

Coarsened Bayesian VARs
Correcting BVARs for Incorrect
Specification

Florian Huber
Massimiliano Marcellino
Tobias Scheckel

Coarsened Bayesian VARs

Correcting BVARs for Incorrect Specification

Florian Huber (University of Salzburg), Massimiliano Marcellino (Bocconi University), Tobias Scheckel (WIFO)

WIFO Working Papers 714/2025
October 2025

Abstract

Model misspecification in multivariate econometric models can strongly influence estimates of quantities of interest such as structural parameters, forecast distributions or responses to structural shocks, even more so if higher-order forecasts or responses are considered, due to parameter convolution. We propose a simple method for addressing these specification issues in the context of Bayesian VARs. Our method, called coarsened Bayesian VARs (cBVARs), replaces the exact likelihood with a coarsened likelihood that takes into account that the model might be misspecified along important but unknown dimensions. Since endogenous variables in a VAR can feature different degrees of misspecification, our model allows for this and automatically detects the degree of misspecification. The resulting cBVARs perform well in simulations for several types of misspecification. Applied to US data, cBVARs improve point and density forecasts compared to standard BVARs.

E-Mail: tobias.scheckel@wifo.ac.at

2025/1/W/25070

© 2025 Austrian Institute of Economic Research

Media owner (publisher), producer: Austrian Institute of Economic Research
1030 Vienna, Arsenal, Objekt 20 | Tel. (43 1) 798 26 01 0 | <https://www.wifo.ac.at>
Place of publishing and production: Vienna

WIFO Working Papers are not peer reviewed and are not necessarily based on a coordinated position of WIFO. The authors were informed about the Guidelines for Good Scientific Practice of the Austrian Agency for Research Integrity (ÖAWI), in particular with regard to the documentation of all elements necessary for the replicability of the results.

Free download: <https://www.wifo.ac.at/publication/pid/63690251>

Coarsened Bayesian VARs

Correcting BVARs for Incorrect Specification *

Florian HUBER^a, Massimiliano MARCELLINO^{b,c}, and Tobias SCHECKEL^{a,b}

^a*University of Salzburg*

^b*Bocconi University*

^cCEPR, IGER, BAFFI, BIDS

September 1, 2025

Abstract

Model misspecification in multivariate econometric models can strongly influence estimates of quantities of interest such as structural parameters, forecast distributions or responses to structural shocks, even more so if higher-order forecasts or responses are considered, due to parameter convolution. We propose a simple method for addressing these specification issues in the context of Bayesian VARs. Our method, called coarsened Bayesian VARs (cBVARs), replaces the exact likelihood with a coarsened likelihood that takes into account that the model might be misspecified along important but unknown dimensions. Since endogenous variables in a VAR can feature different degrees of misspecification, our model allows for this and automatically detects the degree of misspecification. The resulting cBVARs perform well in simulations for several types of misspecification. Applied to US data, cBVARs improve point and density forecasts compared to standard BVARs.

JEL Codes: C11, C32, C53

Keywords: Approximate inference, Asymmetric Conjugate Prior, Bayesian VARs, likelihood tempering.

* *Corresponding author:* Florian Huber, Department of Economics, University of Salzburg. *Address:* Mönchsberg 2A, 5020-Salzburg, Austria. *Email:* florian.huber@plus.ac.at. Huber gratefully acknowledges financial support from the Austrian Science Fund (FWF, grant no. ZK 35). Scheckel gratefully acknowledges funding from the Jubiläumsfond of the Oesterreichische Nationalbank (OeNB, grant no. JF-18763). This study was funded by the European Union (NextGenerationEU, Mission 4, Component 2), in the framework of the GRINS (Growing Resilient, INclusive and Sustainable) project (GRINS PE00000018, CUP B43C22000760006). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them. We would like to thank Atsushi Inoue, the handling editor, an anonymous associate editor, two referees, Luca Barbaglia, Tony Chernis, Todd Clark, Niko Hauzenberger, Gary Koop, Dimitris Korobilis, Fabian Krüger, Christian Matthes, Stuart McIntyre, Elmar Mertens, Luca Onorante, Anthoulla Phella, Francesco Ravazollo, Cathy Yi-Hsuan Chen and participants at research seminars at the University of Strathclyde and University of Glasgow for helpful comments and suggestions. The authors acknowledge the computational resources and services provided by Salzburg Collaborative Computing (SCC), funded by the Federal Ministry of Education, Science and Research (BMBWF) and the State of Salzburg.

1 Introduction

When working with multivariate econometric models, challenges like measurement errors, structural breaks, outliers, or non-Gaussian characteristics, including asymmetric and fat-tailed errors, frequently arise. Overlooking these in the model formulation results in misspecification. In a Bayesian framework, this suggests that the posterior distribution concentrates on the parameter value that reduces the Kullback-Leibler divergence between the true data-generating process and the chosen model. Consequently, the estimate found is essentially a pseudo-true parameter, which can be deceptive, especially if misspecification is significant (see, e.g., Müller, 2013). Under such circumstances, posterior uncertainty assessments, such as credible intervals, might become excessively narrow around incorrect values, creating a false sense of certainty.

Especially in recessionary episodes or when large shocks hit the economy (e.g., the global financial crisis or the Covid-19 pandemic), violations of the standard model assumptions become prevalent. Moreover, selecting appropriate variables is necessary to strike a balance between a large model, which might include many irrelevant predictors, and a small tractable model, which could omit important information. These issues can be partly addressed by making the model larger, more flexible, and non-Gaussian, but this typically leads to substantial increases in computational complexity and risks overfitting the data. For example, Cogley and Sargent (2005); Primiceri (2005); Kalli and Griffin (2018); Huber, et al. (2020); Clark, et al. (in-press); Goulet Coulombe, et al. (2022); Korobilis, et al. (2021); Karlsson, Mazur, and Nguyen (2023); Huber and Koop (2024) propose models that allow for time-variation and nonlinearities in the conditional mean, the conditional variance, or both. These methods often outperform their linear counterparts in terms of reliability of estimation and predictive accuracy. However, they are difficult to tune and the computational burden increases markedly with the size of the dataset. Similarly, to address the omitted variable problem, it is common to rely on factor models (e.g., Stock and Watson, 2002; Bernanke, Boivin, and Elias, 2005; Kaufmann and Schumacher, 2019) or large Bayesian VARs (see, e.g., Bańbura, Giannone, and Reichlin,

2010; Koop, 2013; Chan, 2022; Gefang, Koop, and Poon, 2023). However, factor models require the selection of the number of factors and call for filtering techniques to estimate the latent factors. Large VARs, on the other hand, are often specified to be linear and homoskedastic. Both assumptions are necessary to retain conjugacy and thus enable fast estimation, but could be unrealistic. And one could introduce a separate stochastic process to capture measurement errors (see, e.g., Cogley and Sargent, 2015), but doing so requires knowledge of the nature of the measurement error and is thus not easy to apply in general.

In summary, while there exist methods to handle specific types of misspecification of standard, simple, econometric models, these methods are not commonly used in practice, due to their complexity and computational costs, in addition to the fact that the type of misspecification is often unknown. Small, linear, Gaussian econometric models, such as vector autoregressions (VARs), remain the workhorse of applied macroeconomists.

There is also a literature on robust estimation of VARs and Bayesian VARs (BVARs). For instance, Schorfheide (2005) analyzes multi-step-ahead forecasting with VARs under a dynamically misspecified data generating process (DGP). More recently, González-Casasús and Schorfheide (2025) propose setting the hyperparameters of a Bayesian VAR using robust loss functions that can be tailored to the specific application at hand (i.e. whether the focus is on using the VAR for forecasting or structural inference).

Our objective in this paper is to modify BVARs to make them more robust to general and unknown forms of misspecification, without changing the simple model specification and retaining computational simplicity and efficiency. We build on a recent paper, Miller and Dunson (2018), and propose a robust version of a conjugate VAR. We call this model coarsened Bayesian VAR (cBVAR). The cBVAR replaces the exact likelihood with a coarsened likelihood that takes general and unknown forms of model misspecification into account. The key idea is that instead of conditioning on the observed data \mathbf{Y} , one conditions on the event that the difference in the sampling distribution of the observed data $P_{\mathbf{Y}}$ and of the idealized data $P_{\mathbf{Y}^*}$ (with \mathbf{Y}^* denoting the idealized data) is smaller than a constant number c , where the idealized data are such that all the model assumptions are

valid. The conditioning event effectively requires a distance metric and we use relative entropy. Miller and Dunson (2018) provide a simple approximation of the coarsened likelihood under relative entropy distance that reduces to raising the standard likelihood to a fraction, called the learning rate.¹ This approximation links the coarsening approach to the literature on power (or Gibbs) posteriors (Holmes and Walker, 2017; Grünwald and van Ommen, 2017; Bhattacharya, Pati, and Yang, 2019). The learning rate has a particularly simple interpretation. The smaller the extent of the misspecification, the closer the fraction is to one. Vice versa, for large misspecification the fraction is close to zero, putting less weight on the data.

What sets us apart from the literature on power posteriors is that these papers consider univariate regression models whereas our goal is to develop robust multivariate time series models. Using a single coarsening parameter for the full system induces the same degree of misspecification for all endogenous variables. We avoid this issue by introducing a separate coarsening parameter per equation. This implies that different equations can feature different amounts of misspecification. However, the introduction of separate coarsening parameters breaks up the convenient conjugacy of the multivariate Gaussian likelihood with the Normal-inverse Wishart prior commonly used in BVARs. As a solution, we use the asymmetric conjugate prior proposed in Chan (2022) which gives more flexibility with respect to the prior and likelihood specification.

However, this raises a serious issue. As noted in Miller and Dunson (2018), setting the learning rate appropriately is challenging. This issue is exacerbated in our framework since we have to decide on various learning rates, calling for a solution that requires little input from the researcher. We achieve this using the SafeBayes algorithm (Grünwald and van Ommen, 2017) to set the learning rate on an equation-by-equation basis. This works since the asymmetric conjugate prior of Chan (2022) avoids full system estimation by augmenting each equation with the contemporaneous values of the endogenous variables of the preceding equations while maintaining conjugacy plus order-invariance. This enables

¹Baumeister and Hamilton (2019) downweight earlier observations in their sample by raising the corresponding likelihood contributions to a fraction smaller than one.

equation-by-equation estimation of all model parameters, including the hyperparameters of the prior and the equation-specific learning rates.

To assess the empirical performance of the cBVAR, we first use simulated data, generated from a large variety of data generating processes (DGPs), characterized by different types of misspecification. For each of the DGPs, we compare density forecasts obtained from the cBVAR and from a standard BVAR. It turns out that in the presence of misspecification using coarsening improves predictive accuracy appreciably, particularly so for longer forecast horizons. When comparing differences across DGPs we find that coarsening helps particularly in the case of omitted (exogenous) variables or not considered MA errors.

Next, we consider a forecasting application to show that coarsening can often lessen the effects of misspecification for actual US data. Specifically, we focus on point and density forecasting monthly US unemployment, inflation and short-term interest rates, using three model sizes (small, medium, and large). Our findings indicate that gains from coarsening the likelihood are largest if models are small (i.e. the risk of omitted variable bias is particularly pronounced), with relative gains decreasing if the model becomes larger, but still present and systematic, in particular for longer forecast horizons. As macroeconomists tend to prefer small models for empirical analysis, this finding can be of substantial practical use. Moreover, the gains from coarsening are generally larger when the Covid period is included in the evaluation sample, in line with the increased misspecification of the standard BVAR during uncommon times.

The remainder of the paper is structured as follows. The next section provides an intuitive introduction to the coarsening idea of [Miller and Dunson \(2018\)](#) and links it to the literature on generalized posteriors. This section sets the stage for our model developments that follow in [Section 3](#) in which we derive the coarsened likelihood for the BVAR and then back out the corresponding posterior distributions. In addition, this section also includes how we set the learning rates and hyperparameters of the prior. [Section 4](#) provides evidence from synthetic data, while [Section 5](#) includes our real data forecast exercise. The final section offers a brief summary and conclusions.

2 Coarsened Posterior Distributions

This section motivates our use of the coarsened likelihood function to derive the coarsened posterior distribution in the univariate case. To set the stage, the data we observe is defined as $\mathbf{y} = (y_1, \dots, y_T)'$ which is a $T \times 1$ matrix with empirical sampling distribution $P_{\mathbf{y}} = \frac{1}{T} \sum_{t=1}^T \delta_{y_t}$. However, the unobserved idealized dataset is given by \mathbf{y}^* with associated sampling distribution $P_{\mathbf{y}^*}$ defined analogously to $P_{\mathbf{y}}$. Suppose that \mathbf{y} arises from \mathbf{y}^* through a stochastic process which is, unfortunately, unknown and let $d(P_{\mathbf{y}}, P_{\mathbf{y}^*})$ denote a distance function with $d(\cdot, \cdot) \geq 0$. We assume that $d(P_{\mathbf{y}}, P_{\mathbf{y}^*}) < c$ for some threshold parameter $c \geq 0$. Particular examples of such stochastic processes are simple measurement error models that assume that $y_t = y_t^* + \varpi$ with ϖ denoting a random measurement error with a particular distribution such as a Gaussian or multivariate student t or nonlinear models that assume that $y_t = g(y_t^*)$ for some nonlinear function g .

Standard Bayesian practice would specify a likelihood $p(\mathbf{y}|\boldsymbol{\vartheta})$, with $\boldsymbol{\vartheta}$ denoting a vector of parameters, and a prior on $\boldsymbol{\vartheta} \sim p(\boldsymbol{\vartheta})$, both of which are then used to back out the posterior distribution:

$$p(\boldsymbol{\vartheta}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\vartheta}) p(\boldsymbol{\vartheta}).$$

This procedure, however, neglects the fact that \mathbf{y} is a corrupted version of \mathbf{y}^* . An alternative would be to set up an auxiliary model $p(\mathbf{y}|\mathbf{y}^*)$. However, this is not feasible since, in applications with actual data, the process giving rise to \mathbf{y} from \mathbf{y}^* is not known (or might render the resulting model computationally involved). [Miller and Dunson \(2018\)](#) propose a simple alternative. Instead of conditioning on \mathbf{y} when forming the posterior, one could condition on $d(P_{\mathbf{y}}, P_{\mathbf{y}^*}) < C$ instead. Since C is typically unknown to the researcher one could specify a prior on it, i.e. $C \sim \pi$. Doing so leads to the coarsened posterior:

$$p(\boldsymbol{\vartheta}|\hat{d}(P_{\mathbf{y}}, P_{\mathbf{y}^*}) < C) \propto \mathbb{P}(\hat{d}(P_{\mathbf{y}}, P_{\mathbf{y}^*}) < C|\boldsymbol{\vartheta}) p(\boldsymbol{\vartheta}),$$

where the probability $\mathbb{P}(\hat{d}(P_{\mathbf{y}}, P_{\mathbf{y}^*}) < C|\boldsymbol{\vartheta})$ can be interpreted as a likelihood function.

