EUROPEAN CENTRAL BANK
EUROSYSTEM

# Working Paper Series

Andrian Yambolov

## How to conduct joint Bayesian inference in VAR models?

No 3100

**Abstract**

When economic analysis requires simultaneous inference across multiple variables and time horizons, this paper shows that conventional pointwise quantiles in Bayesian structural vector autoregressions significantly understate the uncertainty of impulse responses. The performance of recently proposed joint inference methods, which produce noticeably different error band estimates, is evaluated, and calibration routines are suggested to ensure that they achieve the intended nominal probability coverage. Two practical applications illustrate the implications of these findings: (i) within a structural vector autoregression, the fiscal multiplier exhibits error bands that are 51% to 91% wider than previous estimates, and (ii) a pseudo-out-of-sample projection exercise for inflation and gross domestic product shows that joint inference methods could effectively summarize uncertainty for forecasts as well. These results underscore the importance of using joint inference methods for more robust econometric analysis.

*Keywords:* vector autoregressions, impulse responses, forecasts, pointwise inference, joint inference
*JEL Codes:* C22, C32, C52

## Non-technical summary

Standard methods for constructing error bands around impulse response functions consider them in isolation, neglecting the estimation uncertainty that arises across variables and time horizons due to the joint nature of the underlying structural parameters. For example, one approach to assessing the impact of government expenditure on economic activity—known as the fiscal multiplier—is to estimate the ratio between the cumulative responses of gross domestic product and government expenditure. By convention, practitioners use marginal error bands to address this and similar economic questions, which leads to an underestimation of uncertainty.

This paper quantifies the extent to which conventional error bands understate uncertainty and focuses on methods for conducting joint inference to support more robust economic analysis. It conducts a series of simulation experiments using widely adopted Bayesian vector autoregressions to evaluate the performance of several estimators, discussed by Montiel Olea and Plagborg-Møller (2019) and Inoue and Kilian (2022), that account for the joint uncertainty of impulse response functions. In addition, it proposes strategies to improve upon those joint inference methods, which tend to overstate the uncertainty.

The policy relevance of the paper is illustrated through two examples. The first draws on a widely used fiscal vector autoregression estimated with U.S. data. It shows that taking into account the joint uncertainty between gross domestic product and government expenditure can uncover stronger and earlier effects of fiscal policy shocks than conventional approaches suggest. In particular, the results indicate that the fiscal multiplier could exceed one as early as four quarters before standard estimates would imply. This highlights the potential benefits of timely and proactive fiscal interventions—especially during periods of economic stress when swift action is essential. Moreover, as Blanchard and Leigh (2013) emphasize, misestimating the fiscal multiplier can result in substantial forecast errors, particularly during periods of sharp changes in government expenditure.

The second example demonstrates how joint inference methods can be employed to assess the probability that both the growth rate of gross domestic product and inflation remain within specified bounds over the projection horizon. This is especially important for central banks with a dual mandate, but more broadly also for any forecasting exercise which strives to properly measure and report economic risks. In practice, the uncertainty around economic projections is often adjusted at the discretion of forecasters or policy makers to account for risks that macroeconometric models may not fully capture. For instance, if uncertainty around the growth rate of gross domestic product seems to be too low or unbalanced, its error bands may be widened or skewed. Since economic activity and inflation are linked, the uncertainty about the latter must also be adjusted—something that is often overlooked (see Bernanke, 2024). The methods developed in this paper provide a practical way to capture this type of estimation uncertainty and improve the reliability of macroeconomic projections.

# 1 Introduction

The impulse response function (IRF) is the most widely used tool for analyzing economic questions within a structural vector autoregression (VAR) framework. Regardless of the estimation methodology, whether frequentist or Bayesian, practitioners typically construct pointwise (individual) bands[1] around the impulse response coefficients at each horizon separately. However, most economic questions necessitate simultaneous evaluation across multiple structural parameters.

For example, a standard monetary VAR model could be used to examine whether the exchange rate overshoots in response to a monetary policy shock (see Eichenbaum and Evans, 1995). Alternatively, the same model could study the conditional exchange rate pass-through by computing the ratio between the responses of the price level and the exchange rate following a monetary policy shock (see Forbes et al., 2018). While the former question involves a consideration of cross-horizon relationship in the IRF of a single variable, the latter requires inference method that accounts for cross-variable dependence.

This example illustrates why error bands should be viewed as context-specific objects, whose proper specification requires thorough discussion before any empirical analysis. Lütkepohl et al. (2015, 2018) have evaluated the empirical properties of frequentist methods for joint inference in VARs using Monte Carlo experiments. These studies are comprehensive, covering a wide range of methods that have been discussed and assessed. Overall, Lütkepohl and his coauthors demonstrate that some of the methods they examine could be used to conduct joint inference across multiple parameters. However, the practical relevance of these findings may be somewhat limited, given that large share of the current empirical research using VARs tends to rely on Bayesian estimation. This raises the main question addressed in the paper: how can joint error bands be constructed to achieve a pre-specified probability content in Bayesian VAR models?

In the Bayesian literature, two recently proposed methods for conducting joint inference have emerged, with significant implications due to their adaptability to various time series models. The sup-t estimator introduced by Montiel Olea and Plagborg-Møller (2019), employs numerical optimization to determine the optimal pointwise quantile that achieves joint coverage across multiple parameters. Another method independently introduced by Akram et al. (2016) and Inoue and Kilian (2022) treats the multivariate posterior draws as individual models satisfying structural restrictions. This method uses a vector-valued loss function to retain a specified share of draws (corresponding to the credibility level) with the lowest loss values. An envelope around the IRFs is then constructed by taking the minimum and maximum responses across horizons and structural parameters—a step that motivates the use of the label min-max for the method in this paper. While both methods are readily applicable in practice, their empirical properties

[1] In the frequentist framework, these are confidence intervals, while in the Bayesian framework, they are credible intervals.

have not been thoroughly evaluated, raising questions whether the error bands achieve the desired probability content. More importantly, the sup-t and the min-max estimator produce considerably different uncertainty estimates.

By conducting series of simulation experiments using four widely adopted VAR models, I demonstrate that conventional pointwise quantiles understate estimation uncertainty when the analysis requires joint inference across multiple parameters. Moreover, the sup-t estimator achieves the nominal probability content jointly across multiple parameters, but the performance of the min-max estimator depends on the specific vector-valued loss function, which is employed. In particular, the min-max estimator relying on absolute, quadratic, or angular loss functions, as suggested by Inoue and Kilian (2022), tends to overstate joint uncertainty and produces more conservative estimates of the error bands. Conversely, the min-max estimator using the Chebyshev loss function, as proposed by Akram et al. (2016), shows similar coverage to the sup-t estimator. A further important contribution of this paper is that, I enhance the min-max estimator under the three loss functions proposed by Inoue and Kilian (2022) by adding an additional calibration step. The error bands computed with the enhanced min-max estimator are subsequently shown to achieve the nominal probability content, but they are tighter in comparison to the error bands computed with the sup-t estimator. Finally, I also show that the min-max estimator preserves the covariance among structural parameters, which are within the joint error bands. This supports the argument put forward by Inoue and Kilian (2022), who motivate the usage of their method by its ability to capture the shape and comovement of IRFs, rather than concentrating only on error bands that have the desired probability content.

The remainder of the paper is organized as follows. The next section reviews the relevant literature, also discussing the frequentist methods, as advancements in this framework have significantly contributed to enhancing the Bayesian procedures proposed in this study. Section 3 describes the joint inference methods evaluated in this paper and introduces the experimental design, which follows the simulation-based calibration procedure recently proposed by Talts et al. (2020). To the best of my knowledge, this is the first paper to apply it in macroeconometrics. Its primary practical application lies in evaluating whether any Bayesian estimation routine properly samples from the underlying posterior distribution. Section 4 outlines the data generating processes (DGPs) and presents the analysis of results. Section 5 highlights the practical contributions of the joint inference methods, showing that error bands for the fiscal multiplier computed with joint inference methods could be between 51% and 91% wider, depending on the selected credibility level, compared to conventional pointwise estimates. The second example demonstrates the similarities between IRFs and forecasts by applying the joint inference methods to a pseudo-out-of-sample projection exercise for inflation and GDP in the United States and euro area using two distinct time series models. Section 6 concludes the paper.

## 2 Literature Review

Joint inference methods have long been a focal point of interest for econometricians. For example, the Wald test is commonly used to test the validity of a set of linear parameter restrictions in a regression model. Joint inference methods for VAR models can be thought of in a similar way, but they have two key dimensions: they can be used either to construct error bands that achieve the nominal coverage level or to capture the joint distribution among structural parameters. The methods discussed in this section address either the first aspect or both.

Sims and Zha (1999) were among the first to recognize the limitations of pointwise inference for IRFs. Although their solution relies on Bayesian methods, most subsequent studies have employed frequentist techniques (e.g., Staszewska, 2007; Jordà, 2009; Inoue and Kilian, 2016). The simplest technique for joint inference about IRFs is presented in the well-known textbook by Lütkepohl (2005). It builds on Bonferroni's (1936) method, which adjusts significance levels for multiple tests by dividing the overall significance level by the number of structural parameters considered. This adjustment ensures that the probability of making one or more Type I errors across multiple comparisons remains below the desired threshold.

Lütkepohl et al. (2015) provide a review of the most commonly used frequentist approaches for conducting joint inference about IRFs. Their objective is to compare the coverage levels of pointwise and joint confidence bands through Monte Carlo simulations. They demonstrate that conventional confidence bands exhibit very low coverage levels in VARs when the economic analysis requires joint inference across multiple structural parameters. Additionally, they find that Bonferroni error bands are overly conservative in summarizing joint uncertainty. To address this, Lütkepohl et al. (2015) propose an iterative procedure that discards draws from the bootstrap distribution until the resulting error bands achieve the nominal joint coverage level. This procedure has inspired the calibration techniques proposed in this study to enhance Bayesian inference methods. Lütkepohl et al. (2015) also show that the adjusted Bonferroni error bands, along with another method known as the neighboring paths procedure (Staszewska, 2007), systematically achieve the desired nominal joint coverage level. Another significant contribution of their study is the finding that Scheffé bands, as proposed by Jordà (2009), do not reach the prespecified confidence level.[2] Conversely, confidence bands adapted from the method by Wolf and Wunderli (2015) for constructing multi-horizon forecasts do achieve sufficient coverage but tend to be wider compared to other joint inference techniques.[3] In a similar Monte Carlo study, Lütkepohl et al. (2018) demonstrate that joint confidence bands, constructed by evaluating the density of structural models, are also preferable when the researcher seeks a coverage level corresponding to the prespecified confidence level.

[2]The Scheffé is similar to the Bonferroni approach, but adjusts the critical value instead of the significance level.
[3]Both of these methods have only frequentist interpretations and are not discussed further here for brevity.

The Bayesian literature focusing on joint inference about IRFs in VARs remains relatively scarce. Inoue and Kilian (2013) propose a method to estimate a joint credible region by ranking the posterior density of each structural draw. Initially, draws are generated from the joint posterior distribution of the reduced-form VAR parameters, and the posterior density value is computed. Subsequently, draws that do not satisfy the theoretically motivated sign restrictions on the IRFs are discarded, and the remaining draws are ranked by the value of their posterior density. According to Inoue and Kilian (2013), the draw that maximises the posterior density, known as the modal model, serves as an estimator of the central tendency of the IRFs. Furthermore, the $(1-\alpha)100\%$ highest posterior density credible region of admissible models is used as an estimator for the inference about the IRFs.

In a more recent study, Montiel Olea and Plagborg-Møller (2019) propose the so-called sup-t method for conducting joint inference about IRFs. Their main analysis is conducted within the frequentist framework[4], but they also propose a procedure that could be applied within Bayesian estimation. This estimator requires the use of numerical optimization to find the pointwise quantiles that achieve the nominal coverage level for all structural parameters simultaneously. Montiel Olea and Plagborg-Møller (2019) also compare their method with more traditional approaches for joint inference, such as the Bonferroni (1936) and Šidák (1967) error bands. The Šidák method, similar to Bonferroni, offers a slightly more powerful adjustment by assuming the tests are independent, and adjusts the significance level using a formula that accounts for this assumption. In any case both methods are used only as a benchmark comparison to show that the sup-t method produces tighter error bands. Although this is the only study in the Bayesian framework that provides a simulation exercise to evaluate the empirical properties of a joint error band, their empirical model does not resemble any established VAR, but rather serves as a theoretical example. Furthermore, the simulation does not provide a fully Bayesian interpretation, as it is only valid for a single point estimate. A further drawback of this method is that it only provides error bands, preventing the applied econometrician from making probability statements about the conditional shape or comovements of endogenous variables.

Another approach for modeling joint inference within a generic Bayesian framework was originally proposed by Bernardo (2011). More recently, Inoue and Kilian (2022) adapted this technique to VARs to conduct joint inference about IRFs. The method addresses both shortcomings associated with pointwise IRFs. Compared to the joint inference method proposed earlier in Inoue and Kilian (2013), it is less computationally burdensome and, crucially, more easily adaptable to different VAR models, without a need to specify the structural likelihood function, which might not be a trivial task.

In essence, for each posterior draw of the vector of structural parameters, Inoue and Kilian (2022) calculate the average distance to all other draws under a vector-valued loss function,

[4]The frequentist sup-t estimator is also suggested by Inoue et al. (2025) as a reliable method for conducting joint inference about IRFs estimated with local projections.

and then retain the first $(1-\alpha)100\%$ draws, starting with the draw having the lowest value. The constructed lowest posterior risk region visually approximates the joint credible region as the number of draws approaches infinity. The members of this region resemble the shotgun trajectories used in ballistics rather than the conventional credible band.[5] Finally, the joint error bands are computed by constructing the upper and lower envelopes around the shotgun trajectories within the credible region, which is referred to as the min-max estimator.

In a related study, Akram et al. (2016) generate projection fan charts for a Bayesian dynamic stochastic general equilibrium model using a simplified version of the methodology proposed by Inoue and Kilian (2022). Their approach does not depend on the average distance across draws, but rather on the distance of each draw from a pointwise measure of central tendency—specifically, the pointwise mean. Another key difference between the two studies is that Inoue and Kilian (2022) employ a variety of vector-valued loss functions, including absolute, quadratic, angular, and Dirac, while Akram et al. (2016) use only the Chebyshev specification. The choice of loss function proves to be crucial for the performance of the min-max estimator, as demonstrated in the latter sections of this paper.

Despite the clear disadvantages of partially ignoring estimation uncertainty due to interdependence among structural parameters during inference, joint error bands have rarely been used in applied research (e.g., Herrera and Rangaraju, 2020; Zhou, 2020; Kilian and Zhou, 2022; Arias et al., 2022; Güntner et al., 2024). One possible explanation is that joint error bands yield significantly wider uncertainty estimates compared to conventional pointwise methods, yet their performance has not been thoroughly assessed in simulation studies.

# 3 Methodology

## 3.1 Conducting Inference

In this subsection, impulse response coefficients are defined first, followed by description of the uncertainty concepts and an explanation of the construction of the error bands using all inference methods evaluated in this study.

Starting from any reduced-form VAR estimated using Bayesian methods, one can impose structural restrictions motivated by economic theory to identify impulse response coefficients. Specifically, the $(n,s)$-th element of the impulse response coefficient matrix at horizon $h$, $\theta_{n,s,h}$, quantifies the effect on variable $n$ in response to a structural shock $s$ of size one in period $h$. The impulse response matrix at horizon $h$ is defined as:

$$\boldsymbol{\Theta}_h = \{\theta_{n,s,h}\}.$$

[5]The shotgun trajectories were originally proposed in the context of frequentist methods by Inoue and Kilian (2016).

Appendix A provides a detailed description of the reduced-form VAR model, the prior distribution and posterior sampling routines, as well as the structural identification strategies employed in the simulation experiments presented in this paper. This information is omitted from the main text for the sake of simplicity.

**Pointwise Inference.** Pointwise inference about IRFs is conducted by constructing credible interval that achieve predefined probability content for each structural parameter individually. For the sake of simplicity, a generalized vector of stacked structural parameters $\boldsymbol{\theta} \in \mathbb{R}^K$ is defined.

**Definition 1: Credible interval.** *The* $(1-\alpha)100\%$ *credible interval,* $\boldsymbol{\Theta}_j^{1-\alpha}$*, for any structural parameter* $\theta_j \in \mathbb{R}$ *is an interval with lower and upper bounds given by* $\left[\underline{\theta}_j, \bar{\theta}_j\right]$ *and coverage probability* $1-\alpha$ *such that*

$$Pr\left(\theta_j \in \boldsymbol{\Theta}_j^{1-\alpha}\right) = Pr\left(\underline{\theta}_j^{1-\alpha} \leq \theta_j \leq \bar{\theta}_j^{1-\alpha}\right) = 1-\alpha.$$

This credible interval is expected to include the actual value of $\theta_j$ with a probability $1-\alpha$. The conventional approach of constructing the error bands is to compute the corresponding quantiles over all posterior draws.

**Definition 2: PW error band.** *The* $(1-\alpha)100\%$ *pointwise (PW) error band,* $\widehat{\boldsymbol{\Theta}}_j^{1-\alpha,PW}$*, for any structural parameter,* $\theta_j \in \mathbb{R}$*, is an interval given by*

$$\widehat{\boldsymbol{\Theta}}_j^{1-\alpha,PW} = \left[\underline{\theta}_j^{1-\alpha,PW}, \bar{\theta}_j^{1-\alpha,PW}\right],$$

*where* $\underline{\theta}_j^{1-\alpha,PW}$ *and* $\bar{\theta}_j^{1-\alpha,PW}$ *are the individual* $(\alpha/2)100$*-th and* $(1-\alpha/2)100$*-th percentiles, respectively.*

**Joint Inference.** Conducting joint inference results in error bands that must achieve a predefined probability content across multiple structural parameters simultaneously. Unlike previous studies, which apply joint inference across all available variables, horizons, and shocks (e.g., Inoue and Kilian, 2016, 2022) or solely across horizons for each IRF (e.g., Lütkepohl et al., 2015, 2018; Montiel Olea and Plagborg-Møller, 2019), this paper argues that error bands are context-specific objects. The selection of variables included in the joint error bands should be guided by careful consideration of the specific economic question. This is a discretionary choice left to the econometric practitioner, who must similarly make decisions regarding the DGP, prior distribution, and identification strategy. Unfortunately, no definitive metric exists to guide this selection; instead, careful economic reasoning must be applied.

This can be illustrated by considering the evaluation of different pricing theories in international trade (such as dominant currency, producer currency, or local currency pricing), which requires examining the simultaneous response of import and export prices to a foreign monetary policy shock across individual horizons. On the other hand, more frequently, IRF analysis focuses on the persistence of the response, like Jarociński and Karadi (2020), who argue that "real

GDP and the price level both decline persistently by about 10 basis points and 5 basis points, respectively". This example implies the simultaneous evaluation of IRFs across horizons. In practice, any combination of both types of inference can appear. For example, the computation of a fiscal multiplier, often defined as the cumulative sum of the IRF of GDP divided by that of government expenditure, requires simultaneous inference of two variables across several horizons. Furthermore, if the VAR identifies several shocks, such as in Blanchard (1989), then simultaneous inference about a positive demand shock and a favorable supply shock on unemployment and output must be carried out, considering the responses of two variables to two shocks across several horizons. Therefore, depending on the parameters targeted by the error bands, four possible types of joint inference and their corresponding vectors of stacked structural parameters are considered. The definitions that follow are inspired by Akram et al. (2016). First, the construction of error bands that achieve predefined probability content simultaneously for all variables at a specific horizon requires a multivariate (MV) vector of structural parameters.

**Definition 3: MV vector of structural parameters.** *Let* $\boldsymbol{\theta} \in \mathbb{R}^K$ *denotes the MV vector of structural parameters, defined as*

$$\boldsymbol{\theta} = (\theta_{1,s,h}, \theta_{2,s,h}, \ldots, \theta_{n,s,h})' \quad for \quad s = 1, \ldots, S \text{ and } h = 1, \ldots, H.$$

In the same vein, the time-simultaneous (TS) vector of structural parameters is used by a practitioner targeting error bands that achieve predefined probability content simultaneously for all horizons of a given variable.

**Definition 4: TS vector of structural parameters.** *Let* $\boldsymbol{\theta} \in \mathbb{R}^K$ *denotes the TS vector of structural parameters, defined as*

$$\boldsymbol{\theta} = (\theta_{n,s,1}, \theta_{n,s,2}, \ldots, \theta_{n,s,H})' \quad for \quad n = 1, \ldots, N \text{ and } s = 1, \ldots, S.$$

A natural extension of the previous two cases occurs when the error bands must achieve predefined probability content simultaneously for the IRFs at all horizons for every variable reacting to a single shock. This requires a multivariate, time-simultaneous (MV-TS) vector of structural parameters.

