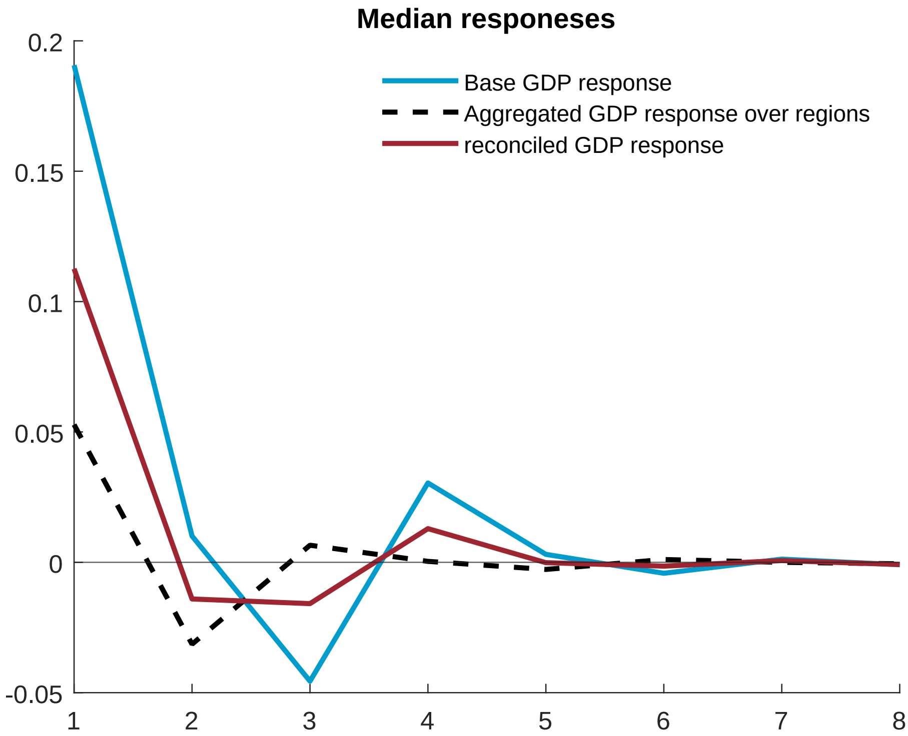

Figure 3: Median responses of base GDP, base aggregated GDP, and reconciled GDP.

show not only that one may generate coherent forecasts but also that the approach can improve both the point and density forecasts in a pseudo out-of-sample forecasting exercise for the UK regions. In a second application, we show how forecasting the German states in seperate models can improve forecasting performance but also increases the need for forecast reconciliation.

Finally, we derive and define the concept of reconciled impulse response functions. By utilizing the property of IRFs to be calculated using forecasting techniques, we can implement forecast reconciliation to ensure that the generated IRFs are coherent across the hierarchical levels. We employ reconciled IRFs using regional UK data demonstrating that the aggregation bias should not be neglected. Thus, our approach allows for studying the potentially heterogeneous impact of various policies and economic developments on the economic regions while maintaining the inherent hierarchy.

References

Antolín-Díaz, J., Petrella, I., and Rubio-Ramírez, J. F. (2021). Structural scenario analysis with SVARs. *Journal of Monetary Economics*, 117:798–815.

Ashouri, M., Hyndman, R. J., and Shmueli, G. (2022). Fast forecast reconciliation using linear models. *Journal of Computational and Graphical Statistics*, 31(1):263–282.

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., Gamakumara, P., Panagiotelis, A., Hyndman, R. J., and Affan, M. (2020). Hierarchical Forecasting. In Fuleky, P., editor, *Macroeconomic Forecasting in the Era of Big Data*, volume 52, pages 689–719. Springer International Publishing, Cham.

Athanasopoulos, G., Hyndman, R. J., Kourentzes, N., and Panagiotelis, A. (2024). Forecast reconciliation: A review. *International Journal of Forecasting*, 40(2):430–456.

Barbaglia, L., Frattarolo, L., Hauzenberger, N., Hirschbuehl, D., Huber, F., Onorante, L., Pfarrhofer, M., and Pezzoli, L. T. (2024). Nowcasting economic activity in European regions using a mixed-frequency dynamic factor model.

Baumeister, C. and Hamilton, J. D. (2019). Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks. *American Economic Review*, 109(5):1873–1910.