Notice that this coarsened likelihood is generally not a probability distribution of \mathbf{y} given $\boldsymbol{\vartheta}$.

Different alternatives for the distance function d can be used. The choice of the discrepancy function can be based on the expected (or rather feared) type of misspecification. For example, as noted by Miller and Dunson (2018), robustness to outliers requires a discrepancy function that is little sensitive to movements of small amounts of probability mass to the outlying region (e.g., the first Wasserstein distance). Among the different distance functions d , one particularly attractive variant stands out, which considers differences in the entire likelihood function: the relative entropy. Moreover, using the relative entropy between $P_{\mathbf{y}}$ and $P_{\mathbf{y}^*}$ and an exponential prior on $C \sim \text{Exp}(\alpha)$ leads to a particularly simple and accurate approximation of the coarsened posterior. In this case, it can be approximated as follows (see Miller and Dunson, 2018, for a proof):

$$p(\boldsymbol{\vartheta} | \hat{d}(P_{\mathbf{y}}, P_{\mathbf{y}^*}) < C) \propto p(\boldsymbol{\vartheta}) \prod_{t=1}^T p(\mathbf{y}_t | \boldsymbol{\vartheta}, \mathbf{y}_1, \dots, \mathbf{y}_{t-1})^\phi, \quad (1)$$

with \propto denoting approximately proportional to. Equation (1) implies that the coarsened posterior is simply equal to the prior times a tempered likelihood.² This is a power posterior which has been shown to be robust to misspecification of a general form (see, e.g. Holmes and Walker, 2017; Grünwald and van Ommen, 2017; Bhattacharya, Pati, and Yang, 2019) and hence using this approximation directly links the coarsening approach to this literature.

The standard likelihood is raised by a learning rate ϕ . To understand the role of this learning rate, it is useful to note that the resulting posterior corresponds to shrinking the sample size from T to ϕT . Hence, when ϕ is small and coarsening is relevant, the sample size is substantially reduced and the posterior is much less concentrated. This is sensible, as substantial coarsening should be associated with a larger extent of misspecification, and hence uncertainty should indeed be larger. Yet, if ϕ is set to a small number when

²This particular form of the coarsened posterior resembles the intermediate approximating distribution of sequential Monte Carlo (SMC) methods. The main difference is that the learning rate in SMC grows from 0 (i.e. the likelihood plays no role) to 1 (i.e. one obtains the uncoarsened posterior). If one would fix the learning rate in SMC to ζ_T , the result would be a sequential coarsened posterior.

the model is instead approximately correct, posterior credible sets will be too large and model complexity likely under-estimated. Therefore, the choice of the parameter ϕ is very important for coarsening to be helpful and we will discuss a method to select it in a fully automatic manner later in the paper.

It is worth stressing that this form of the coarsened posterior does not require the computation of the relative entropy term since it is absorbed in the constant of proportionality (and thus independent of ϑ).

An interesting question is whether the choice of a specific prior, in particular a robust one, could make coarsening less relevant. As discussed by [Miller and Dunson \(2018\)](#), this is not the case since a robust prior makes the results less sensitive to the choice of the prior but the importance of misspecification of the likelihood remains the same, and the likelihood dominates the prior when the sample size grows (even though the coarsened posterior does not concentrate as T diverges just because of coarsening). A related issue is whether the choice of the prior matters more in the coarsened than in the standard case, since the "weight" of the likelihood decreases. Intuitively, this makes sense. Conditional on the prior, if we believe that the information contained in the likelihood is severely corrupted (either through wrongly specifying the likelihood or measurement errors, both of which fall into our definition of misspecification) a good choice would be to downweight this piece of information.

It is also worth discussing the relationship of coarsening with robust control theory, to clarify they they are quite different. In robust control theory (see, e.g., [Hansen, et al., 2006](#); [Hansen and Sargent, 2008](#)) the decision maker has one reference model but she evaluates a decision rule under a set of alternative models that are perturbed versions of the reference model. [Hansen, et al. \(2006\)](#) measure the difference between the reference and perturbed models using relative entropy, taking the maximum value of the difference as a parameter that measures the set of perturbations against which the decision maker seeks robustness (and restricts the extent of model misspecification). They also provide conditions that permit to consider the perturbed models as the multiple priors that appear in the max–min expected utility theory of [Gilboa and Schmeidler \(1989\)](#). Instead, in our

context, the prior is unique, we condition on the fact that the reference model can be at a certain (entropic) distance from the true model, and we contaminate the likelihood to take that into account when forming the posterior distribution of the model parameters.

3 Coarsened Bayesian VAR

In this section we develop the cBVAR model for the multivariate case. Inference is complicated by the fact that the degree of misspecification may vary across equations. This warrants the introduction of a distinct learning rate for each endogenous variable in the VAR. We start by discussing the coarsened likelihood of the VAR in [Subsection 3.1](#). In [Subsection 3.2](#) we sketch our prior setup and provide details on the full conditional posterior distributions. Finally we move on to discussing how to set the learning rates and other hyperparameters in [Subsection 3.3](#).

3.1 Coarsened Likelihood

Let $\mathbf{y}_t = (y_{t,1}, \dots, y_{t,M})'$ be the vector of M economic variables at time t . We consider a linear and homoskedastic VAR model of order P , with reduced form given by:

$$\mathbf{y}_t = \mathbf{a} + \mathbf{A}_1 \mathbf{y}_{t-1} + \dots + \mathbf{A}_P \mathbf{y}_{t-P} + \mathbf{u}_t, \quad \mathbf{u}_t \sim \mathcal{N}(\mathbf{0}, \Sigma), \quad (2)$$

with \mathbf{a} denoting the intercept vector, \mathbf{A}_p ($p = 1, \dots, P$) being $M \times M$ matrices of autoregressive coefficients, and \mathbf{u}_t denotes a Gaussian vector white noise process with zero mean and $M \times M$ -dimensional variance-covariance matrix Σ . The model in [Eq. \(2\)](#) contains $k = (PM + 1)M$ regression coefficients and $v = M(M + 1)/2$ free elements in the error variance-covariance matrix.

Using the Cholesky decomposition of $\Sigma = \mathbf{B}_0^{-1} \tilde{\Sigma} \mathbf{B}_0^{-1'}$, where \mathbf{B}_0 is the lower unitriangular and $\tilde{\Sigma} = \text{diag}(\sigma_1^2, \dots, \sigma_M^2)$ is a diagonal matrix, and multiplying [Eq. \(2\)](#) from

the left with \mathbf{B}_0 , we obtain the structural form of the model:

$$\mathbf{B}_0 \mathbf{y}_t = \boldsymbol{\beta} + \mathbf{B}_1 \mathbf{y}_t + \cdots + \mathbf{B}_P \mathbf{y}_{t-P} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \tilde{\boldsymbol{\Sigma}}),$$

with $\boldsymbol{\beta} = \mathbf{B}_0 \boldsymbol{\alpha}$, $\mathbf{B}_p = \mathbf{B}_0 \mathbf{A}_p$ (for $p = 1, \dots, P$) and $\boldsymbol{\varepsilon}_t = \mathbf{B}_0 \mathbf{u}_t$ being a vector of mutually independent error terms. Since $\tilde{\boldsymbol{\Sigma}}$ is diagonal, it is straightforward to derive the representation for equation i only:

$$y_{t,i} = \hat{\boldsymbol{\vartheta}}_i' \tilde{\mathbf{y}}_{t,i} + \tilde{\boldsymbol{\theta}}_i' \tilde{\mathbf{x}}_{t,i} + \varepsilon_{t,i}, \quad \varepsilon_{t,i} \sim \mathcal{N}(0, \sigma_i^2),$$

where $y_{t,i}$ and $\varepsilon_{t,i}$ denote the i -th elements of \mathbf{y}_t and $\boldsymbol{\varepsilon}_t$, respectively. For later reference, we have introduced two blocks of parameters. First, the $(i-1)$ -dimensional vector $\tilde{\mathbf{y}}_{t,i} = (-y_{t,1}, \dots, -y_{t,i-1})'$ contains contemporaneous observations of the variables ordered above variable i in \mathbf{y}_t and $\hat{\boldsymbol{\vartheta}}_i$ are the corresponding contemporaneous parameters. Note that $\tilde{\mathbf{y}}_{t,i}$ and $\hat{\boldsymbol{\vartheta}}_i$ only exist for $i > 1$. Second, $\tilde{\mathbf{x}}_{t,i} = (1, \mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-P})'$ is a $(MP+1)$ -dimensional vector that contains lagged observations and the intercept term. The corresponding regression coefficients are given by $\tilde{\boldsymbol{\theta}}_i = (\beta_i, B_{1,i,1}, \dots, B_{1,i,M}, \dots, B_{P,i,1}, \dots, B_{P,i,M})'$.

We can rewrite the regression model for equation i more concisely as:

$$y_{t,i} = \boldsymbol{\theta}_i' \mathbf{x}_{t,i} + \varepsilon_{t,i}, \quad \varepsilon_{t,i} \sim \mathcal{N}(0, \sigma_i^2),$$

with $\mathbf{x}_{t,i} = (\tilde{\mathbf{y}}_{t,i}, \tilde{\mathbf{x}}_{t,i})$ and $\boldsymbol{\theta}_i = (\hat{\boldsymbol{\vartheta}}_i, \tilde{\boldsymbol{\theta}}_i)$ being vectors of dimensions $(i+MP)$, respectively.

Stacking observations over time yields

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\theta}_i + \boldsymbol{\varepsilon}_i, \quad \boldsymbol{\varepsilon}_i \sim \mathcal{N}(\mathbf{0}, \sigma_i^2 \mathbf{I}_T),$$

where $\mathbf{Y}_i = (y_{1,i}, \dots, y_{T,i})'$ is a T -dimensional vector and $\mathbf{X}_i = (\mathbf{x}'_{1,i}, \dots, \mathbf{x}'_{T,i})'$ is a matrix of dimension $T \times (i+MP)$.

Since $\tilde{\boldsymbol{\Sigma}}$ is diagonal, the likelihood of the full system is the product of M conditionally

independent Gaussian densities:

$$p(\mathbf{Y}_1, \dots, \mathbf{Y}_M | \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_M, \sigma_1^2, \dots, \sigma_M^2) = \left(\prod_{i=1}^M p(\mathbf{Y}_i | \boldsymbol{\theta}_i, \sigma_i^2) \right).$$

We use this form of the likelihood because it effectively allows us to treat each equation independently and hence introduce separate coarsening parameters per series. This is predicated on the fact that certain series in \mathbf{y}_t might be more prone to misspecification than other series. Hence, using a single coarsening parameter on the joint likelihood implies a trade-off; we can either control for substantial degrees of misspecification or little/no misspecification. Therefore, we set up the coarsened likelihood for each equation of the VAR and introduce separate coarsening parameters ϕ_i :

$$\tilde{p}(\mathbf{Y}_i | \boldsymbol{\theta}_i, \sigma_i^2, \phi_i) = \left(\prod_{t=1}^T p(y_{t,i} | \boldsymbol{\theta}_i, \sigma_i^2)^{\phi_i} \right).$$

Rewriting this equation yields the equation-specific coarsened likelihood:

$$\tilde{p}(\mathbf{Y}_i | \boldsymbol{\theta}_i, \sigma_i^2, \phi_i) = (2\pi\sigma_i^2)^{-\frac{\phi_i T}{2}} \exp \left(-\frac{\phi_i}{2\sigma_i^2} \sum_{t=1}^T (y_{t,i} - \boldsymbol{\theta}_i' \mathbf{x}_{t,i})^2 \right).$$

The joint coarsened likelihood is then given by:

$$\tilde{p}(\mathbf{Y}_1, \dots, \mathbf{Y}_M | \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_M, \sigma_1^2, \dots, \sigma_M^2, \phi_1, \dots, \phi_M) = \prod_{i=1}^M \tilde{p}(\mathbf{Y}_i | \boldsymbol{\theta}_i, \sigma_i^2, \phi_i). \quad (3)$$

For large values of M and P , the number of parameters can quickly exceed the number of available observations, and the OLS estimator ceases to exist. Hence, regularization is necessary, and Bayesian approaches that rely on specifying priors on $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_M$ and $\hat{\sigma}_1^2, \dots, \hat{\sigma}_M^2$ are commonly employed. In the next sub-section we will derive the coarsened posterior under the asymmetric conjugate prior.

3.2 Coarsened Bayesian Analysis of the VAR

The natural conjugate prior (see, e.g., Kadiyala and Karlsson, 1997; Koop, 2013; Carriero, Clark, and Marcellino, 2015) implies that the amount of shrinkage is proportional across equations

(scaled by the corresponding elements of the covariance matrix Σ). Since we start from the assumption that the amount of misspecification can differ across equations, this assumption is overly restrictive. As a solution, we use the asymmetric conjugate prior proposed in Chan (2022). This prior has the convenient property that it leads to analytical posterior results and can be combined with the coarsened likelihood in Eq. (3). Moreover, it leads to a model that is order-invariant since the prior on the contemporaneous terms translates into an inverted Wishart prior on Σ .