**Definition 5: MV-TS vector of structural parameters.** *Let* $\boldsymbol{\theta} \in \mathbb{R}^K$ *denotes the MV-TS vector of structural parameters, defined as*

$$\boldsymbol{\theta} = \text{vec}\begin{pmatrix} \theta_{1,s,1} & \theta_{2,s,1} & \ldots & \theta_{N,s,1} \\ \theta_{1,s,2} & \theta_{2,s,2} & \ldots & \theta_{N,s,2} \\ \vdots & \vdots & & \vdots \\ \theta_{1,s,H} & \theta_{2,s,H} & \ldots & \theta_{N,s,H} \end{pmatrix} \quad for \quad s = 1, \ldots, S.$$

In the most comprehensive case, when the economic question requires error bands that achieve predefined probability content simultaneously at all horizons for all variables reacting to all shocks, it is necessary to construct a multivariate, time-simultaneous and cross-shock (MV-TS-CS) vector of structural parameters.

**Definition 6: MV-TS-CS vector of structural parameters.** *Let* $\boldsymbol{\theta} \in \mathbb{R}^K$ *denote the MV-TS-CS vector of structural parameters, defined as*

$$\boldsymbol{\theta} = \text{vec}\begin{pmatrix} \boldsymbol{\theta}_{1,1} & \boldsymbol{\theta}_{2,1} & \dots & \boldsymbol{\theta}_{N,1} \\ \boldsymbol{\theta}_{1,2} & \boldsymbol{\theta}_{2,2} & \dots & \boldsymbol{\theta}_{N,2} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{\theta}_{1,S} & \boldsymbol{\theta}_{2,S} & \dots & \boldsymbol{\theta}_{N,S} \end{pmatrix},$$

*where* $\boldsymbol{\theta}_{n,s} = (\theta_{n,s,1}, \theta_{n,s,2}, \dots, \theta_{n,s,H})'$.

**Definition 7: Joint credible region.** *The* $(1-\alpha)100\%$ *joint credible region,* $\boldsymbol{\Theta}^{1-\alpha}$*, for any vector of structural parameters,* $\boldsymbol{\theta} \in \mathbb{R}^K$*, is a Cartesian product given by*

$$\boldsymbol{\Theta}^{1-\alpha} = \bigtimes_{j=1}^{K} \left[ \underline{\theta}_j^{1-\alpha}, \bar{\theta}_j^{1-\alpha} \right]$$

*and coverage probability* $1-\alpha$ *such that*

$$Pr\left(\boldsymbol{\theta} \in \boldsymbol{\Theta}^{1-\alpha}\right) = Pr\left(\bigcap_{j=1}^{K} \left( \underline{\theta}_j^{1-\alpha} \leq \theta_j \leq \bar{\theta}_j^{1-\alpha} \right)\right) = 1-\alpha.$$

Bonferroni's (1936) inequality is the standard approach that relates joint probabilities to marginal probabilities. Its application to constructing joint confidence regions for forecasts in VARs is discussed in the seminal textbook by Lütkepohl (2005). In terms of the definitions of the credible interval and region, the inequality implies that

$$Pr\left(\boldsymbol{\theta} \in \boldsymbol{\Theta}^{1-\alpha}\right) \geq 1 - \sum_{j=1}^{K} Pr\left(\theta_j \notin \boldsymbol{\Theta}_j^{1-\alpha}\right).$$

This inequality can be used to construct joint error bands. For example, a 95% TS error band for a variable across 5 horizons requires the marginal probability of each structural parameter to be 0.99, since $5 \cdot (1 - 0.99) = 0.05$. Because the relationship is an inequality, the proposed error band has at least 95% probability content, but it may be more conservative. Additionally, when a large number of structural parameters are of interest, the bounds can become uninformative.

**Definition 8: Bonferroni error band.** *The* $(1-\alpha)100\%$ *Bonferroni error band,* $\widehat{\boldsymbol{\Theta}}^{1-\alpha,B}$*, for the vector of structural parameters,* $\boldsymbol{\theta} \in \mathbb{R}^K$*, is a Cartesian product given by*

$$\widehat{\boldsymbol{\Theta}}^{1-\alpha,B} = \bigtimes_{j=1}^{K} \left[ \underline{\theta}_j^{1-\alpha/K,PW}, \bar{\theta}_j^{1-\alpha/K,PW} \right].$$

Similarly to the Bonferroni's approach, Šidák (1967) proposes another correction method, which assumes that all parameters of interest are mutually independent. This method relies on the following inequality:

$$Pr\left(\boldsymbol{\theta} \in \boldsymbol{\Theta}^{1-\alpha}\right) \geq 1 - \prod_{j=1}^{K} Pr\left(\theta_j \notin \boldsymbol{\Theta}_j^{1-\alpha}\right).$$

This correction approach shares all the drawbacks of the previous method, but it is less conservative for a small number of structural coefficients. The respective credible region is defined similarly.

**Definition 9: Šidák error band.** *The* $(1-\alpha)100\%$ *Šidák error band,* $\widehat{\boldsymbol{\Theta}}^{1-\alpha,S}$*, for any vector of structural parameters,* $\boldsymbol{\theta} \in \mathbb{R}^K$*, is a Cartesian product given by*

$$\widehat{\boldsymbol{\Theta}}^{1-\alpha,S} = \bigtimes_{j=1}^{K} \left[ \underline{\theta}_j^{(1-\alpha)^{1/K},PW}, \bar{\theta}_j^{(1-\alpha)^{1/K},PW} \right].$$

Montiel Olea and Plagborg-Møller (2019) propose constructing joint error bands using the sup-t method, which builds on the previous two approaches. Although their primary analysis is in the frequentist framework, relying on parametric solutions, they also propose a methodology tailored to the Bayesian framework. Specifically, Montiel Olea and Plagborg-Møller (2019) suggest using numerical optimization to solve for the pointwise quantile that achieves the joint nominal coverage level of the posterior draws.[6] The proper definition of the sup-t credible region relies on the notation of the posterior draws of the structural parameters, which, for $D$ draws, is given by $\theta_j^{(1)}, \ldots, \theta_j^{(D)}$. Montiel Olea and Plagborg-Møller (2019) refer to the optimization procedure as calibration, while Lütkepohl et al. (2015) call a similar process adjustment. In this paper, the former terminology is adhered to. Since additional calibration methods will be introduced in the following paragraphs, each will be given a distinct name. The calibration method related to the sup-t will be referred to as pointwise quantile optimization (PQO).

**Definition 10: Sup-t (PQO) error band.** *The* $(1-\alpha)100\%$ *sup-t (PQO) error band,* $\widehat{\boldsymbol{\Theta}}^{1-\alpha,Su}$*, for any vector of structural parameters,* $\boldsymbol{\theta} \in \mathbb{R}^K$*, is is a Cartesian product given by*

$$\widehat{\boldsymbol{\Theta}}^{1-\alpha,Su} = \bigtimes_{j=1}^{K} \left[ \underline{\theta}_j^{1-\widehat{\alpha},PW}, \bar{\theta}_j^{1-\widehat{\alpha},PW} \right],$$

[6] Similar calibration procedures are also discussed in Gafarov et al. (2018) and Kaido et al. (2019).

*where*

$$\widehat{\alpha} = sup\left\{\widehat{\alpha} \in [\alpha/(2K), \alpha/2] \middle| D^{-1}\sum_{d=1}^{D} \mathbb{1}\left(\boldsymbol{\theta}^{(d)} \in \bigtimes_{j=1}^{K}\left[\underline{\theta}_j^{1-\widehat{\alpha},PW}, \bar{\theta}_j^{1-\widehat{\alpha},PW}\right]\right) \geq 1-\alpha\right\}$$

and $\mathbb{1}(\cdot)$ denotes indicator function.

Akram et al. (2016) and Inoue and Kilian (2022) propose two related methods based on vector-valued loss functions to construct joint error bands. In the former study, Akram et al. (2016) evaluate the distance of each posterior draw to a pointwise measure of central tendency, specifically the mean, while in the latter study, Inoue and Kilian (2022) evaluate the average distance of each posterior draw to every other draw. The main results in this paper focus on the proposal by Akram et al. (2016).

**Definition 11: Lowest posterior risk joint credible set.** *The* $(1-\alpha)100\%$ *vector of structural parameters draws,* $\boldsymbol{\theta}^{(d)} \in \mathbb{R}^K$*, which are closest to a specified pointwise central tendency measure,* $\widetilde{\boldsymbol{\theta}} \in \mathbb{R}^K$*, under a vector-valued loss function,* $L\left(\boldsymbol{\theta}^{(d)}, \widetilde{\boldsymbol{\theta}}\right)$ *are members of the lowest posterior risk joint credible set denoted as* $\widehat{\boldsymbol{\Theta}}_L^{1-\alpha}$*.*

**Definition 12: Min-max error band.** *The* $(1-\alpha)100\%$ *closest vector of structural parameters draws,* $\boldsymbol{\theta}^{(d)} \in \mathbb{R}^K$*, to any pointwise central tendency measure,* $\widetilde{\boldsymbol{\theta}} \in \mathbb{R}^K$*, evaluated under a vector-valued loss function, L, is a Cartesian product given by*

$$\widehat{\boldsymbol{\Theta}}^{1-\alpha,M} = \bigtimes_{j=1}^{K}\left[\underline{\theta}_j^{1-\alpha,M}, \bar{\theta}_j^{1-\alpha,M}\right],$$

*where*

$$\underline{\theta}_j^{1-\alpha,M} = min\left\{\theta_j^{1-\alpha,M} \middle| \boldsymbol{\theta}^{1-\alpha,M} \in \widehat{\boldsymbol{\Theta}}_L^{1-\alpha}\right\} \quad and \quad \bar{\theta}_j^{1-\alpha,M} = max\left\{\theta_j^{1-\alpha,M} \middle| \boldsymbol{\theta}^{1-\alpha,M} \in \widehat{\boldsymbol{\Theta}}_L^{1-\alpha}\right\}.$$

The preliminary results indicate that this method of constructing joint error bands may significantly overestimate the nominal credibility level. Inoue and Kilian (2022) are aware of this possibility and suggest the potential use of a calibration technique similar to the proposition of Montiel Olea and Plagborg-Møller (2019). However, their study does not include a simulation exercise or a details on how to implement the calibration step. Two possible routines to optimize the coverage level of the min-max error bands are explored in this paper. The first approach is to optimize the percentage of retained draws of the lowest posterior risk set, ensuring that their effective coverage level is at least equal to the nominal credibility level. This approach is referred to as loss quantile optimization (LQO).

**Definition 13: Min-max (LQO) error band.** *The* $(1-\alpha)100\%$ *min-max (LQO) error band,* $\widehat{\boldsymbol{\Theta}}^{1-\alpha,M(LQO)}$, *for any vector of structural parameters,* $\boldsymbol{\theta} \in \mathbb{R}^K$, *is a Cartesian product given by*

$$\widehat{\boldsymbol{\Theta}}^{1-\alpha,M(LQO)} = \bigtimes_{j=1}^{K} \left[\underline{\theta}_j^{1-\widehat{\alpha},M}, \bar{\theta}_j^{1-\widehat{\alpha},M}\right],$$

*where*

$$\widehat{\alpha} = sup\left\{\widehat{\alpha} \in (\alpha,1) \left| D^{-1}\sum_{d=1}^{D} \mathbb{1}\left(\boldsymbol{\theta}^{(d)} \in \bigtimes_{j=1}^{K} \left[\underline{\theta}_j^{1-\widehat{\alpha},M}, \bar{\theta}_j^{1-\widehat{\alpha},M}\right]\right) \geq 1-\alpha \right.\right\}.$$

The second optimization method applied to the min-max error bands is an iterative procedure proposed by Lütkepohl et al. (2015). After constructing the lowest posterior risk set along with the min-max error band, all draws on the boundary of the hyperrectangular box are discarded until the posterior draws within the new error band achieve the nominal credibility level. This approach is referred to as boundary draw rejection (BDR).

**Definition 14: Min-max (BDR) error band.** *The* $(1-\alpha)100\%$ *min-max (BDR) error band,* $\widehat{\boldsymbol{\Theta}}^{1-\alpha,M(BDR)}$, *for any vector of structural parameters,* $\boldsymbol{\theta} \in \mathbb{R}^K$, *is a Cartesian product given by*

$$\widehat{\boldsymbol{\Theta}}^{1-\alpha,M(BDR)} = \bigtimes_{j=1}^{K} \left[\underline{\theta}_j^{1-\alpha,M(BDR)}, \bar{\theta}_j^{1-\alpha,M(BDR)}\right] \quad s.t.$$

$$D^{-1}\sum_{d=1}^{D} \mathbb{1}\left(\boldsymbol{\theta}^{(d)} \in \bigtimes_{j=1}^{K} \left[\underline{\theta}_j^{1-\alpha,M(BDR)}, \bar{\theta}_j^{1-\alpha,M(BDR)}\right]\right) = 1-\alpha,$$

*where* $\underline{\boldsymbol{\theta}}^{1-\alpha,M(BDR)} \in \check{\boldsymbol{\Theta}}_L^{1-\alpha}$ *and* $\bar{\boldsymbol{\theta}}^{1-\alpha,M(BDR)} \in \check{\boldsymbol{\Theta}}_L^{1-\alpha}$. *The set* $\check{\boldsymbol{\Theta}}_L^{1-\alpha}$ *is obtained as draws which belong to the boundaries of the error bands of the min-max estimator are iteratively discarded from the lowest posterior risk joint credible set,* $\widehat{\boldsymbol{\Theta}}_L^{1-\alpha}$, *until the coverage level condition is satisfied.*

The min-max estimator relies on specification of a vector-valued loss function to calculate the distance of each draw from a measure of a central tendency. The functional forms considered in this study are: absolute, quadratic, angular, and Chebyshev. The first three functions are proposed by Inoue and Kilian (2022), while the last is suggested by Akram et al. (2016). Inoue and Kilian (2022) also propose a Dirac loss function, but it is non-separable, unlike the others, making it more complex to derive the credible region for various models and it has limited practical relevance. Akram et al. (2016) propose computing the deviation of each draw from a single pointwise measure of central tendency, whereas Inoue and Kilian (2022) consider the average distance of each draw from all other draws. Since both approaches yield equivalent error band estimates, and the former is computationally less burdensome, it is adopted in this

paper. The formal definitions of the vector-valued loss functions are provided below.

$$L(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}}) = \begin{cases} \sum_{j=1}^{K} |\theta_j - \widetilde{\theta}_j| & \text{(absolute loss)} \\ \sum_{j=1}^{K} (\theta_j - \widetilde{\theta}_j)^2 & \text{(quadratic loss)} \\ \frac{1}{\pi NS} \sum_{i=1}^{N} \sum_{j=1}^{S} cos^{-1} \left( \frac{\boldsymbol{\theta}_{i,j} \cdot \widetilde{\boldsymbol{\theta}}_{i,j}}{\|\boldsymbol{\theta}_{i,j}\| \|\widetilde{\boldsymbol{\theta}}_{i,j}\|} \right) & \text{(angular loss)} \\ \max_j |\theta_j - \widetilde{\theta}_j| & \text{(Chebyshev loss)} \end{cases}$$

Since the pointwise median is the estimator that solves the real-valued absolute loss function, it is used as a measure of central tendency for the first vector-valued loss function. Similarly, the mean solves the real-valued quadratic loss function, and thus, it is used as the pointwise measure for the second function. The latter two loss functions do not have such pointwise estimators, so the pointwise median is used because it is the most frequently employed measure.

Inoue and Kilian (2022) advocate for the use of the angular loss function because it is independent of the scale of the endogenous variables, which is a highly appealing property in the context of VARs. Similarly, the Chebyshev loss function is applied only to draws that are normalized pointwise, as the maximum distance across different structural parameters is meaningless otherwise. Therefore, this loss function is also scale-invariant, even though this property is imposed in a different way.

### 3.2 Illustrative Example

This section visually compares the proposed joint error bands and the conventional PW error bands for a bivariate normally distributed toy parameter space, which is assumed to represent a posterior distribution given by:

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0.25 \end{pmatrix} \right).$$

The estimates of the error bands are computed for 20,000 draws and plotted in Figure 1. As expected, the PW error bands is the smallest rectangle and is contained within all joint error bands, suggesting that this method provides the least conservative estimates of uncertainty. On the other hand, the Bonferroni, Šidák, and sup-t error bands nearly coincide for both parameters, which is not necessarily the case for parameter space of interest with more parameters. For the sake of simplicity, the analysis considers only the absolute loss function for the class of min-max estimators. The non-calibrated min-max error bands encompass all other error bands, which implies that it is the most conservative estimator. In comparison, the min-max (LQO) error bands are tighter and even coincide with the PW error bands for $\theta_1$. However, the area enclosed within the box of the min-max (BDR) error bands is smaller, making it a more precise

estimator. These conclusions hold true for both the 68% credible region and the 90% credible region, although the error bands for the 90% credible region are more dispersed, reflecting the higher credibility level.

The joint inference methods are formally accessed by comparing the relative difference between the width of the joint error bands and the PW error bands as

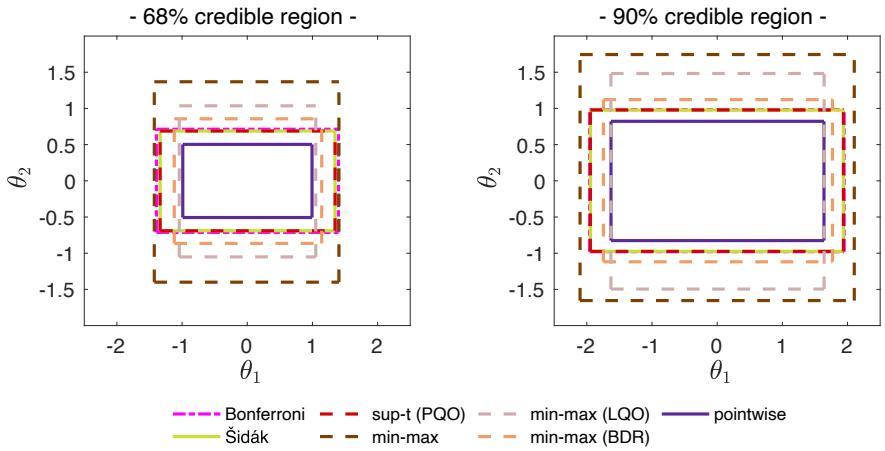
$$\Delta = \left(H\left((\bar{\boldsymbol{\theta}}^{1-\alpha,e} - \underline{\boldsymbol{\theta}}^{1-\alpha,e}) \oslash (\bar{\boldsymbol{\theta}}^{1-\alpha,PW} - \underline{\boldsymbol{\theta}}^{1-\alpha,PW})\right) - 1\right) \cdot 100 \quad \text{for } e \in \{B, S, Su, M, M(LQO), M(BDR)\},$$

where $H(\cdot)$ is the harmonic mean and $\oslash$ denotes the element-wise division. The harmonic mean is selected here as an appropriate measure of the average for ratios.

First, the error bands are computed for a range of covariances in the interval $[-0.45, 0.45]$, which implies correlations in the interval $[-0.9, 0.9]$ for the toy parameter space. The relative width of the error bands against the correlation between the parameters is plotted in Figure 2. The 68% non-calibrated min-max error band is between 70% and 90% wider than the corresponding PW error band, but for the 90% credible interval, this difference falls to the range of 40%-70%. The min-max (LQO) method produces tighter error bands but is strongly dominated by the sup-t and min-max (BDR) methods. The error bands corresponding to the latter two methods nearly coincide being between 18% and 35% wider for the 68% credible region and between 10% and 18% wider for the 90% credible region compared to the PW error bands. The relative width of the min-max error bands and, to a certain extent, the min-max (LQO) error bands does not change smoothly with marginal increases in the correlation among the toy parameters. This is caused by the variability in the error bands of roughly 10% for simulations generated with different seeds of the random number generator. This issue could not be resolved by increasing the number of draws from the posterior distribution. Consequently, the min-max error bands could be less appealing for policy analysis, but the results show that the BDR method does not exhibit the same problem.