Bisaglia, L., Di Fonzo, T., and Girolimetto, D. (2020). Fully reconciled GDP forecasts from Income and Expenditure sides.

Blagov, B. and Schmidt, T. (2022). Schätzung der Wirtschaftsentwicklung in NRW im dritten Quartal 2022: Ein Mixed-Frequency-Ansatz. *RWI Konjunkturberichte*, 73(4):53–59.

Breitung, J. and Knüppel, M. (2021). How far can we forecast? Statistical tests of the predictive content. *Journal of Applied Econometrics*, 36(4):369–392.

Chan, J. C. 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.

Chan, J. C. C., Pettenuzzo, D., Zhu, D., and Poon, A. (2025). Conditional forecasts in large Bayesian VARs with multiple equality and inequality constraints. *Journal of Economic Dynamics and Control*, forthcoming.

Chan, J. C. C., Poon, A., and Zhu, D. (2023). High-dimensional conditionally gaussian state space models with missing data. *Journal of Econometrics*, 236(1):105468.

Di Fonzo, T. and Girolimetto, D. (2023). Cross-temporal forecast reconciliation: Optimal combination method and heuristic alternatives. *International Journal of Forecasting*, 39(1):39–57.

Eckert, F., Hyndman, R. J., and Panagiotelis, A. (2021). Forecasting Swiss exports using Bayesian forecast reconciliation. *European Journal of Operational Research*, 291(2):693–710.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, Princeton, N.J.

Hauber, P. and Schumacher, C. (2021). Precision-based sampling with missing observations: A factor model application. Discussion Paper No. 11, Deutsche Bundesbank.

Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., and Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics & Data Analysis*, 55(9):2579–2589.

Koop, G., McIntyre, S., and Mitchell, J. (2020a). UK regional nowcasting using a mixed frequency vector auto-regressive model with entropic tilting. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 183(1):91–119.

Koop, G., McIntyre, S., Mitchell, J., and Poon, A. (2020b). Regional output growth in the United Kingdom: More timely and higher frequency estimates from 1970. *Journal of Applied Econometrics*, 35(2):176–197.

Koop, G., McIntyre, S., Mitchell, J., and Poon, A. (2024a). Using stochastic hierarchical aggregation constraints to nowcast regional economic aggregates. *International Journal of Forecasting*, 40(2):626–640.

Koop, G., McIntyre, S., Mitchell, J., Poon, A., and Wu, P. (2024b). Incorporating short data into large mixed-frequency VARs for regional nowcasting. *Journal of the Royal Statistical Society: Series A*, 187(2):477–495.

Koop, G., McIntyre, S., Mitchell, J., and Wu, P. (2024c). Measuring sub-regional economic activity: Missing frequencies and missing data. ESCoE Discussion Paper 2024-11, Economic Statistics Centre of Excellence, London, United Kingdom.

Koop, G., Pesaran, M. H., and Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics*, 74(1):119–147.

Lehmann, R. (2024). A real-time regional accounts database for Germany with applications to GDP revisions and nowcasting. *Empirical Economics*, 67(2):817–838.

Lehmann, R. and Wikman, I. (2025). Quarterly GDP estimates for the german states: New data for business cycle analyses and long-run dynamics. *Oxford Bulletin of Economics and Statistics*, forthcoming.

Mariano, R. S. and Murasawa, Y. (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics*, 18(4):427–443.

Mertens, E. (2023). Precision-based sampling for state space models that have no measurement error. *Journal of Economic Dynamics and Control*, 154:104720.

Mertens, K. and Ravn, M. O. (2013). The dynamic effects of personal and corporate income tax changes in the United States. *American Economic Review*, 103(4):1212–1247.

Panagiotelis, A., Athanasopoulos, G., Gamakumara, P., and Hyndman, R. J. (2021). Forecast reconciliation: A geometric view with new insights on bias correction. *International Journal of Forecasting*, 37(1):343–359.

Pesaran, M. H. and Shin, Y. (1998). Generalized Impulse Response Analysis in Linear Multivariate Models. *Economics Letters*, 58(1):17–29.

Van Erven, T. and Cugliari, J. (2015). Game-Theoretically Optimal Reconciliation of Contemporaneous Hierarchical Time Series Forecasts. In Antoniadis, A., Poggi, J.-M., and Brossat, X., editors, *Modeling and Stochastic Learning for Forecasting in High Dimensions*, volume 217, pages 297–317. Springer International Publishing, Cham.