We assume that the priors for each equation $i = 1, \dots, M$ are mutually independent such that $p(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_M, \sigma_1^2, \dots, \sigma_M^2) = \prod_i^M p(\boldsymbol{\theta}_i, \sigma_i^2)$. The prior on the regression coefficients is normally distributed and conditions on σ_i^2 :

$$(\boldsymbol{\theta}_i | \sigma_i^2) \sim \mathcal{N}(\underline{\boldsymbol{\theta}}_i, \sigma_i^2 \underline{\mathbf{V}}_i)$$

with $\underline{\boldsymbol{\theta}}_i$ denoting the prior mean and $\underline{\mathbf{V}}_i$ a prior variance-covariance matrix of dimension $K_i \times K_i$. We follow the Sims and Zha (1998) tradition and assume that, a priori, the elements in \mathbf{y}_t follow M independent random walks. This is achieved by centering \mathbf{B}_1 on \mathbf{I}_M (or to $\mathbf{0}_M$ if the data is stationary) and \mathbf{B}_p ($p \neq 1$) on $\mathbf{0}_{M \times M}$. Let $\underline{\boldsymbol{\theta}}_i = (\underline{\tilde{\boldsymbol{\vartheta}}}_i, \underline{\tilde{\boldsymbol{\theta}}}_i)$, where $\underline{\tilde{\boldsymbol{\vartheta}}}$ and $\underline{\tilde{\boldsymbol{\theta}}}_i$ denote the prior mean for the block of contemporaneous and dynamic coefficients, respectively. For equation $i = 1, \dots, M$, we set the prior mean as $\underline{\tilde{\boldsymbol{\theta}}}_i = (\mathbf{0}'_{(i-1) \times 1}, 1, \mathbf{0}'_{(M-i) \times 1})'$. The prior on the contemporaneous coefficients is centered on zero: $\underline{\tilde{\boldsymbol{\vartheta}}} = \mathbf{0}_{i-1 \times 1}$.

We assume that the prior variance-covariance matrix $\underline{\mathbf{V}}_i = \text{diag}(\underline{\mathbf{v}}_{\tilde{\boldsymbol{\vartheta}},i}^2, \underline{\mathbf{v}}_{\tilde{\boldsymbol{\theta}},i}^2)$ is diagonal. In this paper, we follow Chan (2022) and set $\underline{\mathbf{v}}_{\tilde{\boldsymbol{\vartheta}},i}^2 = (1/s_1^2, \dots, 1/s_{i-1}^2)$, where s_i^2 denotes the OLS estimate of the residual error variance of an $AR(P)$ model for variable i . The elements of $\underline{\mathbf{v}}_{\tilde{\boldsymbol{\theta}},i}^2$ are set using the Minnesota prior, which shrinks coefficients associated to higher-order lags of \mathbf{y}_t and makes a distinction between the own and other lags of a particular target variable. In particular, the k^{th} element of $\underline{\mathbf{v}}_{\tilde{\boldsymbol{\theta}},i}^2$ reads:

$$\underline{v}_{\tilde{\boldsymbol{\theta}},i,k}^2 = \begin{cases} \frac{\kappa_{1,i}}{p^2 s_i^2}, & \text{for the coefficient of the } p\text{-th lag of variable } i \\ \frac{\kappa_{2,i}}{p^2 s_j^2}, & \text{for the coefficient of the } p\text{-th lag of variable } j \neq i \\ \kappa_3, & \text{for the regression intercept} \end{cases}$$

where, for each equation i , $\kappa_{1,i}$ controls the amount of shrinkage induced on the coefficients associated with lags of variable i and $\kappa_{2,i}$ is the shrinkage hyperparameter for coefficients associated with lags of other variables. More information on how we set $\kappa_{1,i}$ and $\kappa_{2,i}$ is provided in Subsection 3.3. The hyperparameter κ_3 is set equal to 10^2 to effectively introduce no prior information on the intercept term.

On the error variances we use an inverse Gamma prior:

$$\sigma_i^2 \sim \mathcal{IG}(\underline{\nu}_i, \underline{S}_i),$$

where ν_i and S_i denote the scale and shape parameter, respectively. We set $\nu_i = 1 + i/2$ and $\underline{s}_i = s_i^2/2$. Chan (2022) shows that this choice, in combination with the prior on $\hat{\vartheta}_i$, leads to a prior on the reduced-form error variance-covariance matrix which follows an inverse Wishart distribution $\Sigma \sim \mathcal{IW}(M+2, \underline{S})$ with $M+2$ degrees of freedom and prior scaling matrix $\underline{S} = \text{diag}(s_1^2, \dots, s_M^2)/(M+2)$.

At this point, it is worth stressing that our coarsened likelihood can be combined with any of the priors commonly used in the VAR literature. For instance, priors based on the use of structural models to inform parameter estimates (see, e.g., Ingram and Whiteman, 1994; Del Negro and Schorfheide, 2004; De Luigi and Huber, 2018; Loria, Matthes, and Wang, 2022), priors utilizing information on the long-term behavior of the time series under scrutiny (Giannone, Lenza, and Primiceri, 2019) or priors that force the VAR towards factor models (Huber and Koop, 2023) can be easily incorporated in our framework.

The joint prior $\prod_{i=1}^M p(\boldsymbol{\theta}_i, \sigma_i^2) = \prod_{i=1}^M p(\boldsymbol{\theta}_i | \sigma_i^2) \times p(\sigma_i^2)$ can be combined with Eq. (3) to obtain the joint coarsened posterior distribution:

$$\begin{aligned} & p(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_M, \sigma_1^2, \dots, \sigma_M^2 | \mathbf{Y}_1, \dots, \mathbf{Y}_M, \phi_1, \dots, \phi_M) \\ &= \prod_{i=1}^M c_i (2\pi)^{-\frac{\phi_i T}{2}} (\sigma_i)^{-(\bar{\nu}_i + \frac{i+MP}{2} + 1)} \exp\left(-\frac{1}{\sigma_i^2} (\bar{S}_i + \frac{1}{2}(\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}}_i)' \bar{\mathbf{V}}_i (\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}}_i))\right), \end{aligned}$$

which is the product of M Gaussian-Inverse-Gamma distributions:

$$(\boldsymbol{\theta}_i, \sigma_i^2 | \mathbf{Y}) \sim \mathcal{NIG}(\bar{\boldsymbol{\theta}}_i, \bar{\mathbf{V}}_i, \bar{\nu}_i, \bar{S}_i). \quad (4)$$

We provide an analytical expression for the normalizing constant c_i of the posterior in [Subsection 3.3](#). The posterior hyperparameters are given by:

$$\begin{aligned}\bar{\mathbf{V}}_i &= (\underline{\mathbf{V}}_i^{-1} + \phi_i \mathbf{X}_i' \mathbf{X}_i)^{-1}, \\ \bar{\boldsymbol{\theta}}_i &= \bar{\mathbf{V}}_i (\underline{\mathbf{V}}_i^{-1} \boldsymbol{\theta}_i + \phi_i \mathbf{X}_i' \mathbf{Y}_i) \\ \bar{\nu}_i &= \underline{\nu}_i + \frac{\phi_i T}{2}, \\ \bar{S}_i &= \underline{S}_i + (\phi_i \mathbf{Y}_i' \mathbf{Y}_i + \underline{\boldsymbol{\theta}}_i' \underline{\mathbf{V}}_i \underline{\boldsymbol{\theta}}_i - \bar{\boldsymbol{\theta}}_i' \bar{\mathbf{V}}_i \bar{\boldsymbol{\theta}}_i) / 2.\end{aligned}$$

As discussed in [Chan \(2022\)](#), it is straightforward to sample from this distribution by first sampling $\sigma_i^2 \sim \mathcal{IG}(\bar{\nu}_i, \bar{S}_i)$ marginally and then drawing $\boldsymbol{\theta}_i | \sigma_i^2 \sim \mathcal{N}(\bar{\boldsymbol{\theta}}_i, \sigma_i^2 \bar{\mathbf{V}}_i)$ conditionally on the current draw of σ_i^2 .

3.3 Deciding on the degree of coarsening and setting the prior hyperparameters

For each equation in the VAR we need to decide on the learning rate as well as the hyperparameters controlling the prior tightness. There are several ways to choose ϕ_i . First, one can use prior information on the amount of misspecification. However, for possibly large-dimensional macroeconomic datasets, this turns out to be unfeasible. Second, we could use cross-validation. This essentially implies that one recomputes the model to obtain quantities of interest (such as forecasts or impulse responses) for different values of ϕ_i and picks the one that minimizes a loss function. Third, from a Bayesian perspective one could integrate out ϕ_i . [Grünwald and van Ommen \(2017\)](#) show, however, that this strategy does not work, since without substantial prior information on ϕ_i the model severely underestimates the true degree of misspecification and we would end up setting ϕ_i close to one.

In this paper, we use the SafeBayes algorithm of [Grünwald and van Ommen \(2017\)](#) that is fully automatic and has excellent theoretical and empirical properties. Since our model is written in its recursive form, the SafeBayes algorithm proceeds on an equation-by-equation basis.

Before starting the algorithm, we need to define a grid of length R of possible learning rates $\boldsymbol{\Phi}_i$. In our empirical work, $\boldsymbol{\Phi}_i = (0.05, 0.1, \dots, 1)'$ for all i . Then, conditional on a particular value $\phi_i^{(j)} \in \boldsymbol{\Phi}_i$, we first decide on the optimal shrinkage parameters of the Minnesota prior.

Algorithm 1 SafeBayes (Grünwald and van Ommen, 2017)

Input: Data \mathbf{Y} , learning rate proposals Φ

Output: learning rates ϕ_1, \dots, ϕ_M , prior hyperparameters $\kappa_{1,1}, \kappa_{2,1}, \dots, \kappa_{1,M}, \kappa_{2,M}$

for i in $1, \dots, M$ **do**

for all $\phi \in \Phi$ **do**

 Select $\kappa_{1,i,\phi}^*, \kappa_{2,i,\phi}^* = \arg \max_{\kappa_{1,i}, \kappa_{2,i}} \tilde{p}(Y_i | \kappa_{1,i}, \kappa_{2,i}, \phi_i)$

for t in $1, \dots, T-1$ **do**

 Obtain coarsened posterior expectation conditional on data up to time t :

$\mathbb{E}_t(\boldsymbol{\theta}_i, \sigma_i^2 | \mathbf{Y}_{i,1:t}, \phi, \kappa_{1,i}^*, \kappa_{2,i}^*)$

 Calculate loss function by predicting actual next outcome:

$l_{i,\phi,t} = -\log f(y_{i,t+1} | \mathbf{x}_{i,t+1}, \mathbb{E}_t(\boldsymbol{\theta}_i, \sigma_i^2 | \mathbf{Y}_{i,1:t}, \phi, \kappa_{1,i}^*, \kappa_{2,i}^*))$

end for

end for

 Choose $\phi_i := \arg \min_{\phi \in \Phi} \sum_{t=1}^{T-1} l_{i,\phi,t}$ and corresponding $\kappa_{1,i,\phi_i}^*, \kappa_{2,i,\phi_i}^*$

end for

This is done by maximizing the coarsened marginal likelihood (ML) with respect to $\kappa_{1,i}, \kappa_{2,i}$ while conditioning on $\phi_i^{(j)}$:

$$\kappa_{1,i}^{(j)}, \kappa_{2,i}^{(j)} = \arg \max_{\kappa_{1,i}, \kappa_{2,i}} \tilde{p}(\mathbf{Y}_i | \phi_i = \phi_i^{(j)}, \kappa_{1,i}, \kappa_{2,i}),$$

with $\tilde{p}(\mathbf{Y}_i | \phi_i = \phi_i^{(j)}, \kappa_{1,i}, \kappa_{2,i})$ denoting the coarsened ML specific to equation i . This is given by:

$$\tilde{p}_i(Y_i | \phi_i, \kappa_{1,i}, \kappa_{2,i}) = (2\pi)^{\frac{\phi_i T}{2}} |\mathbf{V}_i|^{-\frac{1}{2}} |\bar{\mathbf{V}}_i|^{\frac{1}{2}} \frac{\Gamma(\bar{\nu}_i) \underline{s}_i^{\bar{\nu}_i}}{\Gamma(\underline{\nu}_i) \bar{s}_i^{\bar{\nu}_i}}.$$

Here, we let $\Gamma(\bullet)$ denote the Gamma function and $|\mathbf{A}|$ denotes the determinant of a matrix \mathbf{A} . It should be noted that we optimize the coarsened ML for each value of ϕ_i , implying that the hyperparameters are set as a function of the degree of misspecification. This is related to González-Casasús and Schorfheide (2025) who use misspecification robust loss functions instead of the marginal likelihood to select the hyperparameters of a Bayesian VAR. Contrary to our approach, this is done on a full-system basis.

Conditional on $\phi_i^{(j)}$ and the optimal shrinkage parameters $\kappa_{1,i}^{(j)}, \kappa_{2,i}^{(j)}$, we then loop through $t = 1, \dots, T-1$ and derive the coarsened posterior expectations for $\mathbb{E}_t[\boldsymbol{\theta}_i]$ and $\mathbb{E}_t[\sigma_i^2]$ using only data up to time t . We use these to evaluate the loss function $l_{i,\phi,t} = -\log f(y_{i,t+1} | x_{i,t+1}, \mathbb{E}_t[\boldsymbol{\theta}_i], \mathbb{E}_t[\sigma_i^2])$ at time t . In our empirical work, the loss function equals the log predictive likelihood (LPL), a

commonly used forecast evaluation metric. This implies setting $f(\cdot) = \mathcal{N}(y_{i,t+1} | \mathbb{E}_t[\boldsymbol{\theta}_i]' \mathbf{x}_{i,t+1}, \mathbb{E}_t[\sigma_i^2])$. After performing this operation for all t , we choose the learning rate out of Φ that minimizes the sum of loss functions over the entire sample:

$$\phi_i^* = \arg \min_{\phi \in \Phi} \sum_{t=1}^{T-1} l_{i,\phi,t},$$

and choose the associated prior hyperparameters $\kappa_{i,1} = \kappa_{i,1}^*$ and $\kappa_{i,2} = \kappa_{i,2}^*$. This choice of the loss function is also motivated in Grünwald and van Ommen (2017).

This is repeated for each equation until we end up with M optimal learning rates and associated prior hyperparameters. Notice that the conjugate nature of our model makes this step computationally tractable. If posterior quantities such as the predictive mean, the predictive variance or the marginal likelihood would not be available in closed form, one would have to use simulation-based methods for each value of $\phi_i^{(j)}$ in our grid. Algorithm 1 gives an overview of the algorithm.