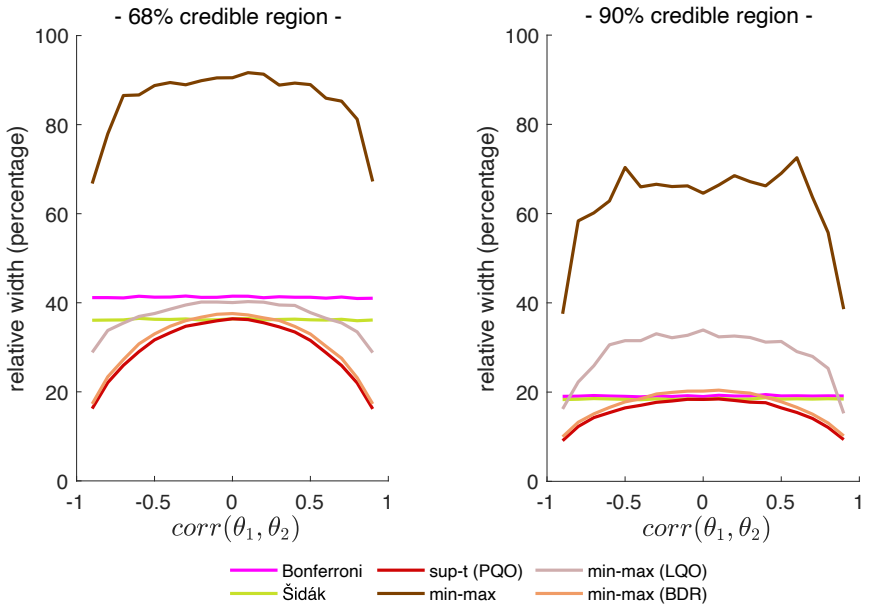
Summing up, the difference between the width of the joint error bands and the PW error bands could be considered significant, but it is not certain that this is an argument to overlook the construction of proper error bands. Except the Bonferroni's and Šidák's error bands, all other joint inference methods produce error bands that shrink with the absolute value of the correlation between the parameters. Since macroeconometric models usually have correlated structural parameters, this strengthens the case for employing joint error bands. Moreover, the loss of precision associated with having joint error bands is smaller for regions with higher credibility.

Figure 1
Pointwise and joint error bands for simulated toy parameters




*Notes: The squares depict PW and MV error bands for simulated toy parameters. The min-max estimators are computed only with absolute loss function. The figure on the left side shows results for 68% credible region and the figure on the right side shows results for 90% credible region.*

Figure 2
Relative width of joint error bands for simulated toy parameters




*Notes: The lines depict relative width of the MV error bands in comparison to the PW error bands for different level of correlation for simulated toy parameters. The vertical axis shows the relative width, while the horizontal axis displays the correlation between $\theta_1$ and $\theta_2$. The min-max class of estimators are computed only with absolute loss function. The figure on the left side shows results for the 68% credible region and the figure on the right side shows results for the 90% credible region.*

### 3.3 Simulation Design

The simulation strategy is designed with three objectives: (i) to evaluate the underestimation of uncertainty when the economic question requires joint inference but the economist reports conventional PW error bands, (ii) to assess the ability of different methods of conducting joint inference to properly account for different types of targeted uncertainty (i.e., MV, TS, MV-TS, and MV-TS-CS error bands), and (iii) to compare the width of the joint error bands.

Recently, Talts et al. (2020) proposed a general procedure called simulation-based calibration[7] for validating the inference of Bayesian estimation algorithms. The error bands are evaluated using the property that the credible regions for parameters retain frequentist coverage on average across the prior if the simulation exercise is specified as in the procedure proposed by Talts et al. (2020) (see also Gelman et al., 1996; Little, 2006; Cook et al., 2006). The simulation-based calibration is a powerful diagnostic tool that can be used to assess whether a software implementation correctly estimates any Bayesian model. As such, it has the potential to eliminate many sources of uncertainty related to coding errors, numerical inaccuracies, or mis-specified algorithms, thereby enhancing the credibility of empirical results.

The procedure begins by drawing model parameters from the prior distribution. Based on these parameter values, the true IRFs are computed, and endogenous variables are simulated. The model is then fitted to the simulated data using $D$ posterior draws, which are saved, and the pointwise and joint error bands are computed. These steps are repeated $C$ times, and the appropriate coverage statistics and average relative width are calculated for each type of targeted uncertainty. The average relative width is calculated as the median of the relative widths—defined as the harmonic mean of the ratio between the joint error band and the PW error band for each structural parameter.

For any credibility level $(1-\alpha)100\%$, the coverage statistics is defined as

$$\Pi = C^{-1} \sum_{c=1}^{C} \mathbb{1}\left( (\boldsymbol{\theta})^{(c)} \in \left[ (\underline{\boldsymbol{\theta}}^{1-\alpha})^{(c)}, (\bar{\boldsymbol{\theta}}^{1-\alpha})^{(c)} \right] \right) \cdot 100,$$

where $\mathbb{1}(\cdot)$ is the indicator function defined over the true structural coefficient, $(\boldsymbol{\theta})^{(c)}$, and the estimated lower, $(\underline{\boldsymbol{\theta}}^{1-\alpha})^{(c)}$, and upper, $(\bar{\boldsymbol{\theta}}^{1-\alpha})^{(c)}$, bounds for each iteration, $c$. If the inference algorithm is correct the coverage statistics tends to the nominal credibility level, $(1-\alpha)100\%$, as the total number of iterations, $C$, increases. Depending on the inference type, which is targeted with the error band the vector of structural parameters will include different elements and there are 5 different types of the coverage statistics: PW, MV, TS, MV-TS, and MV-TS-CS.

[7] Simulation-based calibration refers to the Monte Carlo procedure used in this study, which is different from the calibration procedures used to adjust the nominal coverage of the joint error bands introduced in the previous section.

The simulation strategy requires a well-defined generative model. However, since the DGPs described in the following section are designed to resemble real-world empirical VAR models, their parameter distributions are not known a priori. Therefore, using common prior parametrizations ("non-informative" or "unit root")[8] as initial parameter distributions might cause non-defined or economically implausible models. To solve this problem, the models are first estimated using standard priors and then the the posterior distributions are used as priors in the simulation procedure.[9]

In the empirical literature authors usually report 68% and 95% credible intervals (e.g., Jarociński and Karadi, 2020; Paul, 2020; Arias et al., 2019), but there is no clear cut which one is preferable (see Kilian and Lütkepohl, 2017). The analysis in this paper is conducted for 68% and 90% credible intervals. The latter level is being selected to produce more informative statistics if the coverage is overstated.

Regardless of the large number of iterations, the experiments are still subject to uncertainty; and thus, the standard error of the coverage statistics and the relative width is calculated using bootstrap method.[10] The complete simulation procedure is formally described in Table 1.

Table 1
Steps for simulation-based calibration in Monte Carlo evaluation of coverage and width

set prior distributions $\mathbf{\Sigma} \sim IW_n(\nu_0 \mathbf{S}_0, \nu_0)$ and $vec(\mathbf{A}) \sim \mathcal{N}\left(vec(\bar{\mathbf{A}}_0), \mathbf{\Sigma} \otimes \mathbf{N}_0^{-1}\right)$
**for** $c$ in $C$ **do**
  draw posterior $(\mathbf{\Sigma})^{(c)} \sim IW_n(\nu_T \mathbf{S}_T, \nu_T)$ and $(vec(\mathbf{A}))^{(c)} \sim \mathcal{N}\left(vec(\bar{\mathbf{A}}_T), (\mathbf{\Sigma})^{(c)} \otimes \mathbf{N}_T^{-1}\right)$
  evaluate structural coefficients, $(\theta_{n,s,h})^{(c)} \sim \pi\left(\phi_{n,s,h} | (\mathbf{\Sigma})^{(c)}, (vec(\mathbf{A}))^{(c)}\right)$
  simulate endogenous variables, $(y_t)^{(c)} \sim \pi\left(y_t | (\mathbf{\Sigma})^{(c)}, (vec(\mathbf{A}))^{(c)}\right)$
  **for** $d$ in $D$ **do**
    draw posterior, $(\mathbf{\Sigma})^{(d:c)} \sim IW_n(\nu_{\widetilde{T}} \mathbf{S}_{\widetilde{T}}, \nu_{\widetilde{T}})$ and $(vec(\mathbf{A}))^{(d:c)} \sim \mathcal{N}\left(vec(\bar{\mathbf{A}}_{\widetilde{T}}), (\mathbf{\Sigma})^{(d:c)} \otimes \mathbf{N}_{\widetilde{T}}^{-1}\right)$
    evaluate structural coefficients, $(\theta_{n,s,h})^{(d:c)} \sim \pi\left(\theta_{n,s,h} | (\mathbf{\Sigma})^{(d:c)}, (vec(\mathbf{A}))^{(d:c)}\right)$
  **end for**
  compute the pointwise and joint error bands using $(\theta_{n,s,h})^{(d:c)}$
**end for**
compute the coverage statistics, $\Pi$, and average relative width, $\Delta$.

*Notes: The table outlines the steps for conducting simulation-based calibration as a tool to evaluate coverage statistics and average relative width in a Monte Carlo experiment.*

[8] Detailed definitions of the prior distributions are provided in Appendix A.
[9] This strategy is proposed for a similar experiment by Gabry et al. (2019) providing an example from natural sciences where simulated air pollution data based on loose priors has extreme density.
[10] See Koehler et al. (2009) for a discussion on the uncertainty reporting for Monte Carlo experiments.

# 4 Experiments

## 4.1 Data Generating Processes

This subsection introduces the specifics of the DGPs used in the simulation exercises. In short those are: i) the quarterly fiscal model of Ramey (2019) identified with short-run exclusion restrictions; ii) the quarterly macro model of Peersman (2005) identified with sign restrictions; iii) the monthly monetary model of Jarociński and Karadi (2020) identified with an instrumental variable; and iv) the quarterly macro model of Blanchard and Quah (1989) identified with long-run exclusion restrictions.

**DGP1.** The first DGP follows the fiscal model of Ramey (2019). The VAR model has three endogenous variables: government spending, GDP and tax revenue with all variables in real per capita terms specified in logs. The reduced-form model is estimated initially with "non-informative" prior on quarterly data (from 1939:Q1 to 2015:Q4) with a constant and 4 lags of the endogenous variables. Only government spending shock is identified through recursive identification scheme based on the Cholesky decomposition. The proposed variables order implies that the government spending is affected contemporaneously only by its own structural shock. Figure C.1 in Appendix C shows the IRFs of the three endogenous variables for the simulated realizations of the DGP. A positive shock to government spending increases all endogenous variables on impact, but the effect peaks between the 5th and 12th quarter and by the end of the last observed 20th quarter is already negligible. For DGP1, the experiment is repeated 5000 times on a sample of 200 observations and an impulse response horizon of 20 quarters, with each repetition based on 15000 posterior draws.

**DGP2.** The second DGP resembles the empirical model of Peersman (2005). The VAR model has three endogenous variables: GDP, deflator and 3-month treasury bill yield with the first two variables in real per capita terms specified in logs and the third in percentage points. The reduced-form model is estimated initially with "non-informative" prior on quarterly data (from 1950:Q1 to 2006:Q4) with a constant and 4 lags of the endogenous variables. The sample period of DGP2 is shorter compared to DGP1 since it excludes periods with non-responsive monetary policy, and consequently constant yield. Unlike Peersman (2005), I identify only a supply shock due to computational considerations. The supply shock is identified with sign restrictions and is assumed to raise the interest rate on impact (holding only one period), reduce output for 4 quarters and increase prices for 4 quarters. Figure C.2 in Appendix C shows the IRFs of the three endogenous variables for the simulated realizations of the DGP. In most of the simulated models, the output hits the trough 4 to 8 quarters after impact, while the deflator is experiencing prolonged effects. The interest rate reacts to the economic developments and remains elevated for significant number of periods. For DGP2, the experiment is repeated 5,000 times on a sample

of 200 observations. The impulse response horizon consists of 20 quarters, with each repetition based on 150 posterior draws and 150 rotations of the impact matrix.

**DGP3.** The third DGP considered in this study mimics the monetary model of Jarociński and Karadi (2020). The VAR model has four endogenous variables: one year government bond yield, real GDP, price deflator and excess bond premium. The yield and the excess bond premium are expressed in percentage points and the other two variables in logarithm. The reduced-form model is estimated initially with "unit root" prior on monthly data (from 1990:M2 to 2016:M12) with a constant and 6 lags of the endogenous variables. Only monetary shock is identified through the instrumental variable denoted as poor man's monetary shock by Jarociński and Karadi (2020). It is calculated as the size of the interest rate change around Federal Open Market Committee meetings when stock prices and interest rates move in opposite directions. Figure C.3 in Appendix C shows the IRFs of the four endogenous variables for the simulated realizations of the DGP. For DGP3, the experiment is repeated 5000 times on a sample of 270 observations and an impulse response horizon of 36 months, with each repetition based on 15000 posterior draws.

**DGP4.** The last DGP follows the macro model of Blanchard and Quah (1989). The VAR model has two endogenous variables: growth rate of gross national product (GNP) and unemployment rate; and thus the vector of endogenous variables is potentially stationary. The reduced-form model is estimated initially with "non-informative" priors on quarterly data (from 1948:Q2 to 1987:Q4) with a constant and 4 lags of the endogenous variables. In contrast to the previous DGPs, two structural shocks, namely supply and demand, are identified imposing long-run restrictions. In particular, it is assumed that neither shock has long-run effect on unemployment and also the only supply shock has long-run effect on output. Figure C.4 in Appendix C shows the IRFs of the two endogenous variables in response to both structural shocks for the simulated realizations of the DGP. Structural demand shocks have a hump-shaped effect on GNP and unemployment with the effect peaking by the end of the first year and quickly vanishing by end of the fourth year. The supply shock has an even shorter lived effect on unemployment, but GNP remains elevated in the long run with maximum achieved by the end of the third year after the impact. For DGP4, the experiment is repeated 5000 times on a sample of 200 observations and an impulse response horizon of 20 quarters, with each repetition based on 15000 posterior draws.

## 4.2 Coverage Level and Relative Width

The estimated coverage levels along with their standard deviations of the PW error bands of DGP1-4 are given in Table 2. The first row shows the PW coverage level, i.e., the fraction of individual parameters lying within their respective credible intervals. As expected, the conventional pointwise inference achieves the nominal coverage level for the 68% and 90% credible intervals, implying that the Monte Carlo design is properly specified and the estimation

procedure is correctly implemented. The coverage levels presented in the rows below correspond to the setting in which joint error bands are required for the economic question at hand, but the applied econometrician instead naïvely reports conventional PW error bands. Even in the mildest case—where only multivariate dependence is ignored—the corresponding coverage level, $\Pi^{MV}$, of the PW error bands ranges from 26% to 57% for the 68% error band, and from 70% to 85% for the 90% error band. The subsequent rows indicate that as the number of ignored structural parameters in the inference procedure increases, the degree of uncertainty underestimation associated with PW error bands rises substantially. For example, DGP1 has the worst TS coverage level, $\Pi^{TS}$, which is 21% and 64% for the 68% and 90% error bands, respectively. In the most severe case, where MV-TS-CS dependence is ignored, the estimated coverage levels for DGP4 are only 5% for the 68% credibility level and 39% for the 90% level.[11]

To conclude, the PW error band is an appropriate choice for conducting inference about IRFs only when the economic question does not involve a comparison of multiple structural parameters simultaneously. If this is not the case, the uncertainty is significantly underestimated. The loss associated with reporting incorrectly specified error bands is larger for the 68% credible region compared to the 90% credible region.

The results of the experiments on joint inference methods across all DGPs are presented in Tables 3 to 6. Each table reports the estimated coverage levels and average relative widths, along with their standard deviations, under four different assumptions regarding the targeted mutual dependence between the structural parameters: MV, TS, MV-TS, and MV-TS-CS.

[11] DGP1-3 have only one identified shock; thus, the MV-TS-CS coverage level is not defined.

Table 2
Coverage levels of PW error bands

| Coverage level | 68% PW error band | | | | 90% PW error band | | | |
|---|---|---|---|---|---|---|---|---|
| | DGP1 | DGP2 | DGP3 | DGP4 | DGP1 | DGP2 | DGP3 | DGP4 |
| $\Pi^{PW}$ | 68.07 | 67.25 | 67.63 | 67.70 | 90.17 | 89.47 | 89.94 | 89.71 |
| | (0.34) | (0.29) | (0.34) | (0.32) | (0.22) | (0.19) | (0.22) | (0.21) |
| $\Pi^{MV}$ | 44.64 | 36.18 | 26.05 | 57.06 | 80.00 | 74.92 | 69.75 | 85.31 |
| | (0.42) | (0.48) | (0.46) | (0.37) | (0.36) | (0.41) | (0.54) | (0.27) |
| $\Pi^{TS}$ | 21.23 | 43.83 | 31.65 | 26.34 | 63.98 | 75.09 | 71.82 | 66.41 |
| | (0.40) | (0.41) | (0.40) | (0.42) | (0.48) | (0.35) | (0.42) | (0.48) |
| $\Pi^{MV-TS}$ | 2.90 | 17.14 | 3.24 | 16.43 | 36.12 | 51.98 | 37.70 | 56.43 |
| | (0.25) | (0.44) | (0.24) | (0.37) | (0.67) | (0.64) | (0.72) | (0.54) |
| $\Pi^{MV-TS-CS}$ | - | - | - | 4.68 | - | - | - | 38.86 |
| | | | | (0.28) | | | | (0.65) |

*Notes: Pointwise ($\Pi^{PW}$), multivariate ($\Pi^{MV}$), time-simultaneous ($\Pi^{TS}$), multivariate, time-simultaneous ($\Pi^{MV-TS}$), and multivariate, time-simultaneous, cross-shock ($\Pi^{MV-TS-CS}$) coverage levels of 68% and 90% PW error bands for DGP1, DGP2, DGP3 and DGP4. Standard deviations computed with bootstrap method are in parentheses.*

First, the results for the MV error bands are presented in Table 3. As anticipated, the Bonferroni and Šidák error bands rank among the most conservative approaches to measuring joint uncertainty. This conservatism is particularly pronounced in the case of the Bonferroni method, while the Šidák method achieves the nominal 90% coverage level for DGP2 and DGP3. Conversely, the sup-t error bands method meets the nominal rate for all DGPs and both credibility levels, indicating that it effectively summarizes multivariate estimation uncertainty. The average relative width of the sup-t error bands ranges from 19% to 65% for the 68% credible region and from 11% to 35% for the 90% credible region. The results indicate that MV error bands exhibit only a modest increase in width relative to conventional PW error bands, implying that the cost of employing them is limited. Additionally, a comparison of the MV error bands for DGP4 in the 90% credibility region reveals that Šidák error bands have a 92% effective coverage level and the sup-t error band has a 90% effective coverage level, with average relative widths of 21% and 11%, respectively. This demonstrates that properly constructed error bands achieving the nominal coverage level can lead to more precise uncertainty estimation compared to ad-hoc methods.