Wickramasuriya, S. L. (2021). Properties of point forecast reconciliation approaches.

Wickramasuriya, S. L. (2024). Probabilistic Forecast Reconciliation under the Gaussian Framework. *Journal of Business & Economic Statistics*, 42(1):272–285.

Wickramasuriya, S. L., Athanasopoulos, G., and Hyndman, R. J. (2019). Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *Journal of the American Statistical Association*, 114(526):804–819.

Appendix A Proof of convergence of $\mathbf{W_h}$

Let $\hat{\mathbf{y}}_{T+h}$ denote a vector of $\mathbf{h}$-step ahead forecasts based on the information set $\mathcal{I}_T$ up to time T, and let $\mathbf{e}^{true} = \hat{\mathbf{y}}_{T+h} - \mathbf{y}^{true}_{T+h}$ be the forecast error based on the true, unobserved data. The true forecast error variance-covariance matrix is defined as:

$$\mathbf{W_h} = \text{Var}\left(\hat{\mathbf{y}}_{T+h} - \mathbf{y}^{true}_{T+h} \mid \mathcal{I}_T\right) \tag{A.1}$$

When the true data is unobserved, we approximate $\mathbf{W_h}$ with $\mathbf{W_h^m}$ by substituting $\mathbf{y}^{true}_{T+h}$ with a vector of estimated data, $\hat{\mathbf{y}}^{\mathbf{m}}_{T+h}$. We state this convergence property asymptotically.

Proposition 1. *Let $\mathbf{y}^m$ be an estimator of $\mathbf{y}^{true}$ derived from a state-space latent variable model. If $\mathbf{y}^m \xrightarrow{d} \mathbf{y}^{true}$ as $T \to \infty$, then by continuous mapping and Slutsky's theorem, the tracking errors and conditional variance matrices converge in distribution:*

$$\hat{\mathbf{y}}_{T+h} - \hat{\mathbf{y}}^{\mathbf{m}}_{T+h} \xrightarrow{d} \hat{\mathbf{y}}_{T+h} - \mathbf{y}^{true}_{T+h} \quad \textit{as } T \to \infty, \tag{A.2}$$

$$\mathbf{e^m} \xrightarrow{d} \mathbf{e}^{true}, \tag{A.3}$$

$$\mathbf{W_h^m} \xrightarrow{d} \mathbf{W_h}. \tag{A.4}$$

To establish the validity of A.4, we outline the explicit convergence proofs for both the Kalman-Bucy filter and the precision-based sampler under the following structural setup and dicsuss their shortcomings in practice.

Appendix A.1 The Kalman-Bucy filter

Consider the time-invariant state-space system:

$$x_{t+1} = \mathbf{F}x_t + \nu_t, \qquad \nu_t \sim N(0, \mathbf{Q}), \tag{A.5}$$

$$y_{t+1} = \mathbf{H}'x_t + w_t, \qquad w_t \sim N(0, \mathbf{R}). \tag{A.6}$$

Assumption 1. *The system matrices $\mathbf{F}$ and $\mathbf{H}$ are time-invariant. The pair $(\mathbf{F}, \mathbf{Q}^{1/2})$ is stabilizable, and the pair $(\mathbf{F}, \mathbf{H})$ is detectable.*

Assumption 2. *The true parameter vector $\theta_0 = \{\mathbf{F}, \mathbf{Q}, \mathbf{R}\}$ is either known or replaced by a $\sqrt{T}$-consistent estimator $\hat{\theta}_T$ such that $\hat{\theta}_T \xrightarrow{p} \theta_0$ as $T \to \infty$.*

Define $\hat{x}_t := E\left(x_t \mid \psi_{t-1}\right)$ as the conditional expectation given the history ψ_{t-1}. The sequential updating step is:

$$\hat{x}_{t+1} = \left(\mathbf{F} - \mathbf{K}_t\mathbf{H}'\right)\hat{x}_t + \mathbf{K}_t y_t, \tag{A.7}$$

where $\mathbf{K}_t$ is the time-varying Kalman gain. Given Assumption 1, Proposition 13.1 in Hamilton (1994) guarantees that the Kalman gain uniquely converges to a steady-state, time-invariant matrix: $\lim_{t\to\infty} \mathbf{K}_t = \mathbf{K}$. The steady-state filter is thus:

$$\hat{x}_{t+1} = (\mathbf{F} - \mathbf{K}\mathbf{H}')\,\hat{x}_t + \mathbf{K}y_t. \tag{A.8}$$