4 Evidence based on synthetic data

In this section, we carry out a structured evaluation of the empirical performance of our coarsened BVAR model using a detailed Monte Carlo exercise. First, we employ a straightforward illustrative example grounded in a DGP with non-Gaussian errors. Subsequently, we examine the predictive capacity of our method in various DGPs.

4.1 An illustrative example

To investigate whether the estimated coarsening parameter effectively picks up model misspecification, we simulate $T = 500$ observations from a small linear VAR where the errors are Student's t distributed with varying degrees of freedom ν .³ In particular, we decrease the degrees of freedom from 20 (in which case the shape of the Student-t distribution generates a moderate amount of outliers such that the standard BVAR is only lightly misspecified) to 3 (the lowest value for which the moments of the Student-t distribution exist; implying a maximum amount of misspecification). This leads to 18 different DGPs, from each of which we take 100 draws to estimate

³Information on the precise DGP used can be found in Subsection A.2 of the Online Appendix.

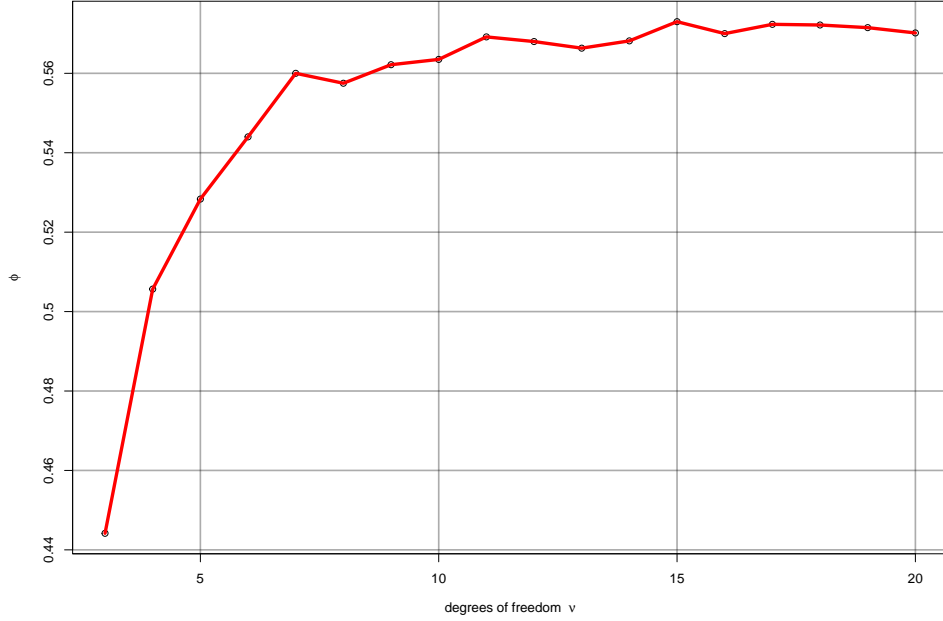


Figure 1: Estimated Coarsening parameter vs. degrees of freedom of the Student-t distribution generating the model errors.

Notes: Estimates are the mean over 100 draws from each DGP and over all three VAR equations.

Legend: — represents a smoothed spline.

the cBVAR.

Figure 1 shows the mean estimate (averaged over all realizations from each DGP and all VAR equations) for each of these DGPs. Starting with the extreme case of three degrees of freedom, our model selects an average learning rate of around 0.44, controlling for the fact that the shocks are heavy-tailed and the misspecified model we estimate has much lighter tails. This learning rate results in a significantly wider predictive distribution, suggesting that the cBVAR model is more capable of accommodating outliers compared to the conventional BVAR model.

Considering larger degrees of freedom reveals that the learning rate increases quickly (to around 0.5) before reaching a level of around 0.57 for larger degrees of freedom. This is surprising given that, in this case, the standard BVAR is actually well-specified. However, notice that the number of observations is large (given the size of the model), and outliers are still more likely under the t distribution with moderate degrees of freedom. In this case, the SafeBayes algorithm would set the learning rate below one. This is driven by the loss function that depends on the log predictive score.

4.2 Simulation results based on a large universe of data generating processes

4.2.1 Design of the Monte Carlo exercise

We evaluate the performance of our coarsened BVAR against that of a standard Bayesian VAR without likelihood tempering across a range of different data generating processes (DGPs). We vary these DGPs across two dimensions. Firstly, we allow for different conditional mean specifications. We consider a standard linear VAR (labeled "Linear"), a VAR with structural breaks in the parameters ("Breaks"), a VAR augmented with exogenous variables ("Exo") and a smooth-transition VAR ("Transition"). Secondly, for the error term we consider standard, Gaussian errors ("Gaussian"), errors arising from a Student-t distribution ("Student"), heteroskedastic errors whose variance arises from a stochastic volatility process ("SV") and errors featuring a moving average component ("MA"). The latter specification allows us to test whether coarsening helps for dynamically misspecified models along the lines of [Schorfheide \(2005\)](#) and [González-Casasús and Schorfheide \(2025\)](#). Details on the specifications of these DGPs can be found in [Subsection A.2](#) of the Online Appendix.

Moreover, to investigate the relationship between model performance and size, we consider three different model sizes. The first is a small VAR with three variables ("small"), the second is a medium-sized VAR containing nine variables ("medium") and the third is a large 27-variable VAR ("large").

These combinations of conditional mean and variance specifications reflect a large share of models commonly applied by empirical macroeconomists. Practitioners are also often interested in estimating larger models, motivating our wish to understand how coarsening performs for models of different sizes. Each of the DGPs is then calibrated by estimating the respective model parameters on US macro data taken from the FRED-MD data base ([McCracken and Ng, 2016](#)). We estimate all models with $P = 2$ lags.

From each DGP, we draw 100 different realizations of \mathbf{y}_t of size $T = 200$ to estimate the coarsened and standard BVAR. For each of these runs, we then use the DGP to simulate 1,000 out-of-sample datasets with $H = 12$ observations each. This yields a total of $100 \cdot 1,000 = 100,000$ evaluation points at each forecast horizon. For each evaluation point in the holdout, we compute the LPL of a particular cBVAR vis-à-vis the BVAR (in differences such that numbers

above zero mean better performance of the cBVAR).

4.2.2 Predictive results

Table 1 displays the mean differences in LPLs averaged across all test sets and all draws from each DGP.⁴ Table 1 is divided into three blocks containing results for the small, medium and large models, respectively. Each of these, in turn, consists of three blocks for the different forecast horizons. We also report the results of a Diebold and Mariano (1995) test with the null hypothesis of equal forecast performance.

In general, the coarsened VAR exhibits a superior performance relative to the traditional Bayesian VAR. We observe a clear and steady pattern: the benefits of using the coarsened likelihood as a substitute for the standard version grow in parallel with the forecast horizon. As VAR predictive distributions are generated iteratively, errors due to model misspecification tend to accumulate and intensify the more extended the forecast period is (see, e.g., Marcellino, Stock, and Watson, 2006).

Another salient observation is that the forecast gains increase with model size. As outlined above, the misspecified standard VAR fits a model based on data that did not arise from the postulated DGP. Since larger models tend to fit the data more strongly, the negative implications of misspecification convolute and have a deleterious impact on forecasting performance that increases with model size.

When we focus on specific selections for model size, conditional mean, and variance, some noteworthy patterns emerge. Firstly, in the case of correct model specification (Linear + Gaussian), it appears that coarsening only adversely affects predictive accuracy when the model size is small. However, these effects are minimal and the predictive distributions remain similar to those of the standard BVAR. Conversely, if the model continues to be correctly specified but increases in size, coarsening can enhance predictive performance. This unexpected outcome aligns with the simulation findings of Grünwald and van Ommen (2017), who, while examining Bayesian ridge regressions, discover that SafeBayes significantly reduces risk even in the correctly specified case.

For combinations of the conditional mean and variance that imply a misspecified model, we find substantial gains, but only for forecast horizons greater than one-step-ahead. The largest

⁴Results for point forecasts are provided in the Online Appendix.

Table 1: Differences in joint LPL scores for predictive distributions.

		Gaussian		Student		SV		MA		
		Δ LPLS		Δ LPLS		Δ LPLS		Δ LPLS		
Small-sized	<i>One-month-ahead predictions</i>									
	Linear	-0.01		-0.02		-0.02		-0.01		
	Break	0.03	***	0.02	***	0.03	***	0.04	***	
	Exo	0.03	***	-0.04		0.01	***	0.01	***	
	Transition	-0.03		-0.05		-0.03		-0.02		
	<i>One-quarter-ahead predictions</i>									
	Linear	0.02	***	0.11	***	0.04	***	0.43	***	
	Break	0.38	***	0.54	***	0.39	***	0.92	***	
	Exo	0.91	***	0.15		0.89	***	0.31	***	
	Transition	0.10	***	0.01		0.12	***	0.46	***	
	<i>One-year-ahead predictions</i>									
	Linear	-0.46		-0.03		-0.58		6.50	***	
Break	2.67	***	2.66	***	2.16	***	9.61	***		
Exo	29.94	***	10.08	***	42.86	***	7.84	***		
Transition	1.32	***	-1.28		1.41	***	6.44	***		
Medium-sized	<i>One-month-ahead predictions</i>									
	Linear	0.05	***	0.06	***	0.06	***	-0.13		
	Break	0.63	***	0.89	***	0.74	***	0.58	***	
	Exo	0.79	***	0.65	***	1.01	***	1.24	***	
	Transition	0.30	***	0.18	***	0.25	***	0.40	***	
	<i>One-quarter-ahead predictions</i>									
	Linear	1.47	***	1.80	***	1.54	***	4.65	***	
	Break	5.13	***	7.02	***	6.25	***	5.35	***	
	Exo	24.95	***	17.04	***	23.81	***	30.37	***	
	Transition	2.30	***	1.42	***	2.00	***	3.10	***	
	<i>One-year-ahead predictions</i>									
	Linear	6.69	***	5.74	***	6.15	***	28.52	***	
Break	20.46	***	19.39	***	22.95	***	22.22	***		
Exo	955.96	***	425.01	***	772.48	***	1030.49	***		
Transition	14.12	***	0.72		12.78	***	20.77	***		
Large-sized	<i>One-month-ahead predictions</i>									
	Linear	3.91	***	4.94	***	3.94	***	5.76	***	
	Break	5.96	***	7.23	***	6.56	***	8.98	***	
	Exo	21.30	***	12.98	***	17.42	***	31.54	***	
	Transition	0.53	***	0.96	***	0.77	***	0.76	***	
	<i>One-quarter-ahead predictions</i>									
	Linear	22.55	***	26.31	***	23.83	***	42.96	***	
	Break	16.98	***	20.25	***	18.59	***	23.89	***	
	Exo	103.53	***	66.36	***	92.34	***	157.14	***	
	Transition	4.49	***	5.54	***	5.10	***	8.69	***	
	<i>One-year-ahead predictions</i>									
	Linear	69.99	***	85.33	***	74.48	***	121.29	***	
Break	52.76	***	58.37	***	54.51	***	71.30	***		
Exo	495.35	***	284.06	***	471.56	***	713.78	***		
Transition	15.80	***	17.71	***	17.53	***	26.73	***		

Notes: Differences are averaged over all draws and test sets for each DGP. Stars indicate p-values from a one-sided t-test with alternative hypothesis that cBVAR produces larger LPL scores.

Legend: . $\sim p < 0.16$; * $\sim p < 0.1$; ** $\sim p < 0.05$; *** $\sim p < 0.01$

gains can be found in the case of omitted variables ("Exo"). In this case, the cBVAR produces much more precise density forecasts than the benchmark specification for forecast horizons greater than one-step-ahead. Interestingly, these gains do not always increase if we allow for misspecification in the error terms: the largest gains can be found if we ignore a MA structure in the shocks while the second-largest gains are observed if the shocks are Gaussian. This pattern is consistent across model sizes.

For the other DGPs which imply non-linearities in the conditional mean, we also observe gains from coarsening. However, these are less pronounced than in the case of (neglecting) exogenous factors. Interestingly, we find a different pattern if we allow for distributional misspecification. If we consider structural breaks or smooth transition DGPs, we find that solid gains can already be obtained for the one-step-ahead forecast if the errors feature stochastic volatility or a MA structure. For larger horizons (and model sizes), a pattern similar to the DGP featuring exogenous covariates shows up. In most cases, we find the largest improvements in forecast accuracy if the DGP has MA shocks or homoskedastic Gaussian disturbances.

In conclusion, we find strong evidence in favor of our proposed approach. In the vast majority of cases the cBVAR significantly outperforms its non-coarsened competitor. The only case where the results are somewhat mixed are one-step-ahead LPLs arising from the small model. For larger model sizes, improvements increase further and tend to be present for short-term forecasts as well.

5 Macroeconomic forecasting with coarsened VARs

5.1 Data, model specification and design of the forecasting exercise

We employ our cBVAR to forecast a range of US macroeconomic variables. Again, we focus on three model sizes. First, we consider a small VAR in three variables. These three variables are the unemployment rate (UNRATE), CPI inflation (CPIAUCSL) and the short-term interest rate (FEDFUNDS). Second, the three series of the small model are complemented with an additional six series. These are average weekly working hours in manufacturing (AWHMAN), the real M2 money stock (M2REAL), the S&P500 stock price index (S&P 500), the industrial

production index (INDPRO), the spread between the federal funds rate and 10-year treasury rate (T10YFFM) as well as the consumer price index for commodities (CUSR0000SAC). Finally, we consider a large model that builds on the medium-scale model and adds additional 18 series such that the large model comprises 27 variables in total. The data is sourced from the FRED-MD database (McCracken and Ng, 2016) and transformed to stationarity using the transformation codes suggested in McCracken and Ng (2016). Precise information on the series and transformations is provided in Table B.1 of the Online Appendix. All models we consider include $P = 2$ lags.

Model comparison is done using LPLs and mean squared forecast errors (MSFEs). This is done in two ways. We consider two types of LPLs. First, we focus on univariate LPLs for three target series (the unemployment rate, UNRATE; CPI inflation, CPIAUCSL; and short-term interest rates, FEDFUNDS). This gives us information on how well coarsening helps when predicting each of the series individually. Second, we consider the joint LPL over all series in \mathbf{y}_t . This gives an impression on the overall predictive performance of a model. All loss functions are computed for the one-month-ahead, one-quarter-ahead and one-year-ahead forecast horizon.