Table 3
Multivariate coverage levels ($\Pi^{MV}$) and average relative widths ($\Delta^{MV}$) of MV error bands

| Method | 68% MV error band | | | | | | | | 90% MV error band | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | DGP1 | | DGP2 | | DGP3 | | DGP4 | | DGP1 | | DGP2 | | DGP3 | | DGP4 | |
| | $\Pi$ | $\Delta$ | $\Pi$ | $\Delta$ | $\Pi$ | $\Delta$ | $\Pi$ | $\Delta$ | $\Pi$ | $\Delta$ | $\Pi$ | $\Delta$ | $\Pi$ | $\Delta$ | $\Pi$ | $\Delta$ |
| Bonferroni | 78.72 (0.37) | 63.39 (0.01) | 73.49 (0.41) | 43.36 (0.01) | 75.07 (0.51) | 79.64 (0.01) | 77.30 (0.33) | 44.64 (0.01) | 93.18 (0.23) | 30.40 (0.01) | 90.49 (0.26) | 18.22 (0.01) | 91.17 (0.32) | 39.26 (0.01) | 92.47 (0.19) | 21.90 (0.01) |
| Šidák | 76.14 (0.39) | 57.05 (0.01) | 70.51 (0.43) | 39.62 (0.01) | 71.88 (0.53) | 72.57 (0.01) | 75.26 (0.35) | 39.08 (0.01) | 92.94 (0.24) | 29.51 (0.01) | 90.15 (0.26) | 17.73 (0.01) | 90.88 (0.32) | 38.23 (0.01) | 92.31 (0.19) | 21.11 (0.01) |
| sup-t (PQO) | 67.78 (0.44) | 38.41 (0.01) | 67.18 (0.44) | 35.61 (0.01) | 68.18 (0.55) | 65.06 (0.02) | 67.75 (0.37) | 18.89 (0.01) | 90.36 (0.27) | 21.08 (0.01) | 89.17 (0.28) | 16.34 (0.01) | 89.69 (0.34) | 34.82 (0.01) | 89.73 (0.22) | 11.16 (0.01) |
| min-max: absolute loss (median) | 88.25 (0.28) | 102.33 (0.03) | 89.82 (0.28) | 81.62 (0.04) | 88.62 (0.36) | 159.88 (0.06) | 76.03 (0.28) | 47.25 (0.02) | 98.00 (0.13) | 64.33 (0.03) | 98.17 (0.13) | 44.07 (0.04) | 97.38 (0.18) | 95.38 (0.04) | 93.08 (0.18) | 32.88 (0.01) |
| min-max (LQO): absolute loss (median) | 68.27 (0.40) | 47.26 (0.02) | 67.28 (0.49) | 41.63 (0.02) | 67.20 (0.55) | 85.92 (0.04) | 67.74 (0.31) | 27.17 (0.01) | 90.33 (0.25) | 27.13 (0.01) | 89.49 (0.28) | 24.49 (0.02) | 89.52 (0.34) | 57.02 (0.04) | 89.65 (0.21) | 19.66 (0.02) |
| min-max (BDR): absolute loss (median) | 67.85 (0.44) | 38.97 (0.01) | 66.86 (0.48) | 35.05 (0.02) | 67.80 (0.58) | 67.63 (0.02) | 67.66 (0.34) | 20.63 (0.01) | 90.39 (0.27) | 21.14 (0.01) | 89.24 (0.31) | 16.95 (0.01) | 89.56 (0.35) | 37.19 (0.01) | 89.62 (0.22) | 13.38 (0.01) |
| min-max: quadratic loss (mean) | 75.69 (0.38) | 72.05 (0.02) | 77.64 (0.44) | 61.13 (0.02) | 74.64 (0.54) | 117.61 (0.03) | 70.20 (0.30) | 37.70 (0.02) | 93.66 (0.20) | 47.52 (0.02) | 95.04 (0.20) | 36.19 (0.02) | 92.33 (0.31) | 77.47 (0.03) | 90.75 (0.20) | 26.64 (0.02) |
| min-max (LQO): quadratic loss (mean) | 68.30 (0.39) | 54.17 (0.02) | 66.91 (0.50) | 43.32 (0.02) | 67.13 (0.58) | 94.14 (0.04) | 67.66 (0.30) | 31.49 (0.02) | 90.24 (0.25) | 35.22 (0.02) | 89.46 (0.29) | 25.88 (0.02) | 89.41 (0.36) | 66.68 (0.04) | 89.63 (0.21) | 22.79 (0.02) |
| min-max (BDR): quadratic loss (mean) | 68.19 (0.42) | 43.66 (0.02) | 66.78 (0.50) | 36.41 (0.02) | 67.29 (0.57) | 75.89 (0.03) | 67.65 (0.31) | 24.95 (0.01) | 90.24 (0.26) | 24.74 (0.01) | 89.24 (0.30) | 18.25 (0.01) | 89.54 (0.35) | 45.65 (0.03) | 89.60 (0.21) | 16.96 (0.02) |
| min-max: angular loss (median) | 99.93 (0.02) | 310.81 (0.15) | 99.87 (0.03) | 132.56 (0.07) | 99.93 (0.03) | 335.78 (0.18) | 99.97 (0.01) | 376.96 (0.25) | 99.95 (0.01) | 151.50 (0.08) | 99.92 (0.03) | 63.35 (0.05) | 99.95 (0.02) | 165.21 (0.10) | 99.97 (0.01) | 187.07 (0.14) |
| min-max (LQO): angular loss (median) | 97.08 (0.14) | 171.06 (0.21) | 86.99 (0.30) | 73.62 (0.12) | 95.00 (0.23) | 179.40 (0.24) | 97.87 (0.09) | 195.08 (0.29) | 97.08 (0.14) | 62.67 (0.13) | 91.86 (0.24) | 26.69 (0.06) | 95.02 (0.23) | 66.27 (0.14) | 97.87 (0.09) | 72.58 (0.17) |
| min-max (BDR): angular loss (median) | 67.65 (0.44) | 37.83 (0.01) | 67.18 (0.46) | 34.92 (0.01) | 68.00 (0.56) | 64.17 (0.02) | 67.72 (0.37) | 18.83 (0.01) | 90.35 (0.27) | 20.84 (0.01) | 89.14 (0.28) | 16.22 (0.01) | 89.67 (0.34) | 34.61 (0.01) | 89.71 (0.22) | 11.09 (0.01) |
| min-max: Chebyshev loss (median) | 67.70 (0.44) | 37.91 (0.01) | 67.43 (0.46) | 35.23 (0.02) | 67.96 (0.56) | 63.93 (0.02) | 67.64 (0.37) | 18.23 (0.01) | 90.50 (0.28) | 20.81 (0.01) | 88.99 (0.30) | 17.17 (0.01) | 89.67 (0.33) | 34.83 (0.01) | 89.66 (0.25) | 12.02 (0.01) |
| min-max (LQO): Chebyshev loss (median) | 67.70 (0.44) | 37.91 (0.01) | 67.42 (0.46) | 35.23 (0.02) | 67.96 (0.56) | 63.93 (0.02) | 67.64 (0.37) | 18.23 (0.01) | 90.50 (0.28) | 20.81 (0.01) | 88.99 (0.30) | 17.17 (0.01) | 89.67 (0.33) | 34.83 (0.01) | 89.66 (0.25) | 12.02 (0.01) |
| min-max (BDR): Chebyshev loss (median) | 67.70 (0.44) | 37.91 (0.01) | 67.43 (0.46) | 35.23 (0.02) | 67.96 (0.56) | 63.93 (0.02) | 67.64 (0.37) | 18.23 (0.01) | 90.50 (0.28) | 20.81 (0.01) | 88.99 (0.30) | 17.17 (0.01) | 89.67 (0.33) | 34.83 (0.01) | 89.66 (0.25) | 12.02 (0.01) |

*Notes: Multivariate coverage levels ($\Pi^{MV}$) and average relative widths ($\Delta^{MV}$) of 68% and 90% MV error bands for DGP1, DGP2, DGP3 and DGP4, labeled $\Pi$ and $\Delta$, respectively, for simplicity. Standard deviations computed with bootstrap method are in parentheses. The 'median' or 'mean' indicated in parentheses next to each min-max estimator denotes the pointwise measure of central tendency used for the corresponding loss function.*

Next, an examination of the MV error bands constructed using the min-max methodology reveals that the non-calibrated bands are conservative, often yielding coverage levels that exceed the nominal credibility level for the absolute, quadratic, and angular loss functions. Moreover, notable differences emerge in the performance of error bands across these loss functions. In particular, the angular loss function underperforms relative to the absolute loss function, which is itself clearly dominated by the quadratic loss function. This finding is somewhat surprising, as Inoue and Kilian (2022) report no significant differences among the loss functions in their empirical applications. One possible reason for this discrepancy is that their specifications consider MV-TS-CS uncertainty, which increases the dimensionality of the problem and diminishes the importance of the selected loss function. This will be evaluated later in this section when analyzing the simulation results in Table 6. Additionally, it is worth noting that, despite Inoue and Kilian (2022) proposing the angular loss function to address potential issues with evaluating vector-valued distances of structural parameters with different scales, it actually computes the most conservative error bands. Finally, among the non-calibrated min-max error bands, the Chebyshev loss function consistently achieves the nominal coverage level, with a relative width economically indistinguishable from that of the sup-t error bands.

Having examined the non-calibrated min-max error bands, I now shift the focus to other extended routines that implement an intermediate calibration step (i.e., LQO or BDR). Notably, the Chebyshev loss function achieves the nominal coverage level without implementing LQO or BDR, and the performance of the error bands in the corresponding rows aligns with the min-max error bands as the calibration steps are not initiated. While LQO calibration enables the absolute and quadratic loss functions to attain the nominal coverage level, it is ineffective in the case of the angular loss function. However, comparing the average relative width of the error bands for the former two loss functions, the results indicate that these error bands are wider compared to the sup-t estimations, with differences reaching up to 20 percentage points for some DGPs. On the other hand, the error bands with BDR calibration not only achieve the nominal coverage level but also have widths for the absolute and angular loss functions similar to those of the sup-t error bands. The BDR calibration applied to the min-max error bands with the quadratic loss function is slightly larger than the sup-t error bands for some DGPs. Overall, the sup-t, min-max with Chebyshev loss and min-max (BDR) with absolute and angular loss are good estimators of MV uncertainty with very similar performance regarding the estimated coverage level and average relative width.

The results of the experiments regarding the TS error bands are given in Table 4. The increase of the number of elements of the structural vector of interest, changes the relative performance of the joint inference methods. The Bonferroni and Šidák error bands emerge as the most conservative methods, substantially overestimating the uncertainty surrounding the IRFs. In this class of methods, only the sup-t error bands attain the nominal coverage level. The difference in average relative width between the TS and MV error bands estimated using the sup-t method

is substantial for DGP1 and DGP4, but negligible for DGP2 and DGP3. A possible explanation may lie in the different identification methodology. Similarly to the reported results in the previous table, the TS error bands with the min-max error bands with Chebyshev loss function perform the same as the sup-t error bands even without a calibration step. However, this does not extend to the other loss functions in this class of joint inference methods. In the absence of calibration, the angular loss function yields more conservative bands than the absolute loss, which itself is outperformed by the quadratic loss. The LQO calibration method enables all error bands to attain the nominal coverage level, while yielding slightly narrower average relative widths for the absolute and quadratic loss functions. In contrast, the LQO-calibrated min-max error bands with the angular loss function are significantly wider than those produced by the sup-t method. The BDR calibration method also ensures that the min-max estimator achieves nominal coverage, but it further reduces the width of the error bands. Under BDR calibration, the angular loss function yields error bands that are similar to—or slightly wider than—those of the sup-t method. In comparison, the absolute and quadratic loss functions already offer substantial improvements over the sup-t bands. For instance, for the 68% credible region, the relative width of the min-max (BDR) error bands with the absolute loss function is between 7 and 14 percentage points lower than that of the sup-t error bands.

Table 5 presents the result for analysis of MV-TS joint inference methods. An increase in the number of structural parameters of interest does not lead to differences in the relative performance of the various inference methods with respect to coverage levels, as the MV-TS and TS error bands yield comparable results. It is worth noting, however, that the increase of the average relative width does not entirely depends on the number of structural parameters, but also on the specifics of the DGPs. In particular the estimate of the 68% TS error band of the min-max (BDR) method under absolute loss function for DGP1 and DGP2 is 60% and 24%. Even though the increase of the number of structural parameters for both DGPs in the case of MV-TS inference is the same, the average relative width changes to 95% and 56%. The most pronounced increase in average relative width from the TS to the MV-TS error bands is observed for DGP3, where it rises from 45% to 104%. This is consistent with expectations, as DGP3 involves the largest expansion in the number of structural parameters between the two specifications. In contrast, DGP4 exhibits the smallest increase in width between the TS and MV-TS error bands. The comparison between the results of the TS and MV-TS error bands show that an economist interested in joint inference method could observe different width among models depending on the economic question and the corresponding inference methods, which is selected.

Table 4
Time-simultaneous coverage levels ($\Pi^{TS}$) and average relative widths ($\Delta^{TS}$) of TS error bands

| Method | 68% TS error band | | | | | | | | 90% TS error band | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | DGP1 | | DGP2 | | DGP3 | | DGP4 | | DGP1 | | DGP2 | | DGP3 | | DGP4 | |
| | $\Pi$ | $\Delta$ | $\Pi$ | $\Delta$ | $\Pi$ | $\Delta$ | $\Pi$ | $\Delta$ | $\Pi$ | $\Delta$ | $\Pi$ | $\Delta$ | $\Pi$ | $\Delta$ | $\Pi$ | $\Delta$ |
| Bonferroni | 93.03 (0.25) | 147.35 (0.03) | 94.03 (0.19) | 84.63 (0.03) | 96.37 (0.16) | 177.70 (0.03) | 92.79 (0.24) | 165.40 (0.06) | 97.83 (0.15) | 74.47 (0.02) | 97.79 (0.12) | 39.61 (0.02) | 98.77 (0.11) | 92.25 (0.02) | 97.36 (0.14) | 86.47 (0.03) |
| Šidák | 91.77 (0.29) | 140.32 (0.02) | 93.13 (0.21) | 81.64 (0.03) | 95.73 (0.18) | 170.31 (0.03) | 91.59 (0.25) | 156.98 (0.05) | 97.73 (0.15) | 73.42 (0.02) | 97.70 (0.12) | 39.14 (0.02) | 98.74 (0.11) | 91.10 (0.02) | 97.23 (0.15) | 85.17 (0.03) |
| sup-t (PQO) | 68.07 (0.48) | 74.60 (0.01) | 66.60 (0.43) | 31.71 (0.02) | 67.81 (0.46) | 59.47 (0.03) | 67.74 (0.46) | 72.10 (0.07) | 90.58 (0.31) | 40.35 (0.01) | 88.80 (0.28) | 17.58 (0.01) | 89.67 (0.28) | 33.83 (0.01) | 89.56 (0.28) | 41.20 (0.04) |
| min-max: absolute loss (median) | 78.13 (0.45) | 94.48 (0.02) | 77.13 (0.31) | 46.62 (0.03) | 77.06 (0.44) | 79.09 (0.05) | 81.91 (0.38) | 118.89 (0.10) | 94.41 (0.26) | 60.28 (0.01) | 93.17 (0.20) | 25.88 (0.02) | 93.53 (0.24) | 50.74 (0.03) | 94.82 (0.25) | 72.02 (0.06) |
| min-max (LQO): absolute loss (median) | 68.23 (0.54) | 68.00 (0.01) | 67.12 (0.36) | 29.94 (0.02) | 67.47 (0.48) | 53.72 (0.03) | 67.70 (0.46) | 75.90 (0.09) | 90.18 (0.31) | 44.28 (0.01) | 89.56 (0.26) | 18.00 (0.02) | 89.72 (0.29) | 36.28 (0.02) | 89.44 (0.32) | 47.75 (0.05) |
| min-max (BDR): absolute loss (median) | 67.91 (0.52) | 60.27 (0.01) | 66.74 (0.37) | 24.13 (0.01) | 67.16 (0.47) | 45.01 (0.03) | 67.41 (0.48) | 64.55 (0.08) | 90.28 (0.31) | 35.62 (0.01) | 89.23 (0.25) | 13.39 (0.01) | 89.53 (0.29) | 27.64 (0.01) | 89.30 (0.32) | 37.43 (0.04) |
| min-max: quadratic loss (mean) | 73.80 (0.48) | 79.87 (0.02) | 74.02 (0.33) | 39.25 (0.02) | 73.58 (0.46) | 66.98 (0.04) | 76.19 (0.40) | 107.11 (0.10) | 92.53 (0.28) | 50.33 (0.02) | 92.36 (0.22) | 22.48 (0.02) | 92.28 (0.27) | 44.17 (0.02) | 92.80 (0.25) | 65.87 (0.06) |
| min-max (LQO): quadratic loss (mean) | 68.19 (0.53) | 66.13 (0.02) | 66.80 (0.38) | 27.93 (0.02) | 67.39 (0.50) | 51.75 (0.02) | 67.44 (0.48) | 83.69 (0.10) | 90.19 (0.32) | 42.38 (0.02) | 89.42 (0.27) | 16.43 (0.01) | 89.66 (0.29) | 35.34 (0.02) | 89.19 (0.31) | 52.96 (0.06) |
| min-max (BDR): quadratic loss (mean) | 67.93 (0.53) | 60.81 (0.02) | 66.51 (0.37) | 24.40 (0.01) | 67.25 (0.48) | 46.14 (0.02) | 67.41 (0.49) | 72.14 (0.09) | 90.03 (0.32) | 36.45 (0.01) | 89.25 (0.26) | 13.19 (0.01) | 89.46 (0.29) | 28.82 (0.01) | 89.28 (0.31) | 42.41 (0.05) |
| min-max: angular loss (median) | 78.27 (0.45) | 142.13 (0.09) | 80.11 (0.27) | 84.49 (0.07) | 88.28 (0.25) | 185.52 (0.10) | 78.55 (0.37) | 135.78 (0.11) | 92.99 (0.27) | 86.07 (0.05) | 91.49 (0.19) | 38.62 (0.05) | 95.86 (0.15) | 110.96 (0.07) | 93.07 (0.22) | 91.16 (0.08) |
| min-max (LQO): angular loss (median) | 68.93 (0.52) | 113.02 (0.05) | 67.19 (0.33) | 50.90 (0.05) | 69.32 (0.38) | 109.68 (0.12) | 67.58 (0.48) | 100.42 (0.13) | 90.14 (0.34) | 75.31 (0.04) | 89.29 (0.22) | 36.00 (0.05) | 90.00 (0.23) | 79.37 (0.10) | 89.39 (0.25) | 73.33 (0.07) |
| min-max (BDR): angular loss (median) | 68.39 (0.52) | 75.09 (0.03) | 66.86 (0.36) | 32.21 (0.02) | 67.28 (0.41) | 60.19 (0.03) | 67.21 (0.48) | 75.01 (0.09) | 90.12 (0.32) | 47.43 (0.02) | 89.33 (0.22) | 24.65 (0.03) | 89.39 (0.27) | 42.00 (0.05) | 89.19 (0.28) | 51.20 (0.05) |
| min-max: Chebyshev loss (median) | 67.64 (0.47) | 73.13 (0.01) | 66.53 (0.43) | 31.41 (0.02) | 67.40 (0.45) | 58.35 (0.03) | 67.39 (0.47) | 70.33 (0.06) | 90.40 (0.34) | 39.62 (0.01) | 88.81 (0.27) | 15.26 (0.01) | 89.61 (0.27) | 33.35 (0.02) | 89.38 (0.30) | 39.98 (0.03) |
| min-max (LQO): Chebyshev loss (median) | 67.64 (0.47) | 73.13 (0.01) | 66.53 (0.43) | 31.40 (0.02) | 67.40 (0.45) | 58.35 (0.03) | 67.39 (0.47) | 70.33 (0.06) | 90.40 (0.34) | 39.62 (0.01) | 88.81 (0.27) | 15.26 (0.01) | 89.61 (0.27) | 33.35 (0.02) | 89.38 (0.30) | 39.98 (0.03) |
| min-max (BDR): Chebyshev loss (median) | 67.64 (0.47) | 73.13 (0.01) | 66.53 (0.43) | 31.41 (0.02) | 67.40 (0.45) | 58.35 (0.03) | 67.39 (0.47) | 70.33 (0.06) | 90.40 (0.34) | 39.62 (0.01) | 88.81 (0.27) | 15.26 (0.01) | 89.61 (0.27) | 33.35 (0.02) | 89.38 (0.30) | 39.98 (0.03) |

*Notes: Time-simultaneous coverage levels ($\Pi^{TS}$) and average relative widths ($\Delta^{TS}$) of 68% and 90% TS error bands for DGP1, DGP2, DGP3 and DGP4, labeled $\Pi$ and $\Delta$, respectively, for simplicity. Standard deviations computed with bootstrap method are in parentheses. The 'median' or 'mean' indicated in parentheses next to each min-max estimator denotes the pointwise measure of central tendency used for the corresponding loss function.*