Defining the tracking error as $e_t := x_t - \hat{x}_t$ and subtracting (A.8) from (A.5) yields the tracking error dynamics:

$$e_{t+1} = (\mathbf{F} - \mathbf{K}\mathbf{H}')\,e_t + (\nu_t - \mathbf{K}w_t)\,. \tag{A.9}$$

Under Assumption 1, Proposition 13.2 in Hamilton (1994) establishes that all eigenvalues of the matrix $(\mathbf{F} - \mathbf{K}\mathbf{H}')$ lie strictly within the unit circle. Since the innovations ν_t and w_t are stationary white noise processes, the error process e_t is strictly stationary and ergodic. As $t \to \infty$, the initial condition error vanishes, yielding:

$$\hat{x}_t \xrightarrow{d} x_t \implies \mathbf{y}^m \xrightarrow{d} \mathbf{y}^{true}. \tag{A.10}$$

By application of Slutsky's Theorem under Assumption 2, substituting θ with $\hat{\theta}_T$ preserves this asymptotic distribution. $\square$

While the proof holds asymptotically, every state-space system is finite. For any finite dataset of length T, the Kalman filter may not have completely converged to its steady state, meaning $\mathbf{W}_\mathbf{h}^\mathbf{m}$ could still suffer from transient estimation error. Thus, the asymptotic result may not hold perfectly for a finite-sample forecast at time T.

Appendix A.2 Precision-based sampler

Following Chan et al. (2023) and Chan and Jeliazkov (2009), the precision sampler reformulates the full multi-period trajectory.

The joint conditional distribution of the unobserved state vector x given the data vector y and parameters θ is Gaussian:

$$[x \mid y, \theta] \sim N\left(\tilde{x}, \mathbf{K}_P^{-1}\right), \tag{A.11}$$

where $\mathbf{K}_P$ represents the global sparse precision matrix, and $\tilde{x} = E(x \mid y, \theta)$ corresponds exactly to the global conditional mean derived from the fixed-interval Kalman smoother.

To establish convergence for the sampler, a random draw trajectory x^m from this posterior distribution is decomposed into its deterministic conditional mean and a stochastic simulation residual:

$$x^m = \tilde{x} + \boldsymbol{\zeta}, \qquad \boldsymbol{\zeta} \sim N(0, \mathbf{K}_P^{-1}). \tag{A.12}$$

The total tracking error of the sampler can then be structurally decomposed as:

$$x^{true} - x^m = \underbrace{(x^{true} - \tilde{x})}_{\text{Smoothing Error}} - \underbrace{\boldsymbol{\zeta}}_{\text{Simulation Error}} . \tag{A.13}$$

Since the Kalman smoothing error $(x^{true} - \tilde{x})$ is stationary and ergodic under the eigenvalues properties proved in (A.9), and the simulation error $\boldsymbol{\zeta}$ is drawn from a bounded, stable precision matrix $\mathbf{K}_P$ governed by the stable system dynamics, the combined tracking error $(x^{true} - x^m)$ is globally stationary.

Therefore, as $T \to \infty$, the impact of initial state distributions converges to zero, ensuring that realizations from the precision sampler satisfy $\mathbf{y}^m \xrightarrow{d} \mathbf{y}^{true}$. $\square$

Appendix B Data Sources

Note on how industrial production time series was constructed. For the monthly indicators the time of download is not as important since they aren't heavily revised. And if they are, only the most recent data points get overhauled. Data before 2024 should be unaffected.

Table B.1: Reconciled forecast errors from a large MF-VAR versus individual models with regional indicators for the largest German states.