The forecast design is as follows. Our sample starts in 1967:M07 and ends in 2023:12. We use 1967:M07 to 1999:M12 as our initial training sample. After estimating the models, we produce the predictive distributions for 2000:M01 up to 2000:M12. We then add one additional observation to the initial estimation window and compute the corresponding predictive distributions (2000:M02 to 2001:M01). This procedure is repeated until we arrive at the end of the sample.

5.2 Forecasting results

5.2.1 Point forecasting accuracy

We start our discussion with the point forecasting performance for the three focus variables. As a point forecast, we use the posterior median of the respective predictive distributions. The MSFEs for the cBVAR, in ratios relative to the BVAR, are shown in Table 2. Numbers greater than one imply that the cBVAR produces less precise point forecasts, whereas numbers smaller than one indicate the opposite. Raw MSFEs for the BVARs are shown in the columns labeled 'BVAR'.

Table 2: Mean squared error over the entire hold-out period.

Variables	Small		Medium		Large				
	cBVAR	BVAR	cBVAR	BVAR	cBVAR	BVAR			
<i>One-month-ahead predictions</i>									
UNRATE	0.86	0.12	0.88	.	0.12	0.93	0.1		
CPIAUCSL	1	0.02	1.02		0.02	1.05	0.02		
FEDFUNDS	0.33	.	0.01	0.71	**	0	0.65	***	0.01
<i>One-quarter-ahead predictions</i>									
UNRATE	0.8	.	0.43	0.79		0.42	0.85	.	0.34
CPIAUCSL	0.99		0.13	0.98		0.13	0.97		0.13
FEDFUNDS	0.41		0.04	0.84	.	0.04	0.69	***	0.05
<i>One-year-ahead predictions</i>									
UNRATE	0.93	*	1.07	0.91		0.98	0.85	.	1.06
CPIAUCSL	1.05		0.72	0.95		0.75	1.2		0.63
FEDFUNDS	0.91		0.17	0.89	*	0.23	0.78		0.38

Notes: Results for the cBVAR are the ratio relative to the BVAR. Stars indicate p-values from a one-sided DM-test with alternative hypothesis that cBVAR produces smaller forecast errors.

Legend: . $\sim p < 0.16$; * $\sim p < 0.1$; ** $\sim p < 0.05$; *** $\sim p < 0.01$

When assessing point forecasts, we find that coarsening leads to more accurate point forecasts for all variables except inflation. For the unemployment rate, we find improvements of around 14 percentage points (PPs) at the one-month-ahead horizon. For one-quarter-ahead, these improvements reach about 20 PPs before deteriorating appreciably when one-year-ahead predictions and the small dataset are being considered. These gains remain similar when we consider the medium-sized dataset (with very small improvements for one-quarter- and one-year-ahead forecasts). When we consider the large model, we find somewhat weaker gains from coarsening for forecasts up to one-quarter-ahead. However, for longer-run predictions (one year in advance), we observe improvements of around 15 PPs. These somewhat smaller improvements for short-run unemployment rate forecasts are most likely driven by the fact that larger models mitigate misspecification issues by trading unobserved heterogeneity with observed heterogeneity (through the inclusion of many more series that soak up patterns a more flexible econometric model would otherwise pick up).

Turning to inflation forecasts gives rise to a mixed picture. For one-quarter-ahead predictions (and irrespective of the model size), we find MSFE ratios close to (but mostly above) one. Increasing the forecast horizon to one-quarter-ahead slightly changes this picture, suggest-

ing small improvements in forecast performance for the medium and large models. But these improvements are minor and reach around three PPs (for the large model). When we consider one-year-ahead forecasts we find that coarsening hurts point forecast accuracy for small and large models, with pronounced losses of around 20 PPs for the biggest model we consider. Only for the medium model, we observe gains of around 5 PPs.

The time series that profits most from coarsening is the Federal Funds rate. In this case, improvements for the critical one-month-ahead horizon reach staggering 67 PPs for the small model and still sizable improvements of 29 to 35 PPs for the medium and large models, respectively. These strong gains are driven by the zero lower bound (ZLB). Econometrically, the ZLB is a non-linear feature in the FEDFUNDS time series that should be controlled for through, e.g., regime-switching models (Liu, et al., 2019) or models that allow for estimating the shadow interest rate (Carriero, et al., 2025). The BVAR ignores this, leading to misspecification. This results in predictions that display substantial amounts of high-frequency variation even if the short-term interest rate is stuck close to zero. By contrast, the coarsened model produces a predictive median that displays much less high-frequency variation (see Figure C.1g to Figure C.3g of the Online Appendix), and this leads to substantial improvements in MSFEs. Once we increase the forecast horizon, these gains become slightly smaller, but even for the one-year-ahead horizon, remain sizable.

5.2.2 Density forecasting accuracy

We now turn to discussing whether coarsening also helps for higher-order moments by considering LPL differences to the uncoarsened VAR. Table 3 shows average LPL differences between the cBVAR and the BVAR as well as absolute LPLs for the BVAR benchmark. Numbers exceeding zero indicate outperformance of the cBVAR while negative numbers indicate the opposite. Unlike Table 2, we also have rows that show the performance of the joint density forecast.

In terms of density forecasts, we find an overall picture that is closely related to the one observed for point forecasting performance. Starting with the joint density forecast performance, we find improvements relative to the BVAR for all three model sizes and forecast horizons. Notice that the general pattern that performance improvements increase with the predictive horizon is visible for the medium- and large-scale models. For the small model, we find strong improvements for one-quarter-ahead LPLs and less pronounced gains for one-month- and one-

year-ahead LPLs.

Table 3: Average LPL scores over the entire hold-out period.

Variables	Small		Medium		Large	
	cBVAR	BVAR	cBVAR	BVAR	cBVAR	BVAR
<i>One-month-ahead predictions</i>						
Joint	0.28 *	-2.99	0.26 *	-3.69	0.57 **	-3.88
UNRATE	0.01 ***	0.42	-0.01 ***	0.44	-0.01 ***	0.44
CPIAUCSL	0.11 .	0.90	0.00 ***	1.01	0.02 ***	1.01
FEDFUNDS	0.39 *	-1.77	0.42 ***	-1.93	2.64 ***	-16.03
<i>One-quarter-ahead predictions</i>						
Joint	1.26 .	-13.57	1.47 .	-15.71	3.75 *	-18.95
UNRATE	0.13 ***	-3.32	0.27 ***	-3.42	0.80 ***	-4.11
CPIAUCSL	0.72 .	-0.07	0.26 ***	-0.11	0.66 ***	-0.65
FEDFUNDS	1.86 *	-16.73	2.89 **	-37.45	16.81 ***	-141.52
<i>One-year-ahead predictions</i>						
Joint	1.60 *	-39.09	2.21 *	-41.11	15.12 **	-61.13
UNRATE	-0.74 ***	-23.32	2.36 ***	-25.66	-0.59 **	-24.96
CPIAUCSL	0.65 *	-3.63	1.46 ***	-6.14	4.29 **	-12.02
FEDFUNDS	0.85 **	-63.07	11.17 ***	-175.01	58.64 ***	-607.74

Note: Results for the cBVAR are in differences relative to the BVAR. Stars indicate p-values from a one-sided DM-test with alternative hypothesis that cBVAR produces larger LPL scores.

Legend: . $\sim p < 0.16$; * $\sim p < 0.1$; ** $\sim p < 0.05$; *** $\sim p < 0.01$

When we hone in on the variable-specific performance a story similar to the one we have seen in Table 2 emerges. For unemployment rate forecasts, we find muted gains for one-month-ahead predictions. These increase with the forecast horizon. Notice that, for all three forecast horizons, the large model produces the weakest forecasts in absolute terms.

For CPI inflation and, in difference to the point forecasts, we find that the cBVAR improves upon the BVAR for the one-quarter-ahead forecast horizon and for all three model sizes considered. For one-month-ahead and one-year-ahead predictions we find either negligible improvements (in the case of the small model) or losses.

Finally, for short-term interest rate forecasts we find improvements for forecast horizons greater than one-month-ahead. For the one-month-ahead horizon, modest gains are visible if the small model is adopted. For larger model sizes, improvements are minor.

In summary, the good performance for point forecasts is accompanied by more accurate density predictions. This holds not only for the three focus variables but also when we consider

the joint density forecasts.

Similarly to our Monte Carlo exercise, we observe that coarsening gains increase with model size, because larger versions of the standard BVAR suffer more heavily from misspecification. In particular, extreme events such as the global Covid-19 pandemic affect many series at once, leading to large losses in forecast gains if not adequately controlled for.

5.2.3 Density forecast performance over time

The discussion thus far has focused on average forecast performance. To analyze whether the performance gains from coarsening vary over time, we now analyze the evolution of the cumulative LPLs over time. The figure shows the variable-specific and joint cumulative LPLs across all three model sizes and forecast horizons. The blue lines refer to the one-month-ahead, the red lines to the one-quarter-ahead and the green lines to the one-year-ahead forecast horizons.

We again start with the overall density forecast performance, measured in terms of the joint LPLs over time and displayed in [Figure 2a](#). For small datasets, we find little differences for short-term forecasts (one-month- and one-quarter-ahead) until the onset of the pandemic. In this case, we observe that coarsening substantially improves predictive performance. This finding carries over to medium- and large-scale models. During the pandemic, some of the time series (such as the unemployment rate) displayed substantial outliers that are inconsistent with a Gaussian homoskedastic model. Coarsening, by contrast, leads to more conservative credible intervals, attributing a higher probability to these extreme observations and thus effectively improving density forecasting performance.

However, it should be noted that the joint density forecasting performance is not exclusively driven by the pandemic. When we consider the medium and large models, we observe that for higher-order forecasts, improvements relative to the BVAR arose much earlier and appear to increase during (or shortly after) turbulent periods.

Next, we consider variable-specific LPLs. For the unemployment rate (see [Figure 2b](#)), we find a picture similar to the one of the joint LPLs: smaller differences for one-month- and one-quarter-ahead LPLs up to the pandemic and then substantial gains during the pandemic. This holds for the small and medium dataset. For one-year-ahead LPLs we find that coarsening already helps shortly after the global financial crisis (GFC), approximately coinciding with the period where the unemployment rate peaked. This pattern holds for all three model sizes. Only

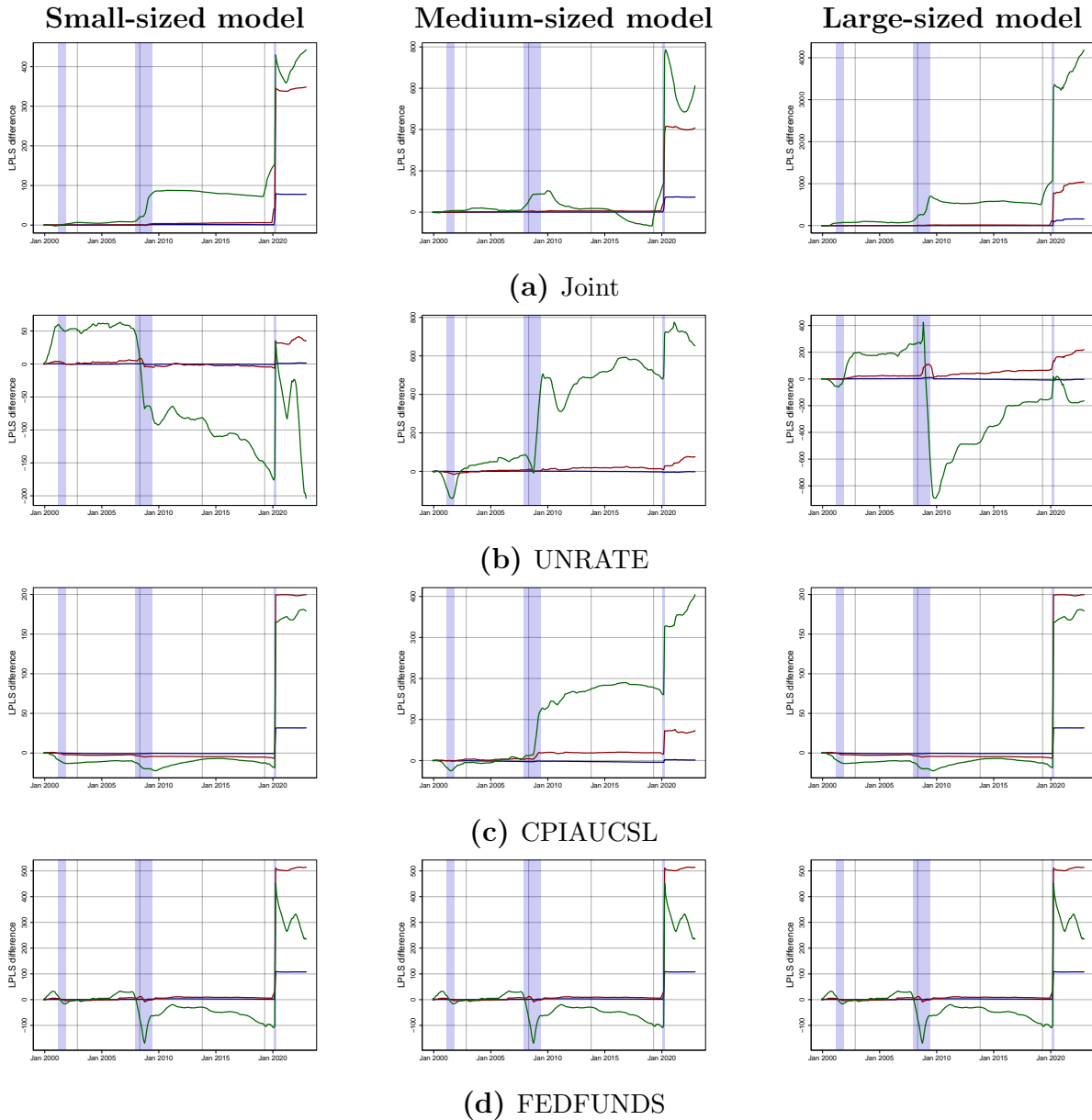


Figure 2: Cumulative LPLS differences over time.