Table 5
Multivariate, time-simultaneous coverage levels ($\Pi^{MV-TS}$) and average relative widths ($\Delta^{MV-TS}$) of MV-TS error bands

| Method | 68% MV-TS error band | | | | | | | | 90% MV-TS error band | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | DGP1 | | DGP2 | | DGP3 | | DGP4 | | DGP1 | | DGP2 | | DGP3 | | DGP4 | |
| | Π | Δ | Π | Δ | Π | Δ | Π | Δ | Π | Δ | Π | Δ | Π | Δ | Π | Δ |
| Bonferroni | 94.56 (0.33) | 188.45 (0.03) | 93.82 (0.29) | 100.97 (0.04) | 96.80 (0.26) | 231.49 (0.04) | 94.32 (0.25) | 197.52 (0.07) | 97.90 (0.21) | 97.03 (0.02) | 97.44 (0.22) | 48.97 (0.03) | 98.68 (0.14) | 122.56 (0.03) | 97.94 (0.16) | 104.43 (0.04) |
| Šidák | 93.18 (0.36) | 181.85 (0.03) | 93.10 (0.30) | 98.46 (0.04) | 96.10 (0.29) | 224.54 (0.04) | 93.44 (0.26) | 189.22 (0.07) | 97.86 (0.21) | 96.03 (0.02) | 97.34 (0.22) | 48.58 (0.03) | 98.66 (0.15) | 121.38 (0.03) | 97.83 (0.16) | 103.13 (0.04) |
| sup-t (PQO) | 68.70 (0.70) | 113.68 (0.02) | 66.42 (0.60) | 60.73 (0.02) | 67.10 (0.68) | 117.81 (0.03) | 67.80 (0.50) | 92.41 (0.07) | 90.40 (0.42) | 60.75 (0.02) | 88.02 (0.42) | 30.32 (0.01) | 88.28 (0.41) | 64.54 (0.02) | 89.44 (0.31) | 52.35 (0.04) |
| min-max: absolute loss (median) | 87.14 (0.50) | 155.65 (0.05) | 84.06 (0.45) | 88.39 (0.06) | 81.24 (0.53) | 165.08 (0.07) | 79.76 (0.41) | 128.67 (0.10) | 96.54 (0.28) | 87.67 (0.03) | 95.68 (0.27) | 46.13 (0.04) | 94.32 (0.35) | 97.53 (0.06) | 93.96 (0.29) | 77.69 (0.06) |
| min-max (LQO): absolute loss (median) | 67.58 (0.66) | 106.37 (0.03) | 66.82 (0.60) | 63.64 (0.03) | 66.94 (0.74) | 121.28 (0.06) | 67.80 (0.52) | 92.76 (0.09) | 90.26 (0.47) | 61.85 (0.02) | 88.76 (0.38) | 35.69 (0.03) | 89.76 (0.44) | 76.83 (0.04) | 89.34 (0.33) | 57.80 (0.05) |
| min-max (BDR): absolute loss (median) | 66.78 (0.66) | 94.75 (0.02) | 66.10 (0.61) | 55.48 (0.03) | 65.88 (0.69) | 104.31 (0.04) | 67.69 (0.51) | 79.77 (0.08) | 89.66 (0.47) | 52.97 (0.02) | 88.42 (0.35) | 29.14 (0.02) | 88.32 (0.43) | 60.62 (0.03) | 88.99 (0.36) | 47.07 (0.04) |
| min-max: quadratic loss (mean) | 78.06 (0.59) | 141.39 (0.04) | 78.96 (0.50) | 79.84 (0.04) | 73.44 (0.72) | 150.64 (0.06) | 75.45 (0.42) | 122.26 (0.10) | 94.28 (0.38) | 81.43 (0.03) | 94.76 (0.25) | 43.52 (0.03) | 91.38 (0.42) | 93.18 (0.04) | 92.28 (0.29) | 74.50 (0.06) |
| min-max (LQO): quadratic loss (mean) | 67.46 (0.65) | 116.92 (0.04) | 66.64 (0.62) | 62.93 (0.03) | 66.80 (0.72) | 131.38 (0.06) | 67.24 (0.51) | 101.23 (0.09) | 90.34 (0.42) | 69.04 (0.03) | 88.96 (0.40) | 35.48 (0.03) | 89.12 (0.45) | 86.47 (0.05) | 88.94 (0.33) | 63.21 (0.06) |
| min-max (BDR): quadratic loss (mean) | 66.62 (0.68) | 100.77 (0.03) | 66.28 (0.61) | 56.32 (0.03) | 65.64 (0.77) | 113.07 (0.05) | 67.20 (0.54) | 86.85 (0.08) | 89.94 (0.46) | 56.80 (0.02) | 88.68 (0.41) | 29.53 (0.02) | 88.68 (0.45) | 68.51 (0.04) | 88.75 (0.35) | 51.47 (0.05) |
| min-max: angular loss (median) | 86.30 (0.52) | 191.84 (0.09) | 74.94 (0.53) | 87.93 (0.06) | 81.98 (0.60) | 201.12 (0.14) | 75.39 (0.44) | 135.15 (0.11) | 95.84 (0.29) | 106.46 (0.06) | 92.62 (0.34) | 46.84 (0.05) | 94.00 (0.34) | 114.76 (0.11) | 92.00 (0.28) | 85.82 (0.08) |
| min-max (LQO): angular loss (median) | 68.64 (0.68) | 137.75 (0.06) | 66.94 (0.60) | 76.22 (0.05) | 66.96 (0.72) | 155.50 (0.13) | 67.73 (0.51) | 110.36 (0.12) | 89.84 (0.43) | 84.36 (0.05) | 88.74 (0.39) | 42.36 (0.04) | 89.54 (0.47) | 98.53 (0.12) | 89.16 (0.33) | 73.43 (0.07) |
| min-max (BDR): angular loss (median) | 67.82 (0.63) | 104.37 (0.02) | 66.48 (0.59) | 64.30 (0.04) | 65.86 (0.70) | 118.21 (0.08) | 67.27 (0.51) | 89.41 (0.10) | 89.42 (0.44) | 59.87 (0.02) | 88.56 (0.40) | 35.02 (0.03) | 89.00 (0.48) | 71.20 (0.07) | 88.74 (0.33) | 56.80 (0.06) |
| min-max: Chebyshev loss (median) | 68.48 (0.72) | 111.33 (0.02) | 66.52 (0.56) | 56.96 (0.02) | 66.88 (0.64) | 115.04 (0.03) | 67.12 (0.53) | 88.83 (0.06) | 91.16 (0.42) | 59.95 (0.01) | 88.34 (0.44) | 29.63 (0.01) | 88.96 (0.44) | 64.41 (0.02) | 89.53 (0.36) | 51.34 (0.03) |
| min-max (LQO): Chebyshev loss (median) | 68.48 (0.72) | 111.33 (0.02) | 66.52 (0.56) | 56.96 (0.02) | 66.88 (0.64) | 115.04 (0.03) | 67.12 (0.53) | 88.83 (0.06) | 91.16 (0.42) | 59.95 (0.01) | 88.34 (0.44) | 29.63 (0.01) | 88.96 (0.44) | 64.41 (0.02) | 89.53 (0.36) | 51.34 (0.03) |
| min-max (BDR): Chebyshev loss (median) | 68.48 (0.72) | 111.33 (0.02) | 66.52 (0.56) | 56.96 (0.02) | 66.88 (0.64) | 115.04 (0.03) | 67.12 (0.53) | 88.83 (0.06) | 91.16 (0.42) | 59.95 (0.01) | 88.34 (0.44) | 29.63 (0.01) | 88.96 (0.44) | 64.41 (0.02) | 89.53 (0.36) | 51.34 (0.03) |

*Notes: Multivariate, time-simultaneous coverage levels ($\Pi^{MV-TS}$) and average relative widths ($\Delta^{MV-TS}$) of 68% and 90% MV-TS error bands for DGP1, DGP2, DGP3 and DGP4, labeled* Π *and* Δ*, respectively, for simplicity. Standard deviations computed with bootstrap method are in parentheses. The 'median' or 'mean' indicated in parentheses next to each min-max estimator denotes the pointwise measure of central tendency used for the corresponding loss function.*

Finally, Table 6 presents the results of the Monte Carlo simulations for MV-TS-CS joint inference, which is conducted exclusively for DGP4, as it is the only model featuring more than one identified shock. The results for the 68% credible region indicate again that Bonferroni and Šidák method are overestimating the uncertainty, but the sup-t error band achieves the nominal coverage level at least within 2 standard deviations of the coverage statistics. The min-max estimators without any calibration step again overestimate the coverage level. The LQO step may yield a sufficiently accurate estimate, but as previously discussed, it can be challenging to obtain stable results—particularly in the presence of fuzzy joint distributions of certain structural parameters. On the other hand, the BDR step could underestimate the uncertainty with approximately 1% depending on the applied loss functions. Even though the simulation exercise gives very clear results for the performance of the method, the economic implications are insignificant. Furthermore, the methods tend to achieve the nominal coverage level for the 90% credible region.

Table 6
Multivariate, time-simultaneous, cross-shock coverage levels ($\Pi^{MV-TS-CS}$) and average relative widths ($\Delta^{MV-TS-CS}$) of MV-TS-CS error bands

| Method | 68% MV-TS-CS error band | | 90% MV-TS-CS error band | |
|---|---|---|---|---|
| | DGP4 | | DGP4 | |
| | $\Pi$ | $\Delta$ | $\Pi$ | $\Delta$ |
| Bonferroni | 94.68 | 228.52 | 98.32 | 121.96 |
| | (0.28) | (0.08) | (0.17) | (0.05) |
| Šidák | 93.42 | 220.37 | 98.28 | 120.68 |
| | (0.32) | (0.08) | (0.17) | (0.05) |
| sup-t (PQO) | 67.16 | 122.92 | 88.94 | 68.65 |
| | (0.66) | (0.07) | (0.41) | (0.03) |
| min-max: absolute loss (median) | 84.86 | 169.29 | 95.08 | 97.11 |
| | (0.54) | (0.15) | (0.32) | (0.06) |
| min-max (LQO): absolute loss (median) | 66.98 | 115.07 | 89.06 | 68.73 |
| | (0.66) | (0.09) | (0.44) | (0.05) |
| min-max (BDR): absolute loss (median) | 65.82 | 104.42 | 88.26 | 59.56 |
| | (0.71) | (0.09) | (0.45) | (0.05) |
| min-max: quadratic loss (mean) | 77.58 | 156.07 | 93.32 | 91.44 |
| | (0.57) | (0.13) | (0.36) | (0.07) |
| min-max (LQO): quadratic loss (mean) | 67.64 | 125.59 | 88.78 | 75.51 |
| | (0.68) | (0.10) | (0.45) | (0.05) |
| min-max (BDR): quadratic loss (mean) | 66.58 | 111.80 | 88.32 | 64.29 |
| | (0.68) | (0.09) | (0.46) | (0.05) |
| min-max: angular loss (median) | 78.22 | 182.89 | 92.84 | 114.39 |
| | (0.59) | (0.21) | (0.37) | (0.13) |
| min-max (LQO): angular loss (median) | 67.00 | 143.98 | 88.80 | 96.16 |
| | (0.68) | (0.18) | (0.49) | (0.13) |
| min-max (BDR): angular loss (median) | 66.32 | 118.35 | 87.96 | 73.46 |
| | (0.70) | (0.12) | (0.46) | (0.08) |
| min-max: Chebyshev loss (median) | 66.84 | 116.46 | 89.56 | 66.04 |
| | (0.76) | (0.07) | (0.42) | (0.03) |
| min-max (LQO): Chebyshev loss (median) | 66.84 | 116.46 | 89.56 | 66.04 |
| | (0.76) | (0.07) | (0.42) | (0.03) |
| min-max (BDR): Chebyshev loss (median) | 66.84 | 116.46 | 89.56 | 66.04 |
| | (0.76) | (0.07) | (0.42) | (0.03) |

Notes: Multivariate, time-simultaneous, cross-shock coverage levels ($\Pi^{MV-TS-CS}$) and average relative widths ($\Delta^{MV-TS-CS}$) of 68% and 90% MV-TS-CS error bands for DGP4, labeled $\Pi$ and $\Delta$, respectively, for simplicity. Standard deviations computed with bootstrap method are in parentheses. The 'median' or 'mean' indicated in parentheses next to each min-max estimator denotes the pointwise measure of central tendency used for the corresponding loss function.

### 4.3 Shape and Comovement

This section discusses whether Bayesian estimation is capable of preserving the joint distributions of non-linear transformations, such as the IRFs, beyond just achieving the desired coverage level for multiple parameters simultaneously. This aligns with the argument of Inoue and Kilian (2022), who advocate for using the lowest posterior risk joint credible set to describe the shape and comovement of IRFs rather than focusing solely on error bands.

Talts et al. (2020) show that if the prior of the Bayesian estimation procedure is specified to coincide with the distribution of the true parameters of the DGP, then the posterior distribution is the same as the prior distribution. The following rank statistics is used to evaluate if both distributions coincide:

$$r(\theta_{n,s,h}) = \sum_{d=1}^{D} \mathbb{1}\Big[(\theta_{n,s,h})^{(d:c)} < (\theta_{n,s,h})^{(c)}\Big], \quad \text{for } n = 1,\dots,N,\ s = 1,\dots,S \text{ and } h = 1,\dots,H,$$

where $\mathbb{1}[\cdot]$ is the indicator function defined over $(\theta_{n,s,h})^{(c)}$, $(\theta_{n,s,h})^{(d:c)}$, which denote for each Monte Carlo iteration $c$, the true impulse response coefficients and the $d$-th posterior draw, respectively. The distribution of the rank statistics is uniform across the integers $[0, D]$ if and only if the sampling procedure is correctly implemented. The rank statistics proposed by Talts et al. (2020) is applicable only to univariate random variables and allows to approach the question if each individual parameter is sampled correctly, but not if their joint distribution is approximated sufficiently well. Since evaluation of the marginal distributions involves significant amount of histograms and also this question is not the main focus of the study, then the analysis will turn towards the ability of the estimation to approximate the joint distributions of the structural parameters.

Thorarinsdottir et al. (2016) propose several vector-valued loss statistics, which are used to evaluate the multivariate distributions of different forecasting procedures applied in the context of weather forecasting.[12] Their comparative analysis shows that the best performing measures are the band depth and average rank. The analysis in this section relies on the latter statistics. Starting with any vector of stacked IRFs $\boldsymbol{\theta} \in \mathbb{R}^K$, the set of vectors $\mathbf{S} = \{(\boldsymbol{\theta})^{(1:c)}, (\boldsymbol{\theta})^{(2:c)}, \dots, (\boldsymbol{\theta})^{(D:c)}, (\boldsymbol{\theta})^{(c)}\}$ is the backbone of the derivation. The first $M$ elements of $\mathbf{S}$ denote draws from the posterior of this particular Monte Carlo iteration, $c$, and the last element denote the true vector of structural parameters. The pre-rank of the $j$-th coordinate of $\boldsymbol{\theta}$ in $\mathbf{S}$ is given by:

$$ru_S\left((\theta_j)^{(i:c)}\right) = \sum_{j=1}^{K} \mathbb{1}\left((\theta_j)^{(d:c)} \leq (\theta_j)^{(i:c)}\right) \quad \text{for } i = 1,\cdots, D+1.$$

[12] A similar procedure is proposed by Knüppel et al. (2022) to evaluate the joint distribution of forecasts in Bayesian VAR models.

Finally, the average rank statistic is simply the average over the univariate ranks

$$r_S\left((\boldsymbol{\theta})^{(i:c)}\right) = \frac{1}{K}\sum_{j=1}^{K} ru_S\left((\theta_j)^{(i:c)}\right) \quad \text{for } i = 1, \cdots, D+1,$$

where ties are resolved at random. The resulting rank of observation $(\boldsymbol{\theta})^{(D+1:c)}$ in $S$ is uniform across the integers $[0, D+1]$ if the elements of $S$ are independent and identically distributed (see Thorarinsdottir et al., 2016).

For simplicity in presenting the results, average rank statistics are calculated for three types of vectors of structural parameters: TS, MV-TS, and MV-TS-CS. The MV vector of stacked IRFs is excluded because it involves summarizing multiple quantities for each DGP and is already implicitly contained in two of the three specifications considered in the analysis. Following this, histograms are constructed to assess whether the statistics are uniformly distributed. This is done by checking if the bins fall within the 0.005 and 0.995 percentiles of a $Binomial(C, 1/B)$ distribution, where $C$ represents the number of Monte Carlo iterations and $B$ the number of bins. Under uniformity, only one bin in a hundred should fall outside these bounds (see Talts et al., 2020).

If the histogram of the average rank statistics is ∪-shaped, the joint distribution of the IRF is underdispersive. On the other hand, if the histogram of the average rank statistics is ∩-shaped, the joint distribution of the IRFs is overdispersive. The average rank statistics under TS inference for all DGPs is shown in Figure D.1 of Appendix D. For each DGP there is one histogram for every IRF - 3 for DGP1 and DGP2, 4 for DGP3 and 4 for DGP4. The bins in all histograms are within the bounds specified by the appropriate binomial distribution, therefore there is no evidence to reject the null hypothesis of uniformity. Analogously, the number of average rank statistics histograms in case of MV-TS inference for DGP1, DGP2 and DGP3 is 1 and for DGP4 is 2, all of which are presented in Figure D.2 of Appendix D. The bins of those histograms are also fully contained within the percentiles of the specified Binomial distribution, which implies that there is no evidence against the uniformity of the statistics. The same conclusion follows the analysis of the single histogram in the case of MV-TS-CS inference for DGP4 shown in Figure D.3 of Appendix D.

Overall, the uniform distribution of the average rank statistics implies that standard Bayesian inference on IRFs is not only capable of jointly achieving the targeted probability content, but also preserves the shapes and comovements of the IRFs. This suggests that the lowest posterior risk joint credible set is a reliable instrument for evaluating competing economic theories, as argued by Inoue and Kilian (2022). On the other hand, the sup-t joint inference method and the respective draws contained within the estimated error bands do not necessarily have this property as it does not evaluate the distance across draws. This question along with the

contribution of the loss function in the estimation of the lowest posterior risk set is left for future study.

# 5 Empirical Applications

Having established the importance of constructing context-specific error bands, this section highlights the benefits of the discussed methods for practitioners and policymakers. First, I demonstrate the implications of conducting joint inference on the fiscal multiplier in a typical VAR model. In the second subsection, I explore the similarities between IRFs and forecasts, applying joint inference methods in an out-of-sample forecasting exercise for inflation and real GDP growth in the United States and the euro area.

## 5.1 Fiscal Multiplier

There are periods when fiscal actions by governments become the main driver of economic activity. For example, the fiscal expenditure and revenue responses to the economic implications of the Covid-19 pandemic in September 2020 amounted to 9 percent of GDP in advanced economies (see Lacey et al., 2021). During such times, the fiscal multiplier becomes a key metric for macroeconomic forecasting and policy design, as it quantifies the impact of discretionary fiscal policy on output. Blanchard and Leigh (2013) argue that underestimation of fiscal multipliers led to persistent forecast errors of output for European economies in the beginning of the 2010s, especially in the early stages of the austerity measures implementation. Although the immediate consequences are evident, the failure to properly quantify the fiscal multiplier when implementing fiscal policy could have adverse long-term implications for the economy. For example, fiscal tightening can worsen the debt ratio in the short run, potentially leading the government to implement additional rounds of tightening in an attempt to stabilize the debt (e.g., Eyraud and Weber, 2013). As a result, confidence in the fiscal authority decreases, and the economy may enter a vicious cycle of slow growth, deflation, and further tightening.