	Region	Source	Time	Release/Download
GDP	DE	Deutsche Bundesbank	1992 - 2024	February 2025
GDP	All States	AK VGRdL	1992 - 2024	March 2025
Truck Toll Mileage Index	DE	Genesis: 42191-0001	2005 - 2024	June 2026
Industrial Production	DE	Genesis: 42153-0001	1992 - 2024	May 2026
Industrial Production	BW	Genesis BW: 42153-0001	2000 - 2024	May 2026
Industrial Production	BY		2008 - 2024	
Industrial Production	NW	Genesis NW: 42153	2005 - 2024	May 2026
New Orders in Manufacturing	DE	Genesis: 42151-0001	1992 - 2024	June 2026
New Orders in Manufacturing	BW	Genesis BW: 42151-0001	1995-2024	May 2026
New Orders in Manufacturing	BY		2008 - 2024	
New Orders in Manufacturing	NW	Genesis NW: 42151	2005-2024	May 2026
CPI	DE	Genesis: 61111-0002	1992 - 2024	June 2026
CPI	BW, BY, NW	Genesis: 61111-0011	1992 - 2024	June 2026
Expectations: Total Business Climate	DE	Ifo Institute	2005 - 2024	June 2026
Expectations: Total Business Climate	BW	L-Bank Staatsbank für Baden-Württemberg	2005 - 2024	May 2026
Expectations: Total Business Climate	NW	NRW Bank	2005-2024	May 2026
Unemployment Rate	DE	Agentur für Arbeit	1992 - 2024	June 2026
Unemployment Rate	BW, BY, NW	Genesis: 13211-0008	1994 - 2024	May 2026
Employees subject to social security contributions	DE	Genesis: 13111-0003	2008 - 2024	June 2026
Employees subject to social security contributions	BW, BY, NW	Genesis: 13111-0006	2008 - 2024	June 2026

Notes: The official statistical data needs to be collected from different sources. German national as well as selected regional data (Genesis): https://genesis.destatis.de/datenbank/online/ Baden-Württemberg (Genesis BW): https://daten.statistik-bw.de/genesisonline/online North-Rhine-Westphalia (Genesis NW): https://www.landesdatenbank.nrw.de/ldbnrw/online/ and the state codes are in line with Table 4: Germany (DE), Baden-Württemberg (BW), Bavaria (BY) and North-Rhine-Westphalia (NW).

Appendix C Structural and reduced form forecasts

Given the equation for the forecast distribution

$$\mathbf{y}_{T+1:T+h} = \mathbf{H}^{-1}\mathbf{c} + \mathbf{H}^{-1}\varepsilon_{T+1:T+h} \sim \mathcal{N}(\mathbf{H}^{-1}\mathbf{c}, \mathbf{H}^{-1}\boldsymbol{\Xi}\mathbf{H}^{-1\prime}). \tag{C.14}$$

The reduced form forecast is given by $\mathbf{H} = \mathcal{H}_{\mathbf{B}}$, $\mathbf{c} = \mathbf{c_b}$, and $\boldsymbol{\Xi} = \boldsymbol{\Sigma}_{T+1:T+h\times T+1:T+h}$ where

$$\mathbf{c_b} = \begin{bmatrix} \mathbf{b} + \sum_{j=1}^{p} \mathbf{B}_j \mathbf{y}_{T+1-j} \\ \mathbf{b} + \sum_{j=2}^{p} \mathbf{B}_j \mathbf{y}_{T+2-j} \\ \vdots \\ \mathbf{b} + \mathbf{B}_p \mathbf{y}_T \\ \mathbf{b} \\ \vdots \\ \mathbf{b} \end{bmatrix}, \mathcal{H}_{\mathbf{B}} = \begin{bmatrix} \mathbf{I}_n & \mathbf{0}_{n\times n} & \cdots & \cdots & \cdots & \cdots & \cdots & \mathbf{0}_{n\times n} \\ -\mathbf{B}_1 & \mathbf{I}_n & \mathbf{0}_{n\times n} & \cdots & \cdots & \cdots & \cdots & \mathbf{0}_{n\times n} \\ -\mathbf{B}_2 & -\mathbf{B}_1 & \mathbf{I}_n & \mathbf{0}_{n\times n} & \cdots & & & \mathbf{0}_{n\times n} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & & \vdots \\ -\mathbf{B}_p & \cdots & & -\mathbf{B}_1 & \mathbf{I}_n & \mathbf{0}_{n\times n} & & \vdots \\ \mathbf{0}_{n\times n} & & & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0}_{n\times n} & \cdots & \mathbf{0}_{n\times n} & -\mathbf{B}_p & \cdots & -\mathbf{B}_2 & -\mathbf{B}_1 & \mathbf{I}_n \end{bmatrix}.$$

Note that the difference here between $\mathcal{H}_{\mathbf{B}}$, $\mathbf{c_b}$ and their counterparts from eq. (2), $\mathbf{H_B}$ and $\mathbf{c_B}$, is simply that the former begin at $T-p$.