Legend: — refers to the one-month-ahead, — to one-quarter-ahead and — to one-year-ahead forecasts. Shaded areas indicate the NBER recession dates.

for the large-sized model and one-year-ahead forecasts we find a much weaker performance of the cBVAR between 2010:01 and the pandemic.

For inflation (Figure 2c), we again find small differences between one-month-ahead and one-quarter-ahead forecast accuracy and all three model sizes. There is again some evidence that the inflation forecasts arising from the cBVAR become more accurate after the Covid-19 pandemic, a period characterized by a sustained increase in inflation. This pattern is visible for all forecast horizons except for the one-year-ahead forecast and the small model. In this

case, we find that cBVAR performs better than the benchmark in the first part of the hold-out period (2020:01 to around 2009:06), is inferior during the second part of the hold-out (2009:07 to 2020:05) before recovering. When we consider the medium-sized model and one-year-ahead predictions, we find that after the GFC, the forecast accuracy premium from coarsening increases, and the cBVAR consistently returns more accurate density predictions than the BVAR. These accuracy improvements also increase during the pandemic and post-pandemic part of the sample. For the large model, the pattern resembles that of the small model.

The LPLs for the short-term interest rate (Figure 2d) evolve similarly to the other two series. Small differences in the pre-pandemic period are followed by substantial variation during the pandemic. Strikingly, we find that there is an appreciable improvement in LPLs during the zero lower bound period only for the medium-sized model. This is not visible for the small and large information sets we consider, and, at first glance, appears to be inconsistent with our findings for the point forecasts. Notice, however, that for LPLs it does not only matter how well we fit the first moment of the predictive density, but higher order moments also play an important role.

5.3 How does misspecification change over time?

The previous sections emphasized that coarsening leads to improvements in predictive performance, particularly at longer forecast horizons. In this section, we ask how much weight is actually put on the likelihood function. To do so, we plot the learning rates for the three focus variables over the hold-out. Doing so gives us a general impression on the degree of misspecification and whether this is subject to time-variation.

Figure 3 shows the estimated ϕ'_i s for the unemployment rate (in blue), inflation (in red), and the short-term interest rate (in green) over the hold-out sample. Across the three model sizes, we observe that ϕ_i is much smaller than one throughout our hold-out period, pointing towards a substantial degree of misspecification. Among the focus series, we find that CPI inflation and short-term interest rates have the lowest learning rates whereas the one of the unemployment rate is much higher. An interesting finding is that, for all three target series, adding information to the model seems to increase the estimated learning rates. This finding is driven by the reduced risk of omitted variables in larger models.

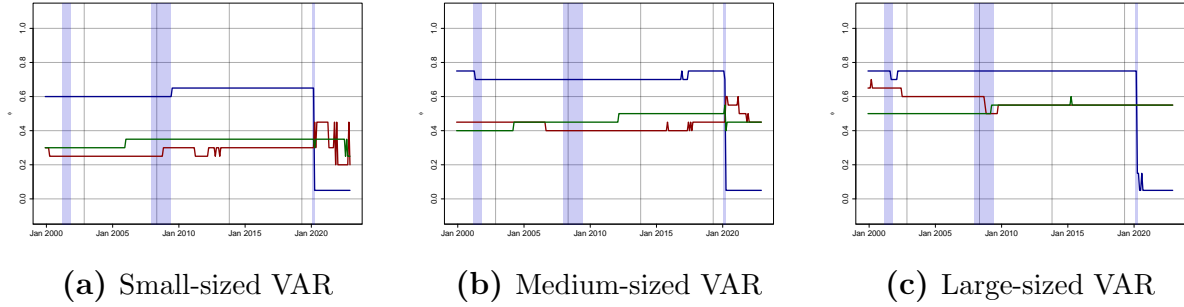


Figure 3: Estimated likelihood weights over the different training samples for the focus variables.

Legend: — refers to UNRATE, — to CPIAUCSL and — to FEDFUNDS. Blue shaded areas indicate the NBER recession dates.

In terms of time variation, we find some rather unsystematic movements in ϕ_i for all three variables prior to the pandemic. In some cases, the learning rate slightly increases while in other cases, we have declining learning rates. However, these movements are relatively small and should not affect the predictive distribution too much.

However, when we focus on the pandemic period, we find a common pattern across all models considered: the unemployment rate equation suddenly requires a much lower learning rate. Again, and consistent with the findings for LPLs over time, this indicates that the sharp increase in unemployment rates (and quick recoveries after the lockdowns subsided) requires a more flexible econometric model. Our cBVAR reflects this necessity by producing a much lower learning rate during these periods.

6 Conclusion

Common practice in empirical macroeconomics is to estimate simple and interpretable models to derive stylized facts and inform decision making. However, these simple models often suffer from misspecification of various, and unknown, types, which has a deleterious effect on estimation, inference, and predictive performance. In this paper, we offer a simple solution to obtain more robust results. Our proposal is to modify standard Bayesian multivariate econometric models. Instead of conditioning on the observed data, we propose to condition on the event that there is a distance between the sampling distribution of the observed data and a hypothetical idealized sampling distribution. Using relative entropy as a loss metric gives rise to a simple approximation that amounts to tempering the likelihood by a learning rate, connecting this approach to the

literature on power posteriors.

Since different endogenous series in a VAR can feature different degrees of misspecification, we allow for different learning rates across equations. The resulting likelihood is then coupled with an asymmetric prior, allowing for conjugate updating. To select the learning rates, we use a fully automatic approach that selects the prior shrinkage parameters and the learning rates jointly.

We consider an extensive Monte Carlo exercise with a wide range of different DGPs that are inspired by models actually used by empirical macroeconomists. These differ in terms of the conditional mean and variance specifications, giving rise to 16 different DGPs. We show that standard BVARs in the presence of misspecification of different forms produce predictive distributions that are substantially less accurate than those produced by the coarsened counterpart.

Using actual US monthly data, we show that the cBVAR produces more accurate point and density forecasts than the standard BVAR, with larger gains at longer horizons and when using small models, and it does not suffer much from the substantial outliers observed during the pandemic. This suggests that our approach can also be used to obtain more robust inference in the presence of large outliers.

References

- Antolin-Diaz, Juan, and Paolo Surico (2025), “The long-run effects of government spending,” *American Economic Review*, 115, p. 2376–2413, <https://www.aeaweb.org/articles?id=10.1257/aer.20231278>, <https://doi.org/10.1257/aer.20231278>.
- Bañbura, Marta, Domenico Giannone, and Lucrezia Reichlin (2010), “Large Bayesian vector auto regressions,” *Journal of Applied Econometrics*, 25, 71–92, <https://doi.org/10.1002/jae.1137>.
- Baumeister, Christiane, and James D Hamilton (2019), “Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks,” *American Economic Review*, 109, 1873–1910.
- Bernanke, Ben S, Jean Boivin, and Piotr Eliaszc (2005), “Measuring the effects of monetary policy: a factor-augmented vector autoregressive (favar) approach,” *The Quarterly Journal of Economics*, 120, 387–422.
- Bhattacharya, Anirban, Debdeep Pati, and Yun Yang (2019), “Bayesian fractional posteriors,” *The Annals of Statistics*, 47.
- Caggiano, Giovanni, Efrem Castelnuovo, and Gabriela Nodari (2022), “Uncertainty and monetary policy in good and bad times: A replication of the vector autoregressive investigation by bloom (2009),” *Journal of Applied Econometrics*, 37, 210–217, <https://doi.org/https://doi.org/10.1002/jae.2861>.
- Carriero, Andrea, Todd E Clark, and Massimiliano Marcellino (2015), “Bayesian vars: specification choices and forecast accuracy,” *Journal of Applied Econometrics*, 30, 46–73.
- Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino (2018), “Measuring uncertainty and its impact on the economy,” *The Review of Economics and Statistics*, 100, 799–815.
- Carriero, Andrea, Todd E Clark, Massimiliano Giuseppe Marcellino, and Elmar Mertens (2025), “Forecasting with shadow-rate vars,” *Quantitative Economics*.
- Chan, Joshua CC (2022), “Asymmetric conjugate priors for large bayesian vars,” *Quantitative Economics*, 13, 1145–1169.
- Check, Adam, and Jeremy Piger (2021), “Structural breaks in u.s. macroeconomic time series: A bayesian model averaging approach,” *Journal of Money, Credit and Banking*, 53, 1999–2036, <https://doi.org/https://doi.org/10.1111/jmcb.12822>.

- Clark, Todd E., Florian Huber, Gary Koop, Massimiliano Marcellino, and Michael Pfarrhofer (in-press), “Tail forecasting with multivariate bayesian additive regression trees,” *International Economic Review*, forthcoming, <https://doi.org/10.1111/iere.12619>.
- Cogley, Timothy, and Thomas J Sargent (2005), “Drifts and volatilities: monetary policies and outcomes in the post wwii us,” *Review of Economic dynamics*, 8, 262–302.
- (2015), “Measuring price-level uncertainty and instability in the united states, 1850–2012,” *Review of Economics and Statistics*, 97, 827–838.
- De Luigi, Clara, and Florian Huber (2018), “Debt regimes and the effectiveness of monetary policy,” *Journal of Economic Dynamics and Control*, 93, 218–238.
- Del Negro, Marco, and Frank Schorfheide (2004), “Priors from general equilibrium models for vars,” *International Economic Review*, 45, 643–673.
- Diebold, Francis X., and Robert S. Mariano (1995), “Comparing predictive accuracy,” *Journal of Business & Economic Statistics*, 13, 253–263, <https://doi.org/10.2307/1392185>.
- Forni, Mario, Luca Gambetti, and Luca Sala (2014), “No news in business cycles,” *The Economic Journal*, 124, 1168–1191, <https://doi.org/10.1111/eoj.12111>.
- Gefang, Deborah, Gary Koop, and Aubrey Poon (2023), “Forecasting using variational bayesian inference in large vector autoregressions with hierarchical shrinkage,” *International Journal of Forecasting*, 39, 346–363.
- Giannone, Domenico, Michele Lenza, and Giorgio E Primiceri (2015), “Prior selection for vector autoregressions,” *Review of Economics and Statistics*, 97, 436–451.
- (2019), “Priors for the long run,” *Journal of the American Statistical Association*, 114, 565–580.
- Gilboa, I., and D. Schmeidler (1989), “Maxmin expected utility with non-unique prior,” *Journal of Mathematical Economics*, 18, 141–153.
- González-Casasús, Oriol, and Frank Schorfheide (2025), “Misspecification-robust shrinkage and selection for var forecasts and irfs,” *arXiv preprint arXiv:2502.03693*.
- Goulet Coulombe, Philippe, Maxime Leroux, Dalibor Stevanovic, and Stéphane Surprenant (2022), “How is machine learning useful for macroeconomic forecasting?” *Journal of Applied Econometrics*, 37, 920–964.
- Grünwald, Peter, and Thijs van Ommen (2017), “Inconsistency of Bayesian Inference for Mis-

- specified Linear Models, and a Proposal for Repairing It,” *Bayesian Analysis*, 12, 1069 – 1103, <https://doi.org/10.1214/17-BA1085>, <https://doi.org/10.1214/17-BA1085>.
- Hansen, L.P., and T.J. Sargent (2008), “Robustness,” *Princeton University Press, Princeton, NJ*.
- Hansen, L.P., T.J. Sargent, G. Turmuhambetov, , and N. Williams (2006), “Robust control and model misspecification,” *Journal of Economic Theory*, 128, 45–90.
- Holmes, Chris C, and Stephen G Walker (2017), “Assigning a value to a power likelihood in a general bayesian model,” *Biometrika*, 104, 497–503.
- Huber, Florian, and Gary Koop (2023), “Subspace shrinkage in conjugate bayesian vector autoregressions,” *Journal of Applied Econometrics*, 38, 556–576.
- (2024), “Fast and order-invariant inference in bayesian vars with nonparametric shocks,” *Journal of Applied Econometrics*, 39, 1301–1320.
- Huber, Florian, Gary Koop, Luca Onorante, Michael Pfarrhofer, and Josef Schreiner (2020), “Nowcasting in a pandemic using non-parametric mixed frequency VARs,” *Journal of Econometrics*, in-press, <https://doi.org/10.1016/j.jeconom.2020.11.006>.
- Ingram, Beth F, and Charles H Whiteman (1994), “Supplanting the ‘minnesota’prior: Forecasting macroeconomic time series using real business cycle model priors,” *Journal of Monetary Economics*, 34, 497–510.
- Kadiyala, K Rao, and Sune Karlsson (1997), “Numerical methods for estimation and inference in bayesian var-models,” *Journal of Applied Econometrics*, 12, 99–132.
- Kalli, Maria, and Jim E. Griffin (2018), “Bayesian nonparametric vector autoregressive models,” *Journal of Econometrics*, 203, 267–282, <https://doi.org/10.1016/j.jeconom.2017.11.009>.
- Karlsson, Sune, Stepan Mazur, and Hoang Nguyen (2023), “Vector autoregression models with skewness and heavy tails,” *Journal of Economic Dynamics and Control*, 146, p. 104580.
- Kaufmann, Sylvia, and Christian Schumacher (2019), “Bayesian estimation of sparse dynamic factor models with order-independent and ex-post mode identification,” *Journal of Econometrics*, 210, 116–134.
- Koop, Gary M (2013), “Forecasting with medium and large bayesian vars,” *Journal of Applied Econometrics*, 28, 177–203.