In the context of VARs, the fiscal multiplier is usually computed as the ratio of the cumulative sum of the GDP response to the cumulative sum of the government expenditure response to an exogenous government expenditure shock (see Ramey, 2019). Identifying these shocks is not straightforward with sign restrictions, but DGP1 considered in the previous section provides a good example of a fiscal model identified with short-run exclusion restrictions. It consists of three endogenous variables ordered in the vector as follows: government expenditure, real GDP, and government revenue. The fiscal shock is identified through Cholesky decomposition of the variance-covariance matrix, which implies that a government expenditure shock does not affect real GDP and government revenues contemporaneously. In particular, using the previously

introduced notation, the fiscal multiplier is given by

$$m_h = \frac{\sum_{i=1}^{h} \theta_{2,1,i}}{\sum_{i=1}^{h} \theta_{1,1,i}} \cdot (Y/G),$$

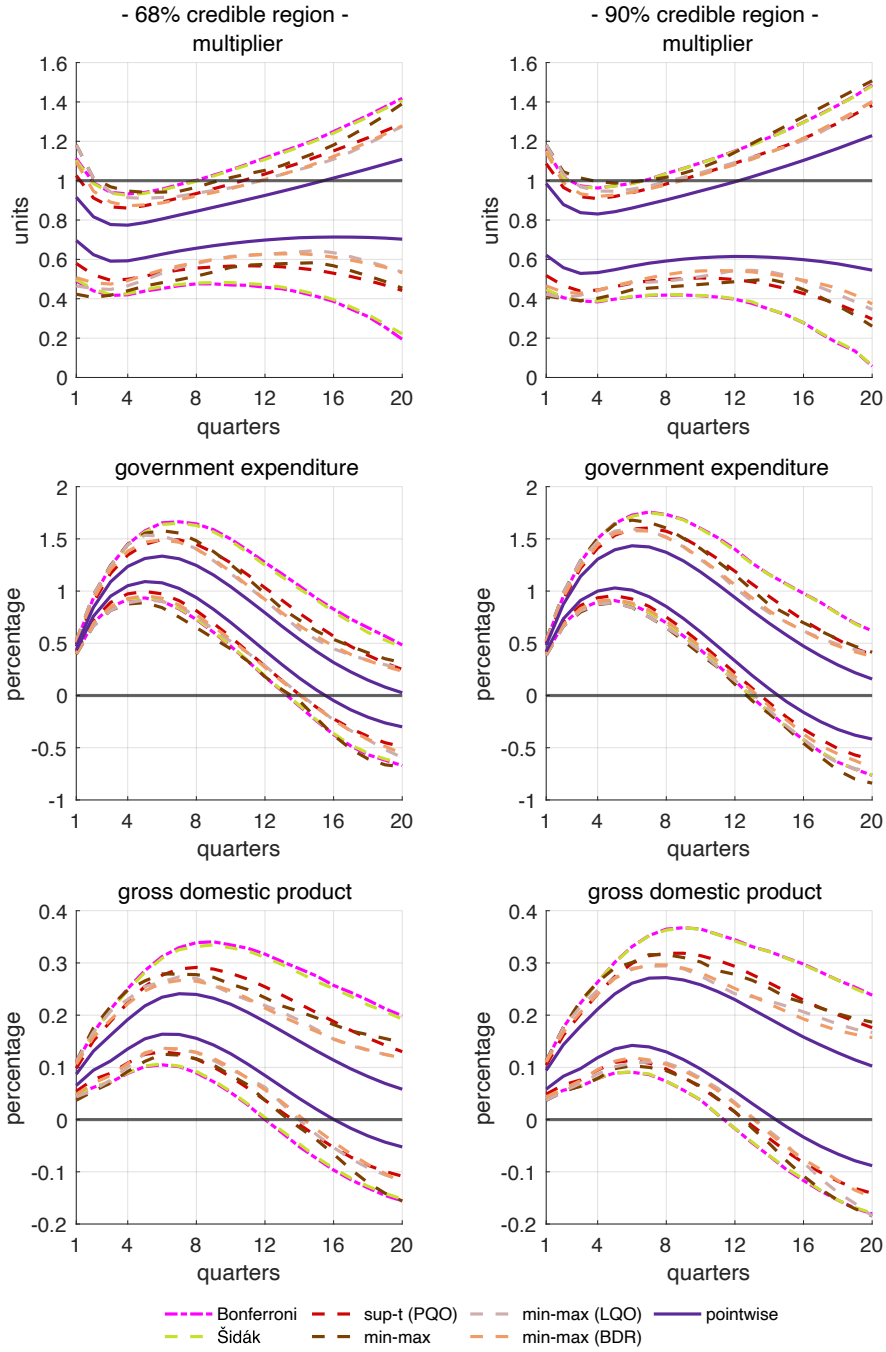
where $\theta_{2,1,i}$ denotes the IRF of GDP in period $i$ to an exogenous shock in fiscal expenditures, $\theta_{1,1,i}$ denotes the IRF of government expenditure in period $i$ with respect to the same shock, and $Y/G$ denotes the average GDP-to-government-expenditure ratio over the entire estimation sample. Therefore, the fiscal multiplier is a non-linear function of two variables that are accumulated over time. If the error bands around the fiscal multiplier are to be constructed by considering the specifics of the economic question, they need to achieve MV-TS coverage. In this case, the vector of structural parameters consists of the fiscal multiplier and the responses of government expenditure and real GDP over 20 quarters, but does not include the response of government revenue. In particular, the vector is given by:

$$\boldsymbol{\theta} = (m_1, m_2, \ldots, m_H, \theta_{1,1,1}, \theta_{1,1,2}, \ldots, \theta_{1,1,H}, \theta_{2,1,1}, \theta_{2,1,2}, \ldots, \theta_{2,1,H}).$$

The model is estimated using U.S. data from 1939:Q1 to 2015:Q4, and Figure 3 illustrates the conventional and joint error bands. The PW error bands suggest that the fiscal multiplier exceeds 1 only after 15 quarters with 68% probability, and after 13 quarters with 90% probability. As expected, the Bonferroni and Šidák error bands are the most conservative. Given their tendency to overstate uncertainty, the implications drawn from these methods are of limited relevance. In contrast, the error bands based on the sup-t method are wider than even the uncalibrated min-max error bands. Therefore, the analysis focuses on the min-max (BDR) specification, as the Monte Carlo exercise demonstrated that this method achieves the desired coverage level.

An important finding is that the LQO and BDR calibration methods yield nearly identical error bands. This similarity between the two calibrations highlights that the choice of intermediate step may not significantly affect the overall conclusions about the fiscal multiplier's behavior. Moreover, when comparing the width of the LQO and BDR error bands to the more conventional PW error bands, the min-max (BDR) bands are, on average, significantly wider—91% for the 68% credibility level and 51% for the 90% level. This additional width reflects the more conservative nature of the BDR method, suggesting a greater level of uncertainty about the multiplier's true value. Importantly, the joint error bands demonstrate that the multiplier is likely to exceed 1 earlier than previously expected,

Figure 3
Pointwise and joint inference about the fiscal multiplier




*Notes: The lines depict PW and MV-TS error bands for government expenditure, GDP, and the fiscal multiplier. For simplicity, the min-max estimators are computed using only the absolute loss function. The figures in the left column display the 68% credible regions, while those in the right column display the 90% credible regions.*

containing the value of 1 within the credible region 4 quarters sooner for both credibility levels. This outcome calls for a re-assessment of fiscal policy, as it implies that government spending could have a more immediate and pronounced effect on output than standard PW error bands suggest. The need for careful policy design is thus underscored, as overlooking these nuances in error band width could lead to misinformed decisions regarding the timing and magnitude of fiscal interventions.

### 5.2 Forecasting

Structural IRFs in VAR models are intended to trace the dynamic effects of a one-time structural shock to one variable on all other variables in the system over time. They provide detailed insights into the dynamic interactions within the system, illustrating how a shock propagates and influences each variable across different periods. In contrast, forecasts in VAR models aim to predict the future values of the variables based on their historical relationships. More formally, as shown in Appendix A, any VAR model can be represented as:

$$\mathbf{y}_{t+l} = \mathbf{J}\mathbf{A}^{l+1}\mathbf{Y}_{t-1} + \sum_{b=0}^{l} \mathbf{J}\mathbf{A}^{b}\mathbf{J}'\mathbf{u}_{t+l-b}.$$

Therefore, the $l$-periods ahead forecast is given by:

$$\mathbb{E}\left(\mathbf{y}_{t+l}|\mathbf{y}_t, \mathbf{y}_{t-1}, \dots\right) = \mathbf{J}\mathbf{A}^{l+1}\mathbf{Y}_{t-1}$$

and the $l$-periods ahead IRF is defined as

$$\mathbb{E}\left(\mathbf{y}_{t+l}|\mathbf{u}_t = \mathbf{B}\boldsymbol{\varepsilon}_t, \mathbf{y}_t, \mathbf{y}_{t-1}, \dots\right) = \mathbf{J}\mathbf{A}^{l+1}\mathbf{Y}_{t-1} + \mathbf{J}\mathbf{A}^{l}\mathbf{J}'\mathbf{B}\boldsymbol{\varepsilon}_t.$$

This illustrates the close relationship between IRFs and forecasts in time series models, suggesting that methods used for joint inference on these quantities are likely to exhibit similar performance. Given these similarities, the joint inference methods are applied in a pseudo-out-of-sample forecasting exercise, even though their forecasting performance is not formally evaluated in the simulation exercises presented in this paper.

For this exercise, I use two time series models that tend to perform reasonably well in out-of-sample forecasting. First, I consider a time-varying bivariate model of inflation and survey-based inflation expectations proposed by Chan et al. (2018) (henceforth CCK18). In this model, the inflation gap is defined as the difference between actual inflation and the unobserved inflation trend. The inflation gap is assumed to follow an AR(1) process, while the inflation trend is modeled as a random walk. Additionally, there is a linear relationship between the survey-based inflation expectations and the inflation trend.[13]

[13] Further details about the model can be found in Chan et al. (2018).

The second model used in the forecasting exercise is a standard time-varying VAR with stochastic volatility, originally proposed by Primiceri (2005) and later refined by Del Negro and Primiceri (2015) (henceforth DP15). This VAR model includes three endogenous variables: the real GDP growth rate, the headline inflation rate, and the 1-year government bond yield. The performance of the joint inference methods in the forecasting exercise is evaluated by constructing TS error bands for inflation forecasts using CCK18 and TS and MV-TS error bands for inflation and GDP growth rates using DP15.

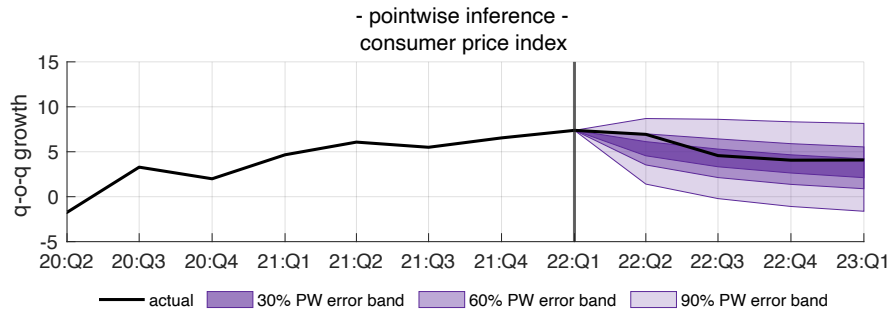
Both models are estimated using U.S. and euro area data, with samples starting in 1982:Q1 and 1999:Q1, respectively. The forecasts are evaluated through an expanding sample pseudo-out-of-sample forecasting exercise, beginning in 2010:Q1 and concluding in 2019:Q4.[14] The headline inflation is annualized quarter-on-quarter growth rate of the consumer price index (CPI)-measured by the Personal Consumption Expenditure (PCE) and the Harmonized Index of Consumer Prices (HICP) for the United States and euro area, respectively. For the United States, the survey-based inflation expectations are derived from the 10-year expected inflation estimated by the Federal Reserve Bank of Cleveland, while for the euro area, the 5-year expected inflation is reported by the Survey of Professional Forecasters of the European Central Bank. For DP15, in addition to the inflation rate, I use the 1-year U.S. Treasury bond yield and the 1-year German government bond yield, along with the annualized quarter-on-quarter real GDP growth rates for both the United States and euro area.

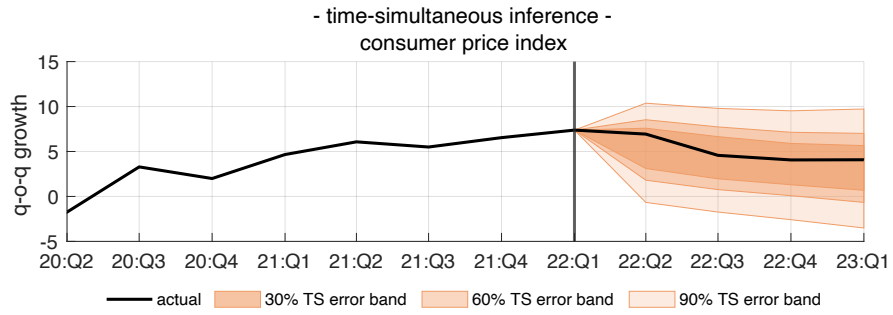
Before analyzing the results of the entire sample for each joint inference method, I first focus on the graphical comparison of the error bands. In the aftermath of the Covid pandemic, inflation in both the United States and the euro area experienced a strong and persistent increase, driven by a combination of supply and demand factors. These effects were further exacerbated by the energy crisis triggered by Russia's invasion of Ukraine in February 2022.

A simple projection conducted at the beginning of 2022:Q2 using CCK18 for the United States and euro area is presented in Figure 4. The figures for the TS error bands (calculated with min-max estimator under absolute loss function) in the United States show that, even after such a significant disruption, the inflation recovery path would have been well-contained within the 30% credible band, except for the second quarter of 2022. This is somewhat surprising, as the model is generally expected to produce better short-term forecasts. On the other hand, the European economy was more severely affected by energy supply disruptions, causing inflation to remain elevated for an extended period. As a result, the PW error band for the headline inflation rate in the euro area is less useful as a measure of the recovery path. As the econometric model is trend-reverting by nature, both PW and TS error band indicate downside risk for the inflation, but the realized path of the series is well contained only within the TS error bands for both economies.

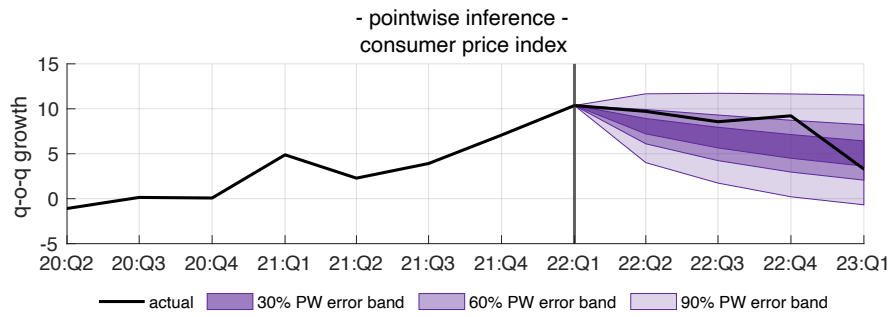
[14]The sample size ends in 2019 to simplify the econometric modeling of the post-Covid heteroskedasticity.

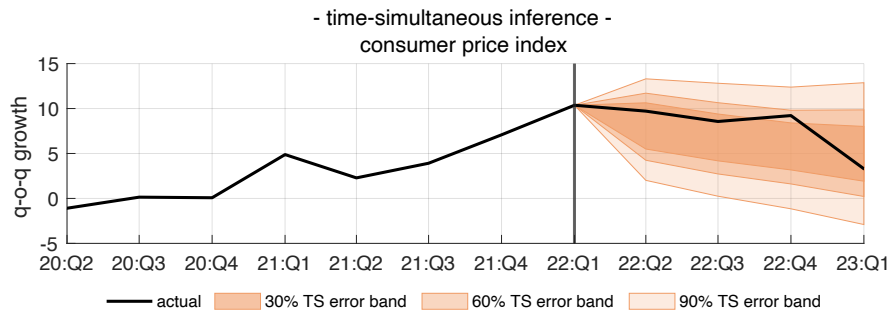
Figure 4
Inflation forecast as of 2022:Q2: PW and TS error bands produced with the CCK18 model for the United States and the euro area




(a) United States




(b) euro area

*Notes: The black solid lines depict the actual annualized quarter-on-quarter growth rate of CPI—measured by the PCE for the United States and the HICP for the euro area. The shaded areas depict the PW and TS error bands produced with the CCK18 model. The TS error bands are constructed using the min-max (BDR) approach with an absolute loss function. Subfigure (a) shows the results for the United States, while subfigure (b) shows the results for the euro area. The model is estimated using data through 2022:Q1.*

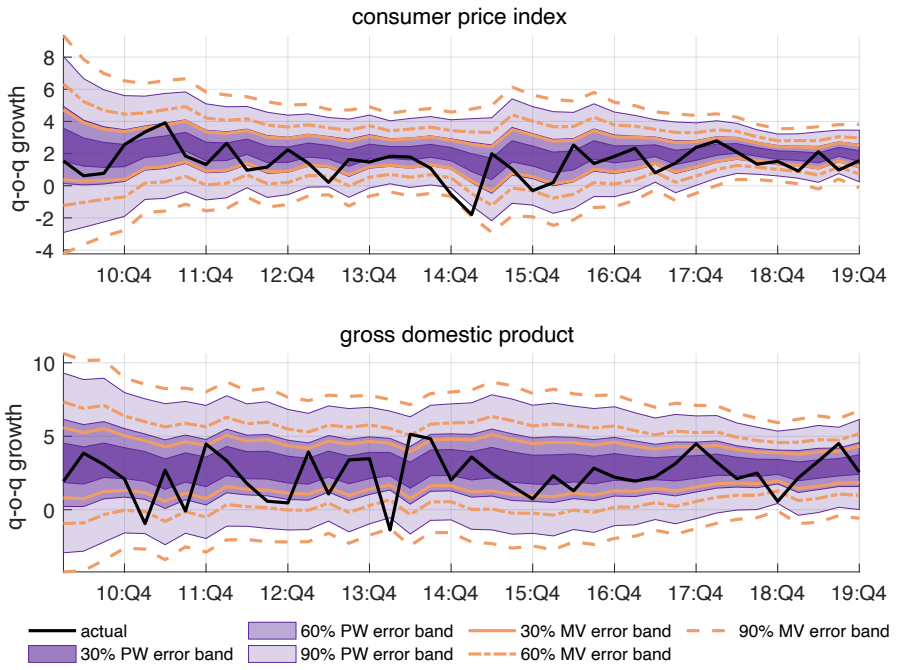
The results of the projection exercises conducted with the DP15 model are presented in Figure 5. It shows a comparison between the PW and MV error bands (using a calibrated min-max estimator under absolute loss function) of the joint projections of two-quarters-ahead inflation and GDP q-o-q growth rates during the period 2010:Q1-2019:Q4. The results show that the 90% MV error bands are not much wider in comparison to the corresponding PW error bands, but the difference between both starts to increase as the credibility level falls. This indicates that underestimation of uncertainty for lower credibility levels will possibly be much larger. A possible cause for that is the nature of the projected variables, namely growth rates. A further important implication of the calibration step implemented for the min-max estimators is that it stabilizes the results across different initialization of the forecasting exercises. This is not caused by not converged estimation procedure, but rather infeasible representation of the joint distribution function.

The coverage rates of the PW error bands for the CCK18 model are presented in Appendix E, and in particular Table E.1 for United States and E.2 for the euro area. The results of the out-of-sample forecasting exercise reveal that even the PW error bands for United States slightly overestimate the uncertainty especially for credible levels of 30% and 60%. On the other hand, the euro area results indicate much better performance of the model. Nonetheless, if the applied researcher reports PW error bands, but requires TS inference the uncertainty will be significantly underestimated. The TS error bands evaluated against TS coverage statistics are presented in Figure E.3 and E.4. The results indicate same conclusions regarding the performance of the models as for the error bands of the IRFs evaluated with the Monte Carlo exercises. The difference, however, is driven by the general (pointwise) forecasting performance of the model for the U.S. and euro area inflation projections. Furthermore, the TS error bands for 8-quarters ahead and smaller credibility level are significantly wider than the PW equivalent, which could make them inapplicable in practical applications.

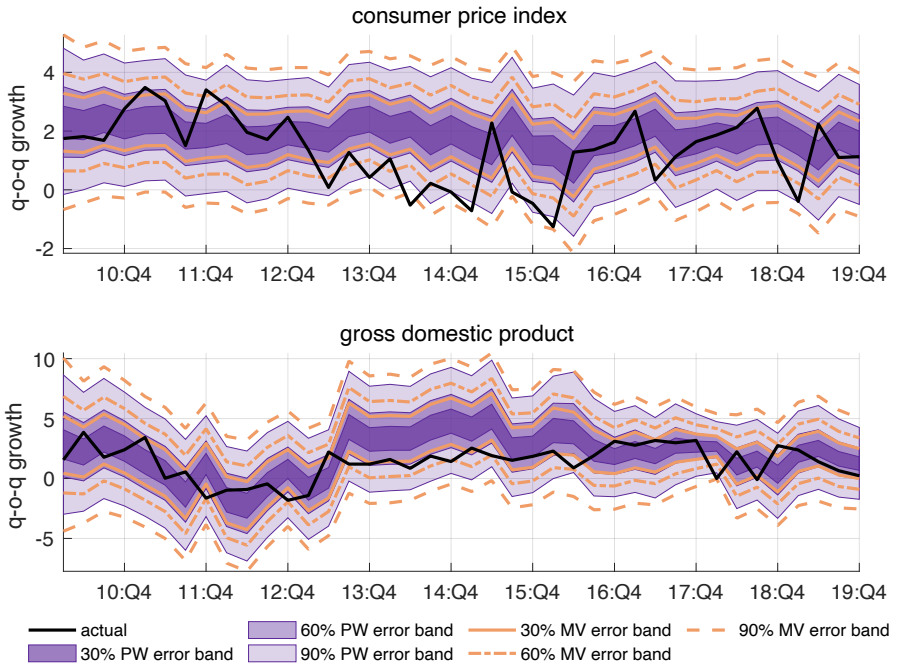
Finally, I evaluate the results of the out-of-sample forecasting exercise based on the DP15 model. The PW error bands evaluated under PW, MV and TS coverage statistics are shown in Table E.5 and E.6. Both tables indicate much better PW coverage of the inflation forecasts in comparison to the GDP growth. Furthermore, there is also difference in the performance across economies - with the projections for United States performing slightly better in comparison to the euro area. The coverage levels of the MV and TS again indicate that the PW error bands are inadequate representations of the joint uncertainty.