The structural forecast model is given by $\mathbf{H} = \mathcal{H}_{\mathbf{A}}$, $\mathbf{c} = \mathbf{c_a}$, and $\boldsymbol{\Xi} = \mathbf{I}$ with

$$\mathbf{c_a} = \begin{bmatrix} \mathbf{a} + \sum_{j=1}^{p} \mathbf{A}_j \mathbf{y}_{T+1-j} \\ \mathbf{a} + \sum_{j=2}^{p} \mathbf{A}_j \mathbf{y}_{T+2-j} \\ \vdots \\ \mathbf{a} + \mathbf{A}_p \mathbf{y}_T \\ \mathbf{a} \\ \vdots \\ \mathbf{a} \end{bmatrix}, \mathcal{H}_{\mathbf{A}} = \begin{bmatrix} \mathbf{A_0} & \mathbf{0}_{n\times n} & \cdots & \cdots & \cdots & \cdots & \cdots & \mathbf{0}_{n\times n} \\ -\mathbf{A}_1 & \mathbf{A_0} & \mathbf{0}_{n\times n} & \cdots & \cdots & \cdots & \cdots & \mathbf{0}_{n\times n} \\ -\mathbf{A}_2 & -\mathbf{A}_1 & \mathbf{A_0} & \mathbf{0}_{n\times n} & \cdots & & & \mathbf{0}_{n\times n} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & & \vdots \\ -\mathbf{A}_p & \cdots & & -\mathbf{A}_1 & \mathbf{A_0} & \mathbf{0}_{n\times n} & & \vdots \\ \mathbf{0}_{n\times n} & & & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0}_{n\times n} & \cdots & \mathbf{0}_{n\times n} & -\mathbf{A}_p & \cdots & -\mathbf{A}_2 & -\mathbf{A}_1 & \mathbf{A_0} \end{bmatrix}$$

for structural matrices $\mathbf{A_0}, \ldots, \mathbf{A_p}$ that correspond to $\mathbf{B_1} = \mathbf{A_0}^{-1}\mathbf{A_1}; \ldots; \mathbf{B_p} = \mathbf{A_0}^{-1}\mathbf{A_p}$.