- Korobilis, Dimitris, Bettina Landau, Alberto Musso, and Anthoulla Phella (2021), “The time-varying evolution of inflation risks,” *ECB Working Paper Series*, 2600.
- Liu, Philip, Konstantinos Theodoridis, Haroon Mumtaz, and Francesco Zanetti (2019), “Changing macroeconomic dynamics at the zero lower bound,” *Journal of Business & Economic Statistics*, 37, 391–404.
- Loria, Francesca, Christian Matthes, and Mu-Chun Wang (2022), “Economic theories and macroeconomic reality,” *Journal of Monetary Economics*, 126, 105–117.
- Marcellino, Massimiliano, James H Stock, and Mark W Watson (2006), “A comparison of direct and iterated multistep ar methods for forecasting macroeconomic time series,” *Journal of econometrics*, 135, 499–526.
- McCracken, Michael W, and Serena Ng (2016), “Fred-md: A monthly database for macroeconomic research,” *Journal of Business & Economic Statistics*, 34, 574–589.
- Miller, Jeffrey W, and David B Dunson (2018), “Robust bayesian inference via coarsening,” *Journal of the American Statistical Association*.
- Müller, Ulrich K (2013), “Risk of bayesian inference in misspecified models, and the sandwich covariance matrix,” *Econometrica*, 81, 1805–1849.
- Primiceri, Giorgio E (2005), “Time varying structural vector autoregressions and monetary policy,” *The Review of Economic Studies*, 72, 821–852, <https://doi.org/10.1111/j.1467-937x.2005.00353.x>.
- Schorfheide, Frank (2005), “Var forecasting under misspecification,” *Journal of Econometrics*, 128, 99–136.
- Sims, Christopher A, and Tao Zha (1998), “Bayesian methods for dynamic multivariate models,” *International Economic Review*, 949–968.
- Stock, James H, and Mark W Watson (2002), “Macroeconomic forecasting using diffusion indexes,” *Journal of Business & Economic Statistics*, 20, 147–162.

A Technical Appendix

A.1 Posterior Distribution

Combining the joint prior and likelihood functions, we can derive the joint posterior distribution as:

$$\begin{aligned}
& p(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_M, \sigma_1^2, \dots, \sigma_M^2 | \mathbf{Y}_1, \dots, \mathbf{Y}_M, \phi_1, \dots, \phi_M) \\
& \propto p(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_M, \sigma_1^2, \dots, \sigma_M^2) \tilde{p}(\mathbf{Y}_1, \dots, \mathbf{Y}_M | \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_M, \sigma_1^2, \dots, \sigma_M^2, \phi_1, \dots, \phi_M) \\
& = \prod_{i=1}^M p(\boldsymbol{\theta}_i, \sigma_i^2) \hat{p}(\mathbf{Y}_i | \boldsymbol{\theta}_i, \sigma_i^2, \phi_i) \\
& = \prod_{i=1}^M c_i (\sigma_i^2)^{-(\nu_i + \frac{i+MP}{2} + 1)} \exp\left(-\frac{1}{\sigma_i^2} (\underline{S}_i + \frac{1}{2} (\boldsymbol{\theta}_i - \underline{\boldsymbol{\theta}}_i)' \underline{\mathbf{V}}_i^{-1} (\boldsymbol{\theta}_i - \underline{\boldsymbol{\theta}}_i))\right) \\
& \quad \times (2\pi\sigma_i^2)^{-\frac{\phi_i T}{2}} \exp\left(-\frac{\phi_i}{2\sigma_i^2} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\theta}_i)' (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\theta}_i)\right) \\
& = \prod_{i=1}^M c_i (2\pi)^{-\frac{\phi_i T}{2}} (\sigma_i^2)^{-(\nu_i + \frac{\phi_i T + i + MP}{2} + 1)} \\
& \quad \times \exp\left(-\frac{1}{\sigma_i^2} (\underline{S}_i + \frac{1}{2} (\boldsymbol{\theta}'_i (\underline{\mathbf{V}}_i^{-1} + \phi_i \mathbf{X}'_i \mathbf{X}_i) \boldsymbol{\theta}_i - 2\boldsymbol{\theta}'_i (\underline{\mathbf{V}}_i^{-1} \underline{\boldsymbol{\theta}}_i + \phi_i \mathbf{X}'_i \mathbf{y}_i) + \underline{\boldsymbol{\theta}}'_i \underline{\mathbf{V}}_i^{-1} \underline{\boldsymbol{\theta}}_i + \phi_i \mathbf{Y}'_i \mathbf{Y}_i))\right) \\
& = \prod_{i=1}^M c_i (2\pi)^{-\frac{\phi_i T}{2}} (\sigma_i)^{-(\bar{\nu}_i + \frac{i+MP}{2} + 1)} \exp\left(-\frac{1}{\sigma_i^2} (\bar{S}_i + \frac{1}{2} (\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}}_i)' \bar{\mathbf{V}}_i (\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}}_i))\right)
\end{aligned}$$

A.2 Simulation of Data Generating Processes

Unless specified otherwise, the parameters below are obtained by estimating the corresponding specification on US data (using the datasets outlined in Table B.1).

All DGPs have the general form:

$$\mathbf{y}_t = \mathbf{A}_{1t} \mathbf{y}_{t-1} + \mathbf{A}_{2t} \mathbf{y}_{t-2} + \mathbf{B} \mathbf{X}_t + \mathbf{u}_t + \boldsymbol{\theta} \mathbf{u}_{t-1}, \quad \mathbf{u}_t \sim p(\mathbf{u}_t | \mathbf{0}_M, \boldsymbol{\Sigma}_t), \quad (5)$$

where \mathbf{y}_t is an M -dimensional vector of endogenous variables at time t and \mathbf{X}_t is a $K = 5$ -dimensional vector of additional exogenous regressors at time t . The associated parameter matrices are given by $\mathbf{A}_{1,t}$, $\mathbf{A}_{2,t}$, \mathbf{B} . The reduced-form shocks at time t are denoted \mathbf{u}_t . It follows some distribution $p(\cdot | \mathbf{0}_M, \boldsymbol{\Sigma}_t)$ with mean $\mathbf{0}_M$ and variance-covariance matrix $\boldsymbol{\Sigma}_t = \mathbf{A}_0 \mathbf{D}_t \mathbf{A}_0'$. Here, \mathbf{A}_0 is a lower triangular matrix with ones on the main diagonal and \mathbf{D}_t a diagonal matrix containing the error variances at time t . Lastly, $\boldsymbol{\theta}$ is a moving average coefficient matrix.

We consider four different kinds of DGPs with respect to modeling the conditional mean and the error structure. These are outlined in further detail below.

A.2.1 Variations with respect to the conditional mean

1. Our baseline is a linear VAR ("Linear") where we set $\mathbf{A}_{1t} = \mathbf{A}_1, \mathbf{A}_{2t} = \mathbf{A}_2, \forall t = 1, \dots, T$ and $\mathbf{B} = \mathbf{0}_{M \times K}$. The parameter matrices \mathbf{A}_1 and \mathbf{A}_2 are obtained by estimating a VAR(2) following [Giannone, Lenza, and Primiceri \(2015\)](#). This model has been applied extensively in the empirical literature, a recent example is [Antolin-Diaz and Surico \(2025\)](#).
2. In the second case, the DGP arises from a smooth transition model ("Transition") and we set $\mathbf{A}_{pt} = s_t \mathbf{A}_{pa} + (1 - s_t) \mathbf{A}_{pb}$, with $\mathbf{A}_{pa}, \mathbf{A}_{pb}$ being the baseline regimes for $p = 1, 2$. The mixture variable $s_t \in [0, 1]$ is given by the following equation:

$$s_t = \frac{1}{1 + \exp((- \gamma) z_t)},$$

where we set the speed of transition parameter $\gamma = 3$ and simulate z_t arise from a random walk with unit variance. We set $\mathbf{B} = \mathbf{0}_{M \times K}$ and, conditional on s_t (which is known given $\{z_t\}_{t=1}^T$), estimate the parameter matrices again using BVAR(2) models on the corresponding macro time series. An recent application of this approach is [Caggiano, Castelnuovo, and Nodari \(2022\)](#).

3. The third DGP follows a break-point model ("Break") such that the VAR coefficients are given by $\mathbf{A}_{pt} = \mathbf{A}_{pk}, p = 1, 2$, where \mathbf{A}_{pk} are changing every 50 observations. We again set $\mathbf{B} = \mathbf{0}_{M \times K}$ and obtain the coefficients through BVAR(2) models estimated over the respective sub-samples. A model of this kind has recently been applied by [Check and Piger \(2021\)](#).
4. In the case of the DGP featuring exogenous regressors ("Exo"), we let $\mathbf{A}_{1t} = \mathbf{A}_1, \mathbf{A}_{2t} = \mathbf{A}_2, \forall t = 1, \dots, T$ (as in the linear baseline model) but allow \mathbf{B} to be a non-zero coefficient matrix which each of its elements drawn from a normal distribution with mean zero and variance $\frac{1}{M^2}$. We simulate each element of X_t from an AR(1) model with persistence parameter 0.95 and error variance 0.0625. An example of this is [Forni, Gambetti, and Sala \(2014\)](#).

A.2.2 Variations with respect to the error structure

1. For the DGP labeled "Gaussian" we let the errors arise from a homoskedastic Gaussian distribution $p(\cdot) = \mathcal{N}(\mathbf{0}_M, \mathbf{D})$ (such that $\mathbf{D}_t = \mathbf{D} \forall t = 1, \dots, T$). \mathbf{D} has been calibrated by estimating a VAR(2) model following [Giannone, Lenza, and Primiceri \(2015\)](#). We set $\boldsymbol{\theta} = \mathbf{0}_{M \times M}$. This simple model has been extensively applied in the empirical literature, a recent example is again [Antolin-Diaz and Surico \(2025\)](#).
2. Additionally, we explore a similar setup (labeled "Student") with $p(\cdot) = t_3(\mathbf{0}_M, \mathbf{D})$ being the Student-t distribution with three degrees of freedom and again $\mathbf{D}_t = \mathbf{D} (\forall t = 1, \dots, T)$ and $\boldsymbol{\theta} = \mathbf{0}_{M \times M}$ ("Student"). Recent work by [Karlsson, Mazur, and Nguyen \(2023\)](#) emphasizes the empirical importance of allowing for heavy tails in the error distribution.
3. The DGP considering stochastic volatility ("SV") sets assumes that the natural logarithm of the error variances evolves according to the following AR(1) model:

$$\ln(\mathbf{D}_t) = \ln(\mathbf{D}) + 0.95 \cdot (\ln(\mathbf{D}_{t-1}) - \ln(\mathbf{D})) + \boldsymbol{\nu}_t$$

where each element of $\boldsymbol{\nu}_t$ is normally distributed with mean zero and variance $\eta^2 = 0.05$. We let $p(\cdot) = \mathcal{N}(\mathbf{0}_M, \mathbf{D}_t)$ and $\boldsymbol{\theta} = \mathbf{0}_{M \times M}$. A recent example of Vector autoregressions featuring stochastic volatility is [Carriero, Clark, and Marcellino \(2018\)](#).

4. For the DGP with a moving-average term ("MA") we set $p(\cdot) = \mathcal{N}(\mathbf{0}_M, \mathbf{D})$, ($\mathbf{D}_t = \mathbf{D} \forall t = 1, \dots, T$). We let $\boldsymbol{\theta}$ be non-zero and draw each of its elements from a normal distribution with mean zero and variance $\frac{1}{M^2}$. A DGP along these lines has recently been considered in [González-Casasús and Schorfheide \(2025\)](#).

B Data appendix

We provide information on the variables comprising the small-, medium- and larged sized models in [Table B.1](#). The table also contains their respective FRED-MD variable codes and information on the transformations we applied. To preserve their cointegrating relationship, all variables (except price series) are typically in (log-)levels

Table B.1: Variables used in the small, medium and larged models alongside their FRED-MD codes and transformations. Transformation codes refer to the following: 1 = level, 4 = log-level, 5 = log-first differences.

FRED-MD Code	Variable	Transformation	Small	Medium	Large
UNRATE	Civilian Unemployment Rate	4	X	X	X
CPIAUCSL	CPI : All Items	5	X	X	X
FEDFUNDS	Effective Federal Funds Rate	1	X	X	X
AWHMAN	Avg Weekly Hours : Manufacturing	4		X	X
M2REAL	Real M2 Money Stock	5		X	X
S&P 500	S&P s Common Stock Price Index: Composite	5		X	X
INDPRO	IP Index	5		X	X
T10YFFM	10-Year Treasury C Minus FEDFUNDS	1		X	X
CUSR0000SAC	CPI : Commodities	5		X	X
CUMFNS	Capacity Utilization: Manufacturing	4			X
CE16OV	Civilian Employment	5			X
CLAIMSx	Initial Claims	4			X
RPI	Real Personal Income	5			X
CES3000000008	Avg Hourly Earnings : Manufacturing	5			X
HOUST	Housing Starts: Total New Privately Owned	5			X
S&P PE ratio	S&P s Composite Common Stock: Price-Earnings Ratio	4			X
DPCERA3M086SBEA	Real personal consumption expenditures	5			X
RETAILx	Retail and Food Services Sales	5			X
TOTRESNS	Total Reserves of Depository Institutions	5			X
NONREVSL	Total Nonrevolving Credit	5			X
GS10	10-Year Treasury Rate	1			X
BAAFFM	Moody s Baa Corporate Bond Minus FEDFUNDS	1			X
EXJPUSx	Japan / U.S. Foreign Exchange Rate	1			X
EXUSUKx	U.S. / U.K. Foreign Exchange Rate	1			X
WPSFD49207	PPI: Finished Goods	5			X
OILPRICEx	Crude Oil, spliced WTI and Cushing	5			X
DSERRG3M086SBEA	Personal Cons. Exp: Services	5			X