The results for the TS error bands are presented in Table E.7 and E.8. As expected the performance of the U.S. forecasting is slightly better. The coverage rate of the min-max estimators under absolute, quadratic and angular loss functions even with BDR calibration step significantly worse in comparison to the Chebyshev loss functions. The estimates under the latter loss function are also significantly tighter. For example, U.S. results indicate that the 60% error band for $h = 4$ and GDP growth rate (i.e., $y_2$) estimated with min-max estimator under absolute loss

Figure 5
Inflation and real GDP growth rate two-quarters-ahead forecast in 2010:Q1-2019:Q4: PW and MV error bands produced with DP15 for the United States and the euro area




(a) United States




(b) euro area

*Notes: The black solid lines depict the actual annualized quarter-on-quarter growth rate of CPI and real GDP. The CPI is measured using the PCE for the United States and the HICP for the euro area. Shaded areas and dashed lines represent PW and MV error bands, respectively, for two-quarters-ahead forecasts produced with the DP15 model. The MV error bands are computed with min-max (BDR) and absolute loss function. Subfigure (a) shows the results for United States and subfigure (b) shows the results for the euro area. The model is estimated using a successively expanding sample, starting in 2010:Q1.*

function with BDR calibration covers 90% and is 188% of the PW error band. On the other hand, the same measures under Chebyshev loss are 79% and 83%. The results in both tables (for United States and euro area) show similar performance in terms of magnitudes of the Chebyshev loss function. The performance of the sup-t error bands is comparable to the Chebyshev min-max estimator. The results for the MV error bands are shown on Table E.9 for United States and Table E.10 for the euro area. As the MV error bands involve less elements to conduct joint inference about, the performance of the min-max estimator under the absolute and Chebyshev loss are comparable among each other but also the sup-t estimator is good alternative.

In summary, the joint error bands about the forecasts have similar performance in comparison to the joint error bands about the IRFs. Since this represents a real-world application rather than an exercise in simulated environment, the performance of the models even in conventional pointwise sense is rather mixed. The results, however, indicate that the applied researcher could use either of the methods - sup-t, min-max under absolute loss with BDR calibration or plain min-max under Chebyshev loss and report the other error bands in robustness exercises.

The importance of conducting joint inference about forecasts is also acknowledged and extensively discussed in Bernanke (2024) as part of the review of the Bank of England's forecasting practices in the context of monetary policymaking and communication. He argues that members of the Monetary Policy Committee have discretionary authority to adjust the width of the error bands or to skew the entire distribution. However, such adjustments do not account for the interdependence between forecasts of different variables, leading to inconsistencies when the error bands of other variables remain unchanged. The joint inference estimators proposed in this paper offer a potential solution to this issue, providing a framework for producing internally consistent and more robust economic projections.

# 6 Conclusion

This paper highlights the critical need to accurately quantify uncertainty in IRFs and forecasts when using VAR models. I discuss the construction of context-specific error bands, tailored to the economic question at hand, as different inquiries may require distinct set of structural parameters or forecasts to effectively capture the underlying uncertainty. Through series of simulation experiments, I demonstrate that conventional pointwise quantiles significantly underestimate the uncertainty in cases that require simultaneous inference about multiple parameters. The results illustrate that joint inference methods, such as the sup-t (see Montiel Olea and Plagborg-Møller, 2019) and min-max (see Inoue and Kilian, 2022; Akram et al., 2016) estimators, offer substantial improvements in achieving the desired coverage levels across horizons and variables, compared to pointwise methods.

The joint error bands, however, are considerably wider, especially for economic questions involving multiple variables and horizons. These findings underline the potential risks associated with relying on traditional pointwise methods, which may lead to overly confident conclusions. On the other hand, the joint inference methods I evaluate provide a more reliable reflection of the underlying uncertainty, making them suitable for applied research where precise estimation of IRFs and forecasts is crucial. This is particularly relevant for policy makers, when understanding the full scope of uncertainty is essential for informed decision-making.

Moreover, the carefully designed simulation exercise reveals that the min-max estimators, which rely on vector-valued loss functions can be overly conservative. Contrary to Inoue and Kilian (2022), I contend that the choice of the loss function plays a crucial role in determining the width of the error bands, and applied econometricians must be meticulous in selecting it. For example, the angular loss function consistently performed the worst, producing the widest error bands. In contrast, calibrated versions of the min-max estimator with quadratic or absolute loss functions performed much better, offering narrower bands while still achieving the desired coverage. Notably, the Chebyshev loss function proposed by Akram et al. (2016) stands out, as it achieves the target coverage level straight off the shelf, without requiring any additional calibration steps.

In addition, this study introduces simulation-based calibration to the field of macroeconometrics - a technique commonly used in other areas of econometrics to assess the performance of Bayesian estimators. Specifically, it allows for evaluating whether the posterior distribution is properly sampled by the proposed algorithm and the software implementation (see Gelman et al., 1996; Talts et al., 2020).

Unfortunately, the post-Covid period is characterized by increased volatility, resulting in excessively wide error bands. Consequently, the implications of this paper are less appealing for practitioners and policymakers. However, understating uncertainty is undesirable when investigating structural relationships or producing robust economic projections. Therefore, applied economists must carefully model heteroskedasticity and rely on more precise estimation techniques. These methods could relate to the identification methodology—for example, using narrative information to shrink error bands with shocks identified via sign restrictions or improving reduced-form estimates by estimating economic relationships in data-rich environment.

## References


Akram, F., Binning, A., and Maih, J. (2016). Joint Prediction Bands for Macroeconomic Risk Management. Working Paper 7-2016, Norges Bank.

Arias, J., Rubio Ramírez, J., and Waggoner, D. (2022). Uniform Priors for Impulse Responses. Working Paper 22-30, Federal Reserve Bank of Philadelphia.

Arias, J. E., Caldara, D., and Rubio-Ramírez, J. F. (2019). The Systematic Component of Monetary Policy in SVARs: An Agnostic Identification Procedure. *Journal of Monetary Economics*, 101:1–13.

Bernanke, B. S. (2024). Forecasting for Monetary Policy-Making and Communication at the Bank of England: A Review. *Bank of England Independent Evaluation Office*.

Bernardo, J. M. (2011). Integrated Objective Bayesian Estimation and Hypothesis Testing. In Bernardo, J. M., Bayarri, M. J., Berger, J. O., Bernardo, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckermann, D., Smith, A. F. M., and West, M., editors, *Bayesian Statistics 9: Proceedings of the Ninth Valencia International Meeting; June 3-8, 2010*, Oxford Science Publications. Oxford University Press, Oxford.

Blanchard, O. J. (1989). A Traditional Interpretation of Macroeconomic Fluctuations. *American Economic Review*, 79(5):1146–1164.

Blanchard, O. J. and Leigh, D. (2013). Growth Forecast Errors and Fiscal Multipliers. *American Economic Review*, 103(3):117–20.

Blanchard, O. J. and Quah, D. (1989). The Dynamic Effects of Aggregate Demand and Supply Disturbances. *American Economic Review*, 79(4):655–673.

Bonferroni, C. E. (1936). *Teoria Statistica Delle Classi e Calcolo Delle Probabilità*. Pubblicazioni Del R. Istituto Superiore Di Scienze Economiche e Commerciali Di Firenze. Seeber.

Chan, J. C., Clark, T. E., and Koop, G. (2018). A New Model of Inflation, Trend Inflation, and Long-Run Inflation Expectations. *Journal of Money, Credit and Banking*, 50(1):5–53.

Cook, S. R., Gelman, A., and Rubin, D. B. (2006). Validation of Software for Bayesian Models Using Posterior Quantiles. *Journal of Computational and Graphical Statistics*, 15(3):675–692.

Del Negro, M. and Primiceri, G. E. (2015). Time Varying Structural Vector Autoregressions and Monetary Policy: A Corrigendum. *The Review of Economic Studies*, 82(4):1342–1345.

Eichenbaum, M. and Evans, C. L. (1995). Some Empirical Evidence on the Effects of Shocks to Monetary Policy on Exchange Rates. *The Quarterly Journal of Economics*, 110(4):975–1009.

Eyraud, L. and Weber, A. (2013). The Challenge of Debt Reduction during Fiscal Consolidation. Working Paper 13-67, International Monetary Fund.

Forbes, K., Hjortsoe, I., and Nenova, T. (2018). The Shocks Matter: Improving our Estimates of Exchange Rate Pass-through. *Journal of International Economics*, 114:255–275.

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., and Gelman, A. (2019). Visualization in Bayesian Workflow. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 182(2):389–402.

Gafarov, B., Meier, M., and Montiel Olea, J. L. (2018). Delta-method Inference for a Class of Set-identified SVARs. *Journal of Econometrics*, 203(2):316–327.

Gelman, A., Meng, X.-L., and Stern, H. (1996). Posterior Predictive Assesment of Model Fitness via Realized Discrepancies. *Statistica Sinica*, 6(4):733–760.

Güntner, J., Reif, M., and Wolters, M. (2024). Sudden Stop: Supply and Demand Shocks in the German Natural Gas Market. *Journal of Applied Econometrics*, 39(7):1282–1300.

Herrera, A. M. and Rangaraju, S. K. (2020). The Effect of Oil Supply Shocks on US Economic Activity: What Have We Learned? *Journal of Applied Econometrics*, 35(2):141–159.

Inoue, A., Jordà, Ò., and Kuersteiner, G. M. (2025). Inference for Local Projections. *The Econometrics Journal*, (forthcoming).

Inoue, A. and Kilian, L. (2013). Inference on Impulse Response Functions in Structural VAR Models. *Journal of Econometrics*, 177(1):1–13.

Inoue, A. and Kilian, L. (2016). Joint Confidence Sets for Structural Impulse Responses. *Innovations in Multiple Time Series Analysis*, 192(2):421–432.

Inoue, A. and Kilian, L. (2022). Joint Bayesian Inference About Impulse Responses in VAR Models. *Special Issue: The Econometrics of Macroeconomic and Financial Data*, 231(2):457–476.

Jarociński, M. and Karadi, P. (2020). Deconstructing Monetary Policy Surprises—The Role of Information Shocks. *American Economic Journal: Macroeconomics*, 12(2):1–43.

Jordà, Ò. (2009). Simultaneous Confidence Regions for Impulse Responses. *The Review of Economics and Statistics*, 91(3):629–647.

Kaido, H., Molinari, F., and Stoye, J. (2019). Confidence Intervals for Projections of Partially Identified Parameters. *Econometrica*, 87(4):1397–1432.

Kilian, L. and Lütkepohl, H. (2017). *Structural Vector Autoregressive Analysis*. Cambridge University Press, 1 edition.

Kilian, L. and Zhou, X. (2022). The Impact of Rising Oil Prices on U.S. Inflation and Inflation expectations in 2020–23. *Energy Economics*, 113:106228.

Knüppel, M., Krüger, F., and Pohle, M.-O. (2022). Score-based Calibration Testing for Multivariate Forecast Distributions. Working Paper 50/2022, Deutsche Bundesbank.

Koehler, E., Brown, E., and Haneuse, S. J.-P. A. (2009). On the Assessment of Monte Carlo Error in Simulation-Based Statistical Analyses. *The American Statistician*, 63(2):155–162.

Lacey, E., Massad, J., and Utz, R. (2021). *A Review of Fiscal Policy Responses to COVID-19*. World Bank.

Little, R. J. (2006). Calibrated Bayes: A Bayes/Frequentist Roadmap. *The American Statistician*, 60(3):213–223.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. New York : Springer, Berlin.

Lütkepohl, H., Staszewska-Bystrova, A., and Winker, P. (2015). Comparison of Methods for Constructing Joint Confidence Bands for Impulse Response Functions. *International Journal of Forecasting*, 31(3):782–798.

Lütkepohl, H., Staszewska-Bystrova, A., and Winker, P. (2018). Calculating Joint Confidence Bands for Impulse Response Functions Using Highest Density Regions. *Empirical Economics*, 55(4):1389–1411.

Montiel Olea, J. L. and Plagborg-Møller, M. (2019). Simultaneous Confidence Bands: Theory, Implementation, and an Application to SVARs. *Journal of Applied Econometrics*, 34(1):1–17.

Paul, P. (2020). The Time-Varying Effect of Monetary Policy on Asset Prices. *The Review of Economics and Statistics*, 102(4):690–704.

Peersman, G. (2005). What Caused the Early Millennium Slowdown? Evidence Based on Vector Autoregressions. *Journal of Applied Econometrics*, 20(2):185–207.

Primiceri, G. E. (2005). Time Varying Structural Vector Autoregressions and Monetary Policy. *The Review of Economic Studies*, 72(3):821–852.

Ramey, V. A. (2019). Ten Years after the Financial Crisis: What Have We Learned from the Renaissance in Fiscal Research? *Journal of Economic Perspectives*, 33(2):89–114.

Šidák, Z. (1967). Rectangular Confidence Regions for the Means of Multivariate Normal Distributions. *Journal of the American Statistical Association*, 62(318):626–633.

Sims, C. A. and Zha, T. (1999). Error Bands for Impulse Responses. *Econometrica*, 67(5):1113–1155.

Staszewska, A. (2007). Representing Uncertainty About Response Paths: The Use of Heuristic Optimisation Methods. *Computational Statistics & Data Analysis*, 52(1):121–132.

Talts, S., Betancourt, M., Simpson, D., Vehtari, A., and Gelman, A. (2020). Validating Bayesian Inference Algorithms with Simulation-Based Calibration.

Thorarinsdottir, T. L., Scheuerer, M., and Heinz, C. (2016). Assessing the Calibration of High-Dimensional Ensemble Forecasts Using Rank Histograms. *Journal of Computational and Graphical Statistics*, 25(1):105–122.

Wolf, M. and Wunderli, D. (2015). Bootstrap Joint Prediction Regions. *Journal of Time Series Analysis*, 36(3):352–376.

Zhou, X. (2020). Refining the Workhorse Oil Market Model. *Journal of Applied Econometrics*, 35(1):130–140.

# Appendix A Bayesian Estimation and Structural Identification

This appendix defines the reduced-form model and introduces identification methodologies used in the simulation experiments.

**Reduced-form model.** The data generating models employed in the Monte Carlo simulations follow a general reduced-form VAR setup:

$$\mathbf{y}_t = \mathbf{c} + \sum_{l=1}^{p} \mathbf{A}_l \mathbf{y}_{t-l} + \mathbf{V}\mathbf{z}_t + \mathbf{u}_t, \quad \text{for } 1 \leq t \leq T,$$

where $\mathbf{y}_t$ is $(N \times 1)$ vector of endogenous variables, $\mathbf{A}_l$ are $(N \times N)$ matrices with parameters, $\mathbf{c}$ is $(N \times 1)$ vector of intercepts, $\mathbf{z}_t$ is $(G \times 1)$ vector of exogenous variables with its corresponding $(N \times G)$ matrix of parameters denoted as $\mathbf{V}$, $N$ is the number of endogenous variables, $G$ is the number of exogenous variables, $p$ is the lag length and $T$ is the sample size. All the models employed in this study are assumed to be stable, necessitating that $det(A(z)) \neq 0$, with the lag polynomial defined as $A(L) := \mathbf{I}_{N\times N} - \mathbf{A}_1 L - \cdots - \mathbf{A}_p L^p$ and an identity matrix $\mathbf{I}_N$ of size $(N \times N)$.

The companion form of the model can be written as

$$\mathbf{Y} = \mathbf{X}\mathbf{A} + \mathbf{u},$$

where $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T)'$, $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_T)'$, $\mathbf{X}_t = (\mathbf{i}_N', \mathbf{y}_{t-1}', \ldots, \mathbf{y}_{t-p}', \mathbf{z}_t')'$, $\mathbf{A} = (\boldsymbol{\mu}, \mathbf{A}_1, \ldots, \mathbf{A}_p)'$, $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_T)'$, and $\mathbf{i}_N$ is $(N \times 1)$ vector of all ones.

**Bayesian estimation.** The reduced-form VAR parameters are specified with normal-inverse Wishart prior distribution following a conventional approach (e.g., Inoue and Kilian, 2013)

$$vec(\mathbf{A})|\boldsymbol{\Sigma} \sim \mathbf{N}\left(vec(\bar{\mathbf{A}}_0), \boldsymbol{\Sigma} \otimes \mathbf{N}_0^{-1}\right) \quad \text{and} \quad \boldsymbol{\Sigma} \sim IW_n(\nu_0 \mathbf{S}_0, \nu_0),$$

where $\mathbf{N}_0$ is $(p \cdot N \times p \cdot N)$ positive definite matrix, $\mathbf{S}_0$ is $(N \times N)$ covariance matrix and $\nu_0 > 0$. This model implies a posterior distribution of the same functional family as the prior given by

$$vec(\mathbf{A})|\boldsymbol{\Sigma} \sim \mathbf{N}\left(vec(\bar{\mathbf{A}}_T), \boldsymbol{\Sigma} \otimes \mathbf{N}_T^{-1}\right) \quad \text{and} \quad \boldsymbol{\Sigma} \sim IW_n(\nu_T \mathbf{S}_T, \nu_T),$$

where

$$\begin{aligned}
&\nu_T = T + \nu, \quad \mathbf{N}_T = \mathbf{N}_0 + \mathbf{X}'\mathbf{X}, \quad \bar{\mathbf{A}}_T = \mathbf{N}_T^{-1}(\mathbf{N}_0\bar{\mathbf{A}}_0 + \mathbf{X}'\mathbf{X}\widehat{\mathbf{A}}),\\
&\mathbf{S}_T = \frac{\nu_0}{\nu_T}\mathbf{S}_0 + \frac{T}{\nu_T}\widehat{\boldsymbol{\Sigma}} + \frac{1}{\nu_T}(\widehat{\mathbf{A}} - \bar{\mathbf{A}}_0)'\mathbf{N}_0\mathbf{N}_T^{-1}\mathbf{X}'\mathbf{X}(\widehat{\mathbf{A}} - \bar{\mathbf{A}}_0),\\
&\widehat{\mathbf{A}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad \text{and} \quad \widehat{\boldsymbol{\Sigma}} = (\mathbf{Y} - \mathbf{X}\widehat{\mathbf{A}})'(\mathbf{Y} - \mathbf{X}\widehat{\mathbf{A}})/T.
\end{aligned}$$

Throughout this paper, two distinct parametrizations of the prior distribution are considered: a "non-informative" prior, characterized by:

$$\nu_0 = 0, \quad \mathbf{S}_0 = \mathbf{O}_{N\times N}, \quad \bar{\mathbf{A}}_0 = \mathbf{O}_{p\cdot N+R\times N}, \quad \text{and} \quad \mathbf{N}_0^{-1} = \mathbf{O}_{p\cdot N+R\times p\cdot N+R},$$

where $\mathbf{O}$ denotes zero matrix with size indicated by the subindex. The second prior parametrization, referred to as a "unit root," is defined by:

$$\nu_0 = N + 2, \quad \mathbf{S}_0 = \begin{pmatrix} \widetilde{\sigma}_1^2 & & \\ & \ddots & \\ & & \widetilde{\sigma}_p^2 \end{pmatrix}, \quad \bar{\mathbf{A}}_0 = \begin{pmatrix} \mathbf{O}'_{R\times N} & \mathbf{I}'_{N\times N} & \mathbf{O}'_{p\cdot(N-1)\times N} \end{pmatrix}', \quad \text{and}$$

$$\mathbf{N}_0^{-1} = \begin{pmatrix} \omega\mathbf{I}_{R\times R} & 1^{-\alpha}\lambda\mathbf{S}_0^{-1} & 2^{-\alpha}\lambda\mathbf{S}_0^{-1} & \dots & p^{-\alpha}\lambda\mathbf{S}_0^{-1} \end{pmatrix}',$$

where $\omega = 10^5$, $\lambda = 0.2$, $\alpha = 2$, and $\widetilde{\sigma}_\cdot^2$ is the error's variance of the regression $y_{t,\cdot} = \gamma_0 + \gamma_1 y_{t-1,\cdot} + e_t$, which is estimated using OLS.