Appendix D Impulse responses for the UK regions

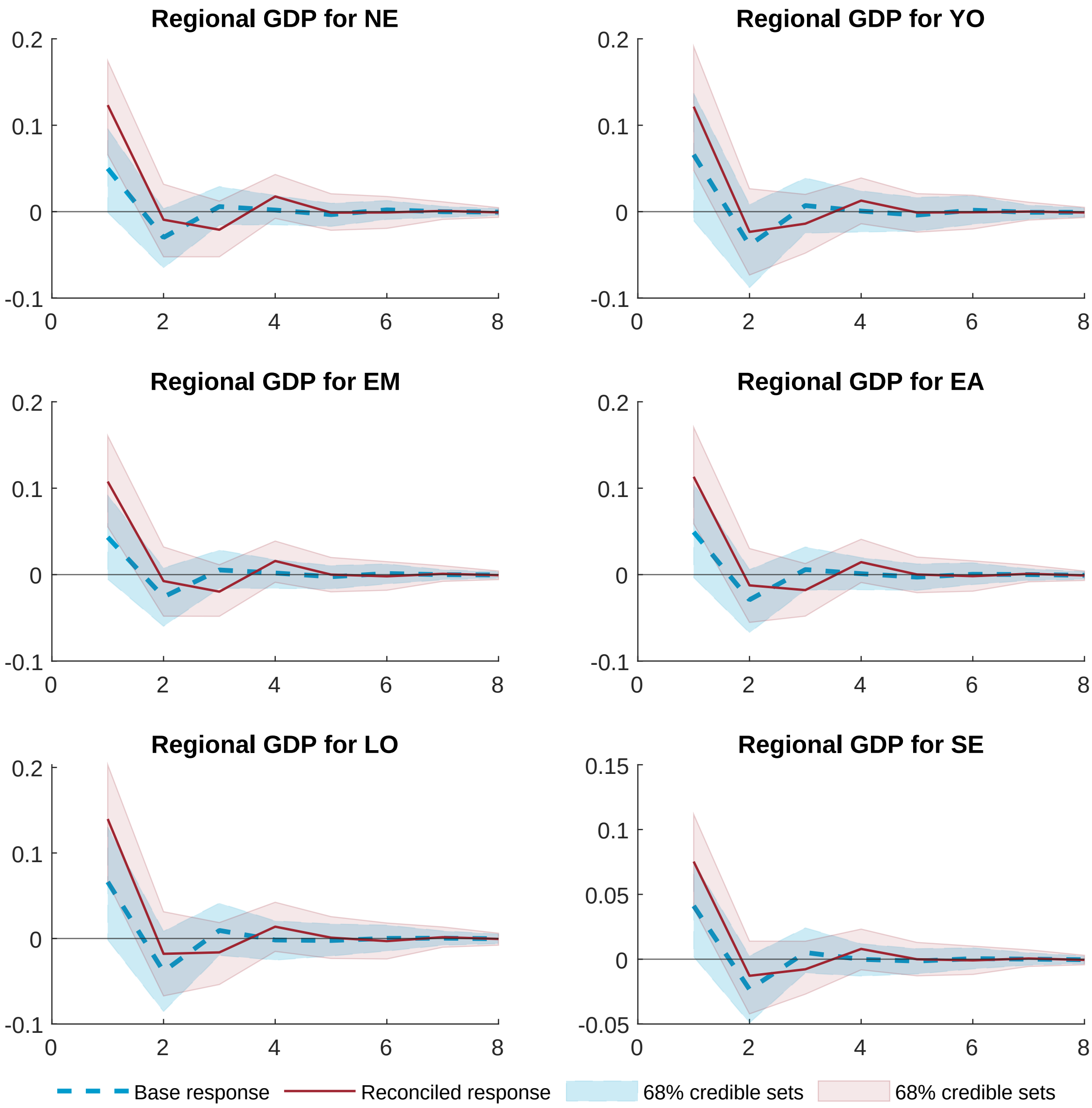

Figure D.1: Base and reconciled impulse response functions for the UK regions.