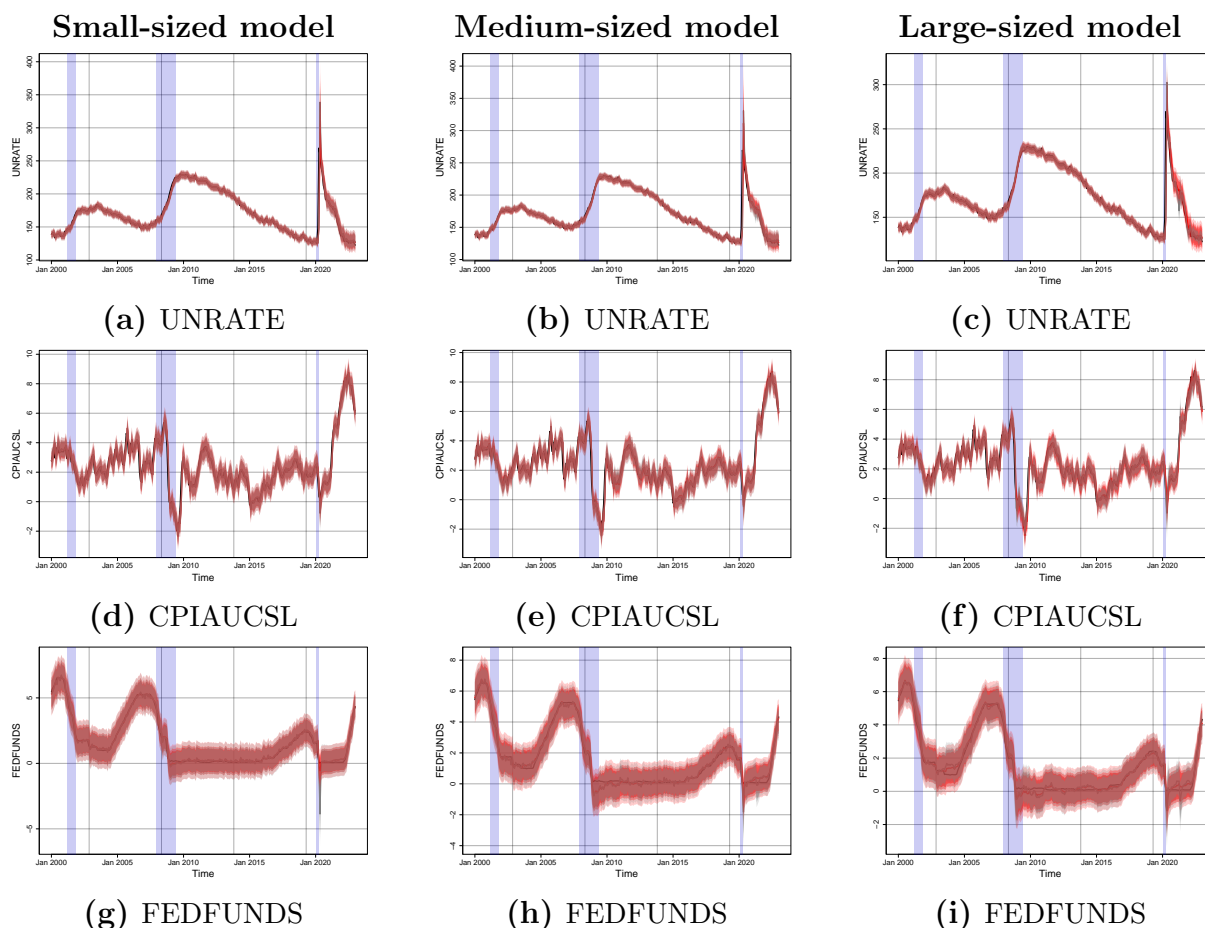


Figure C.1: Observed path versus one-month-ahead predictive distributions for each of the focus variables.

Legend: — are the observed values, — indicates predictions of the cBVAR and — those of the standard BVAR. Lines denote the posterior median and shaded areas the 90%, 95% and 99% credible intervals (from dark to light). Blue shaded areas indicate the NBER recession dates.

C Additional Empirical Results

Table C.2 contains mean squared errors (based on the posterior median) for the Monte Carlo exercise. Again, results for the cBVAR are in relative terms such that scores smaller than one reflect forecast gains.

Additionally, we plot the predictive distributions for each of the focus variables and model sizes at the one-month-ahead (Figure C.1), one-quarter-ahead (Figure C.2) and one-year-ahead (Figure C.3)

Table C.2: Relative MSE scores for predictive distributions.

		Gaussian		Student		SV		MA		
		rel. MSE		rel. MSE		rel. MSE		rel. MSE		
Small-sized	<i>One-month-ahead predictions</i>									
	Linear	1.01		1.02		1.01		1.01		
	Break	0.99	***	1.00		0.99	***	0.99	***	
	Exo	1.00		1.03		1.01		1.01		
	Transition	1.03		1.05		1.03		1.01		
	<i>One-quarter-ahead predictions</i>									
	Linear	1.01		1.01		1.01		1.01		
	Break	0.98	***	0.98	***	0.98	***	0.98	***	
	Exo	0.98	**	1.02		0.99	**	1.00		
	Transition	1.03		1.06		1.03		1.02		
	<i>One-year-ahead predictions</i>									
	Linear	1.04		1.05		1.05		1.00		
Break	0.98	***	0.99		0.99	***	0.99	***		
Exo	0.96	***	1.00		0.96	***	0.99	*		
Transition	1.03		1.11		1.03		0.99	***		
Medium-sized	<i>One-month-ahead predictions</i>									
	Linear	0.99	***	1.00		1.00		1.02		
	Break	0.92	***	0.91	***	0.92	***	0.97	***	
	Exo	0.99	**	0.98	***	0.98	***	1.04		
	Transition	1.01		1.04		1.01		1.02		
	<i>One-quarter-ahead predictions</i>									
	Linear	0.98	***	0.98	***	0.99	***	0.98	***	
	Break	0.89	***	0.88	***	0.89	***	0.93	***	
	Exo	0.95	***	0.95	***	0.95	***	0.99		
	Transition	1.01		1.07		1.02		1.02		
	<i>One-year-ahead predictions</i>									
	Linear	1.02		1.05		1.03		0.98	***	
Break	0.97	***	1.02		1.00		0.98	***		
Exo	0.92	***	0.98		0.95	**	0.92	***		
Transition	1.02		1.18		1.04		1.00	*		
Large-sized	<i>One-month-ahead predictions</i>									
	Linear	0.96	***	0.96	***	0.96	***	0.96	***	
	Break	0.94	***	0.94	***	0.94	***	0.92	***	
	Exo	0.92	***	0.93	***	0.92	***	0.90	***	
	Transition	1.01		1.02		1.01		1.02		
	<i>One-quarter-ahead predictions</i>									
	Linear	0.90	***	0.90	***	0.90	***	0.87	***	
	Break	0.93	***	0.94	***	0.92	***	0.91	***	
	Exo	0.88	***	0.87	***	0.88	***	0.86	***	
	Transition	1.01		1.02		1.01		1.00	***	
	<i>One-year-ahead predictions</i>									
	Linear	0.96	***	0.96	***	0.97	***	0.93	***	
Break	0.99	**	1.01		0.99	***	0.98	***		
Exo	0.91	***	0.92	***	0.92	***	0.88	***		
Transition	1.00	***	1.03		1.00		0.99	***		

Notes: Results are averaged over all draws and test sets for each DGP. Stars indicate p-values from a one-sided t-test with alternative hypothesis that cBVAR produces smaller MSE scores.

Legend: . $\sim p < 0.16$; * $\sim p < 0.1$; ** $\sim p < 0.05$; *** $\sim p < 0.01$

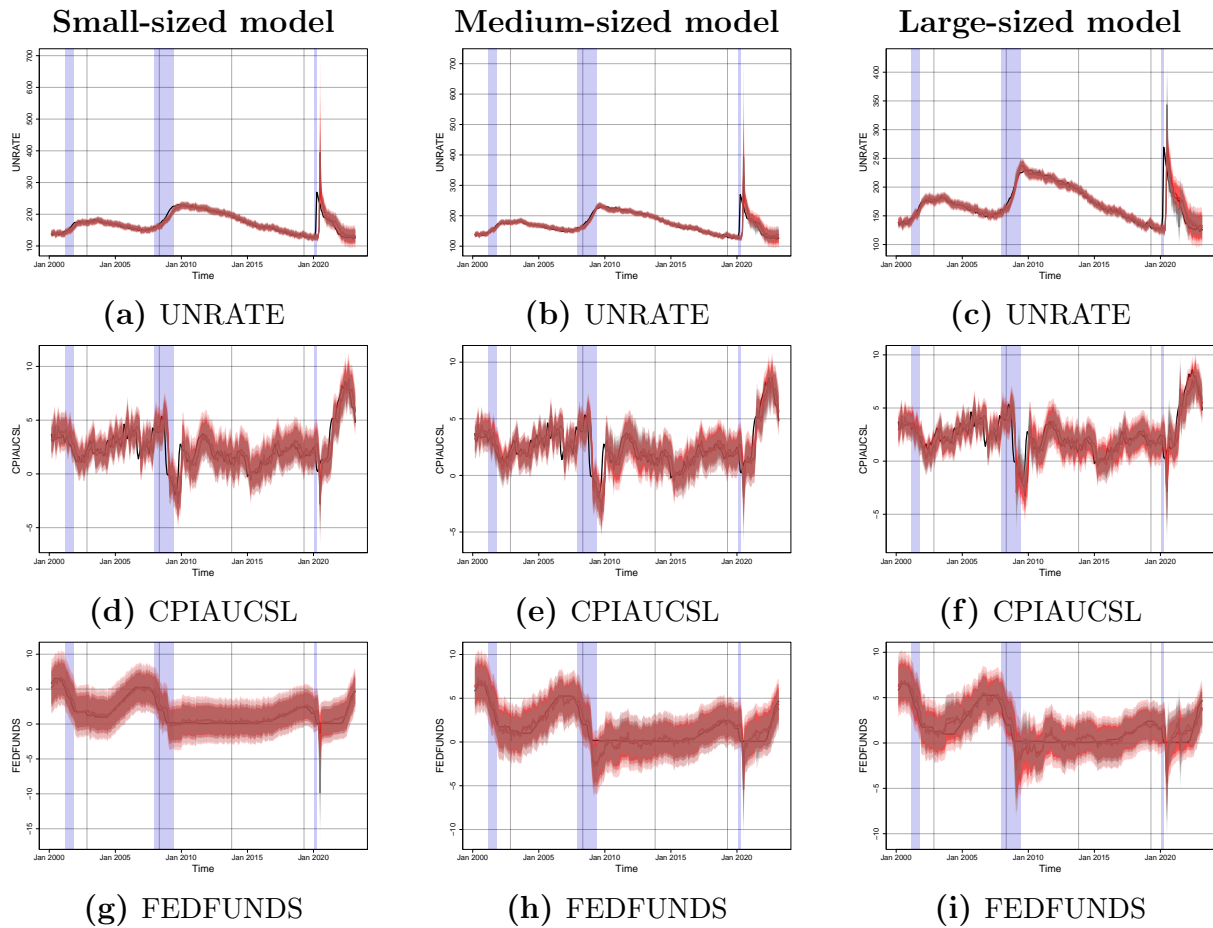


Figure C.2: Observed path versus one-quarter-ahead predictive distributions for each of the focus variables.

Legend: - indicates predictions of the cBVAR and - those of the standard BVAR. Lines denote the posterior median and shaded areas the 90%, 95% and 99% credible intervals (from dark to light). Blue shaded areas indicate the NBER recession dates.

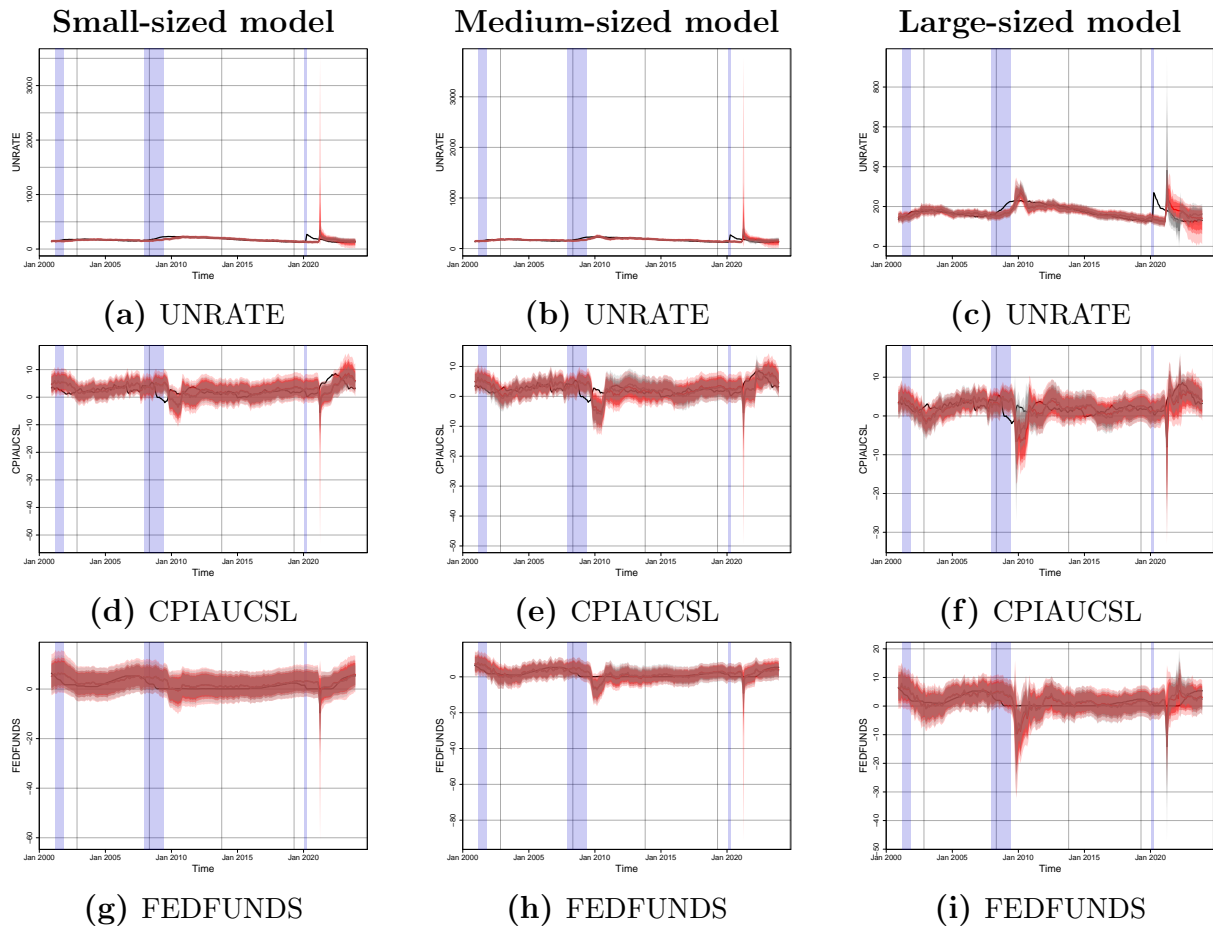


Figure C.3: Observed path versus one-year-ahead predictive distributions for each of the focus variables.

Legend: - indicates predictions of the cBVAR and - those of the standard BVAR. Lines denote the posterior median and shaded areas the 90%, 95% and 99% credible intervals (from dark to light). Blue shaded areas indicate the NBER recession dates.