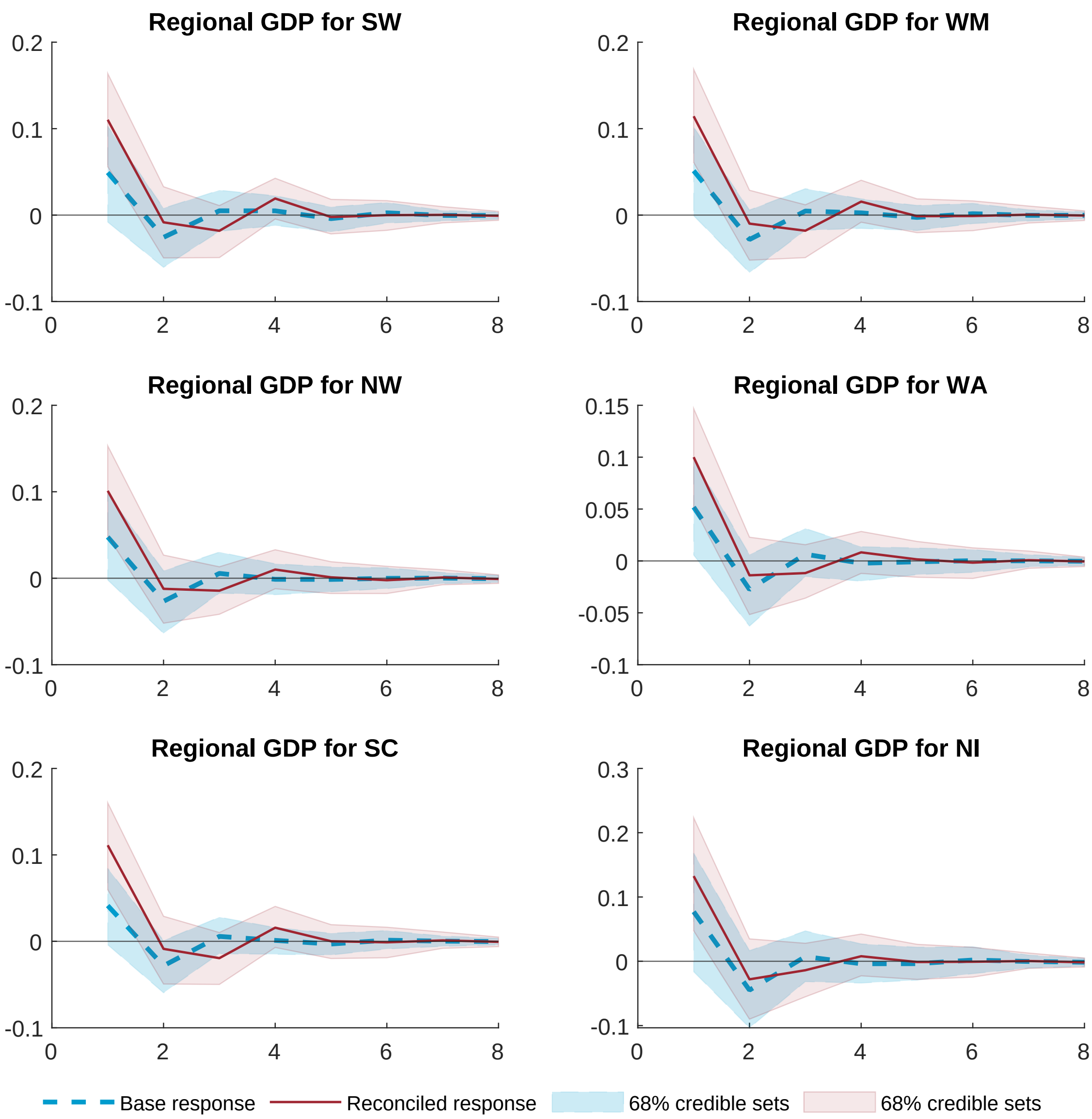

Figure D.2: Base and reconciled impulse response functions for the UK regions (continued).

Appendix E Robustness check for the impulse responses

Here, we estimate the rIRF with the out-of-sample forecast error variance covariance matrix $\mathbf{W}$ from the forecasting application in Section 3.1 instead of the in-sample variance-covariance matrix. We see that the reconciled response is for all intents and purposes the same as Figure 2.

While this choice is considered optimal in the forecast reconciliation literature, it relies on a very long estimation procedure, which might be impractical for structural analysis, where often various specifications are considiered and the model is re-estimated multiple times.

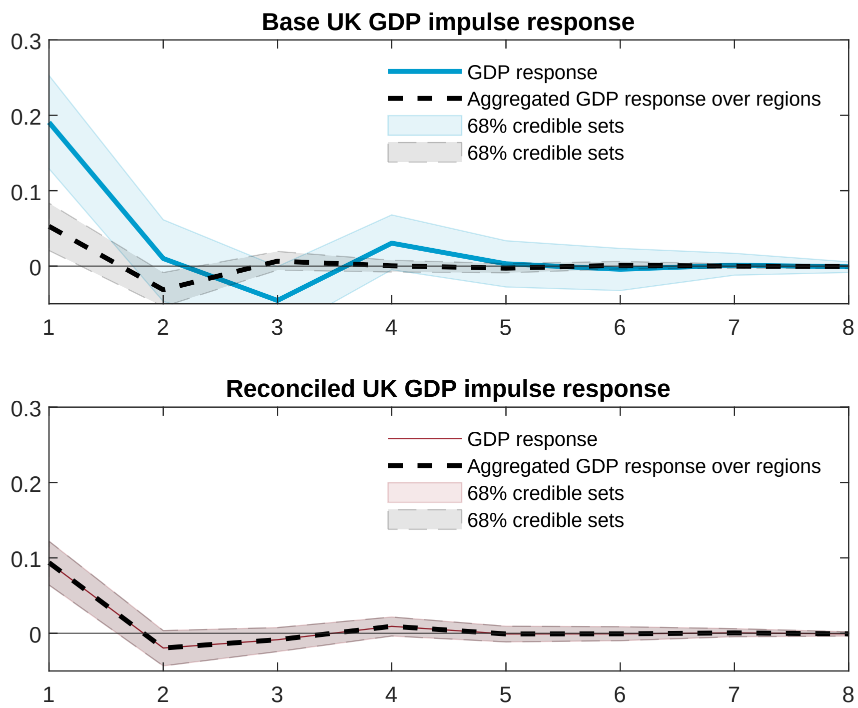

Figure E.3: Base and reconciled impulse response functions for UK GDP, estimated using the out-of-sample forecast error variance covariance matrix.