**Structural Model.** The computation of the impulse response coefficients requires first defining the following companion form

$$\underbrace{\begin{pmatrix} \mathbf{y}_t \\ \mathbf{y}_{t-1} \\ \vdots \\ \mathbf{y}_{t-p} \end{pmatrix}}_{\mathbf{Y}_t} = \underbrace{\begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \dots & \mathbf{A}_{p-1} & \mathbf{A}_p \\ \mathbf{I}_{N\times N} & \mathbf{O}_{N\times N} & & \mathbf{O}_{N\times N} & \mathbf{O}_{N\times N} \\ \mathbf{O}_{N\times N} & \mathbf{I}_{N\times N} & & \mathbf{O}_{N\times N} & \mathbf{O}_{N\times N} \\ \vdots & & \ddots & \vdots & \vdots \\ \mathbf{O}_{N\times N} & \mathbf{O}_{N\times N} & \dots & \mathbf{I}_{N\times N} & \mathbf{O}_{N\times N} \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} \mathbf{y}_{t-1} \\ \mathbf{y}_{t-2} \\ \vdots \\ \mathbf{y}_{t-p-1} \end{pmatrix}}_{\mathbf{Y}_{t-1}} + \underbrace{\begin{pmatrix} \mathbf{u}_t \\ \mathbf{0}_N \\ \vdots \\ \mathbf{0}_N \end{pmatrix}}_{\mathbf{U}_t}.$$

It is well known that by successive substitution for $\mathbf{Y}_{t-l}$ the definition of the companion form reduces to

$$\mathbf{Y}_{t+l} = \mathbf{A}^{l+1}\mathbf{Y}_{t-1} + \sum_{b=0}^{l} \mathbf{A}^b \mathbf{U}_{t+l-b}$$

and left-multiplying by $\mathbf{J} = \begin{pmatrix} \mathbf{I}_{N\times N} & \mathbf{0}_{N\times(p-1)N} \end{pmatrix}$ results in:

$$\mathbf{y}_{t+l} = \mathbf{J}\mathbf{A}^{l+1}\mathbf{Y}_{t-1} + \sum_{b=0}^{l} \mathbf{J}\mathbf{A}^b\mathbf{J}'\mathbf{u}_{t+l-b}.$$

Finally, the $n$-th row and $s$-th column element $\theta_{n,s,h}$ of the impulse response coefficient matrix $\mathbf{\Theta}_h$, which quantifies the effect on variable $n$ in response to a structural shock $s$ of size one in period $h$, is defined as:

$$\mathbf{\Theta}_h = \{\theta_{n,s,h}\} = \mathbf{J}\mathbf{A}^h\mathbf{J}'\mathbf{B},$$

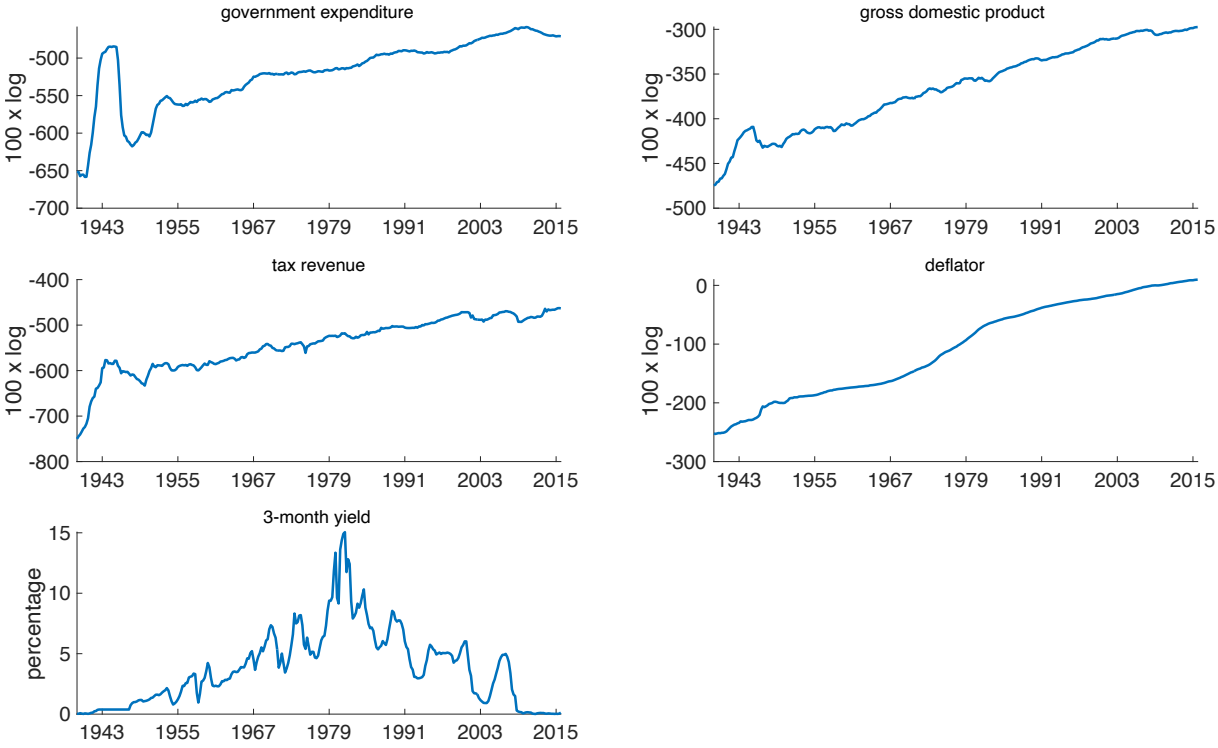
where $\mathbf{B}$ is the structural impact matrix. A linear transformation on the reduced-form errors defines the structural shocks as $\varepsilon_t = \mathbf{B}^{-1}\mathbf{u}_t$, which are assumed to be contemporaneously uncorrelated and have unit variance. Therefore, the relationship $\mathbf{B}\mathbf{B}' = \mathbf{\Sigma}$ must hold. However, the matrix $\mathbf{B}$ is not unique, and identification requires additional assumptions. In the simulation exercises, four different identification methodologies are considered. The first is recursive identification resulting in structural impact matrix $\mathbf{B}^{-1} = \text{chol}(\mathbf{\Sigma})$, i.e., the lower triangular matrix of the Cholesky decomposition. The second identification strategy relies on sign restrictions motivated by economic theory. To this end, for each posterior draw, $i$, of the error's covariance matrix, the lower triangular matrix of the Cholesky decomposition $\text{chol}(\mathbf{\Sigma}_i)$ is calculated and then an orthogonal matrix, $\mathbf{Q}_i$, is drawn such that the structural impact matrix given by $\mathbf{B}_i^{-1} =$

chol($\mathbf{\Sigma}_i$)$\mathbf{Q}_i$ implies IRFs, which are kept only if they satisfy the imposed sign restrictions. The described procedure is repeated $L$ times for each posterior draw of the reduced-form model. The third methodology requires an instrumental variable, $\mathbf{z}_t$, which enters the reduced-form VAR as exogenous variable and the column vector $\mathbf{V}$ gives the first column of the structural impact matrix, $\mathbf{B}^{-1}$, while the remaining columns have zero in all entries.[15] The term $\mathbf{V}\mathbf{z}_t$ is included in the reduced-form VAR setup only for models, which rely on instrumental variables for the identification. The fourth and last identification scheme considered in the Monte Carlo simulations imposes a short-run and long-run restrictions on the IRFs. The implementation follows the methodology of Blanchard (1989) that requires stationary vector of endogenous variables, (i.e., $\mathbf{y}_t \sim I(0)$). Then defining the matrix $\mathbf{\Omega} = \mathbf{J}(\mathbf{I}_{pN\times pN} - \widetilde{\mathbf{A}})^{-1}\mathbf{J}'$ allows to derive the structural impact matrix as $\mathbf{B} = \mathbf{\Omega}^{-1} chol(\mathbf{\Omega}\mathbf{\Sigma}\mathbf{\Omega}')$.

[15]This instrumental variable identification methodology is discussed in details by Paul (2020) and has the usual requirements, namely relevancy with respect to the structural shock of interest and also exogeneity with respect to the remaining structural shocks.

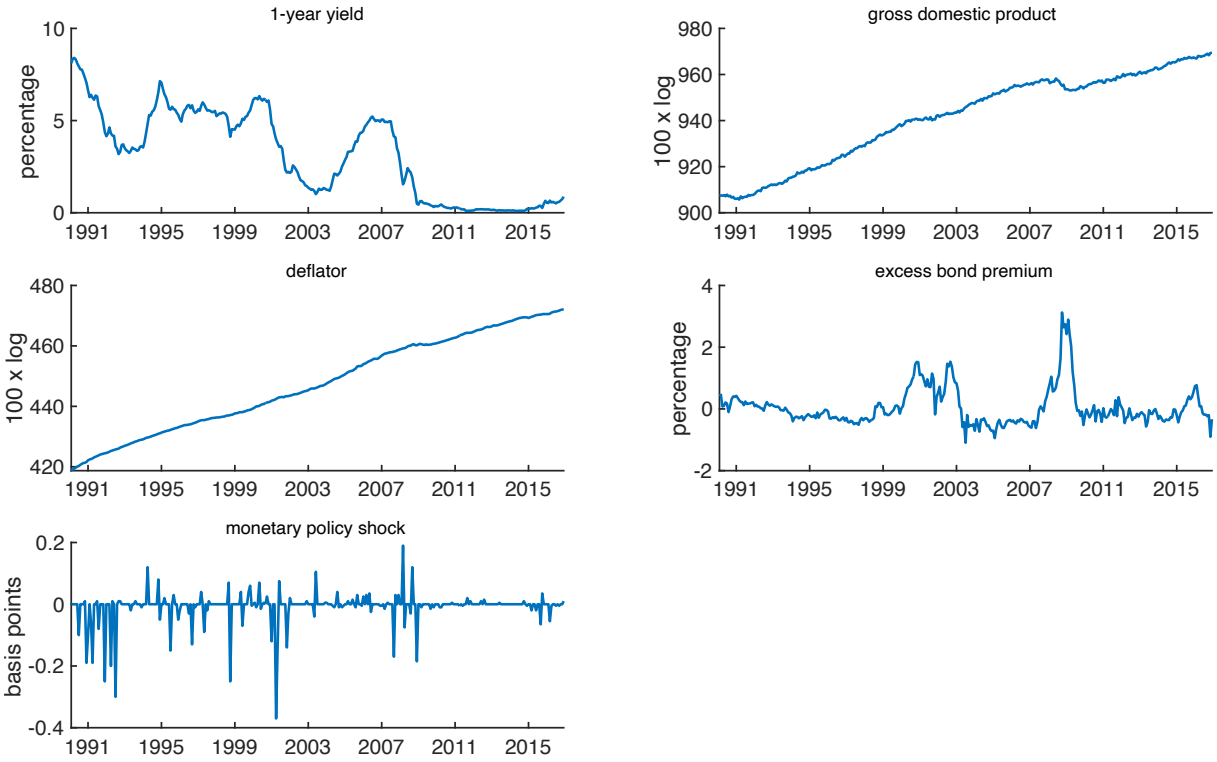
# Appendix B Data

Figure B.1
DGP1 and DGP2: endogenous variables



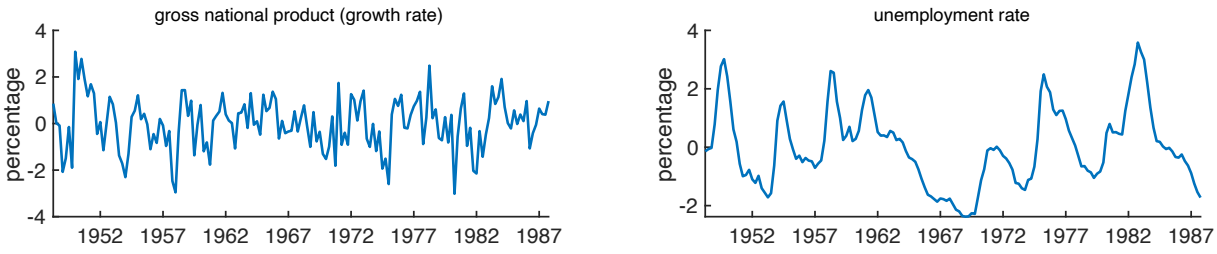

*Notes: Blue solid lines depict the actual time series of the corresponding variable.*

Figure B.2
DGP3: endogenous variables




*Notes: Blue solid lines depict the actual time series of the corresponding variable.*

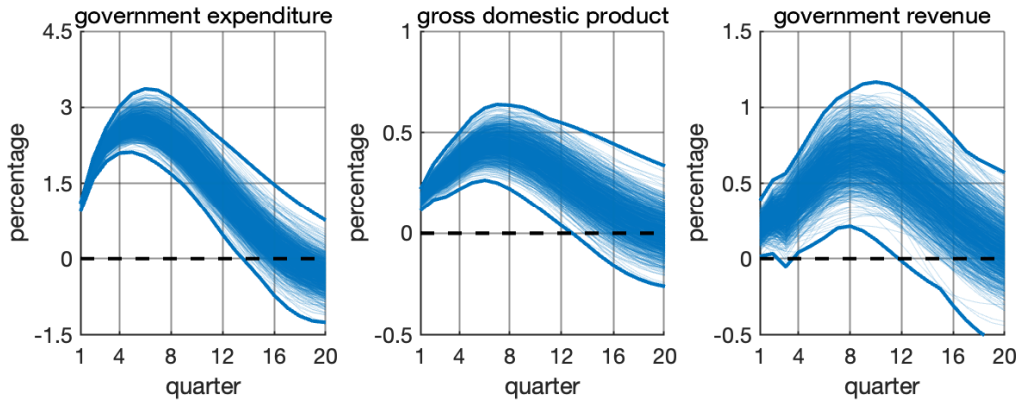
Figure B.3
DGP4: endogenous variables




*Notes: Blue solid lines depict the actual time series of the corresponding variable.*

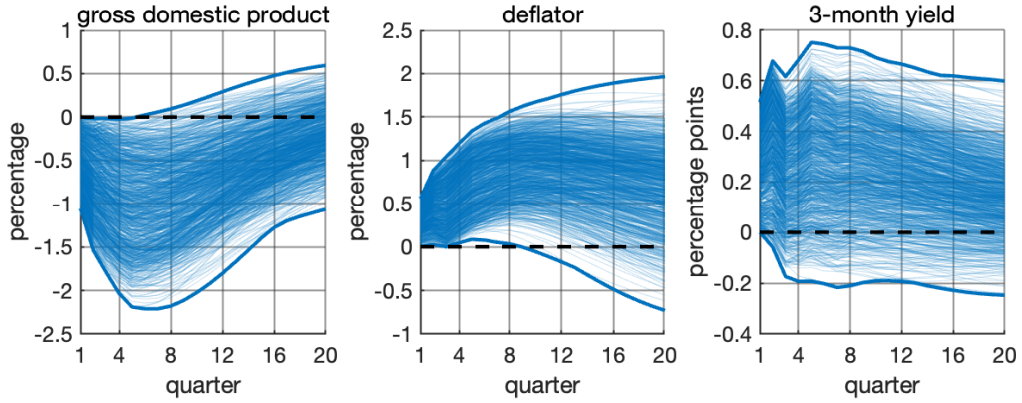
# Appendix C Model Simulations

Figure C.1
DGP1: simulated IRFs



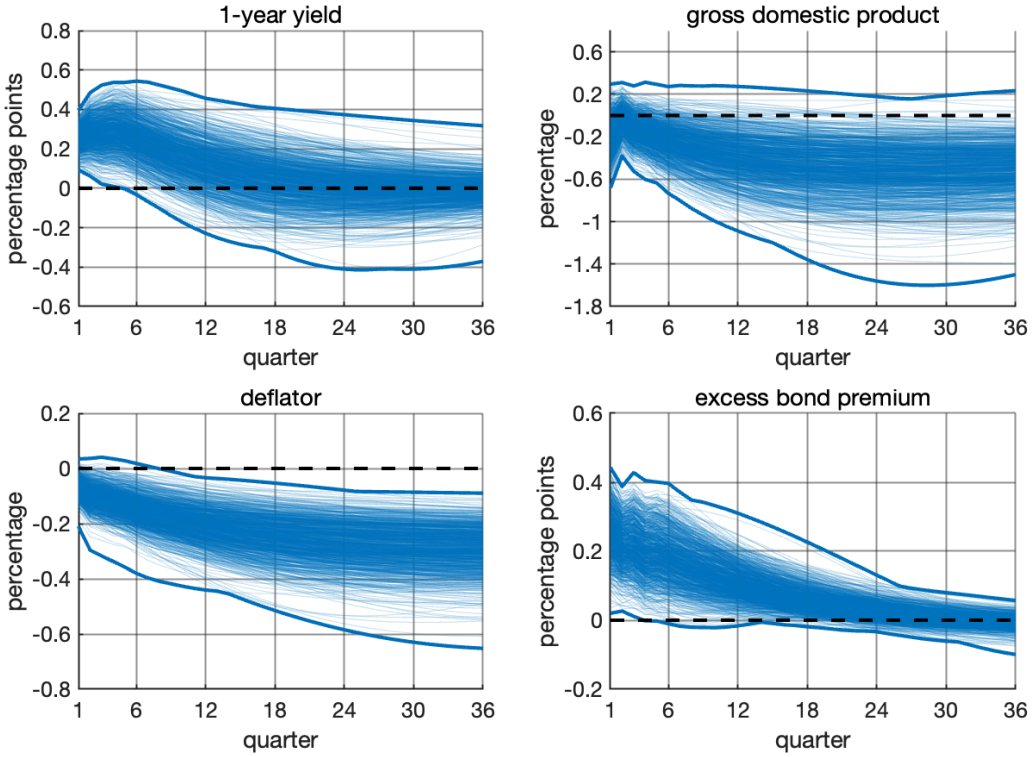

*Notes: Blue solid lines depict IRFs simulated from DGP1.*

Figure C.2
DGP2: simulated IRFs



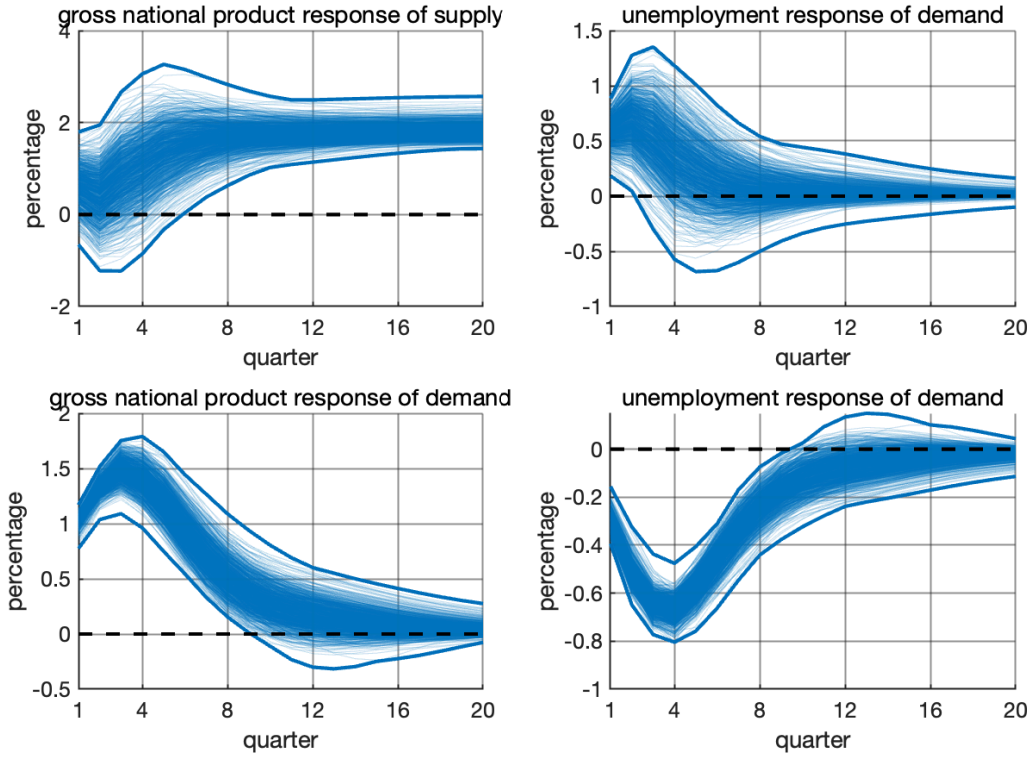

*Notes: Blue solid lines depict IRFs simulated from DGP2.*

Figure C.3
DGP3: simulated IRFs




*Notes: Blue solid lines depict IRFs simulated from DGP3.*

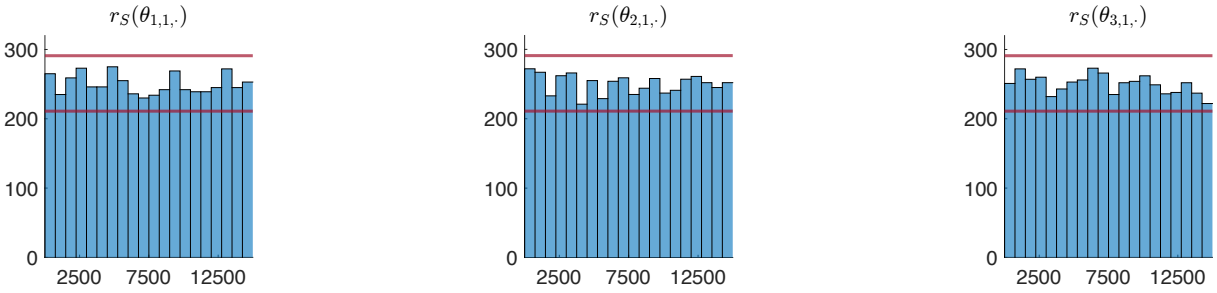
Figure C.4
DGP4: simulated IRFs



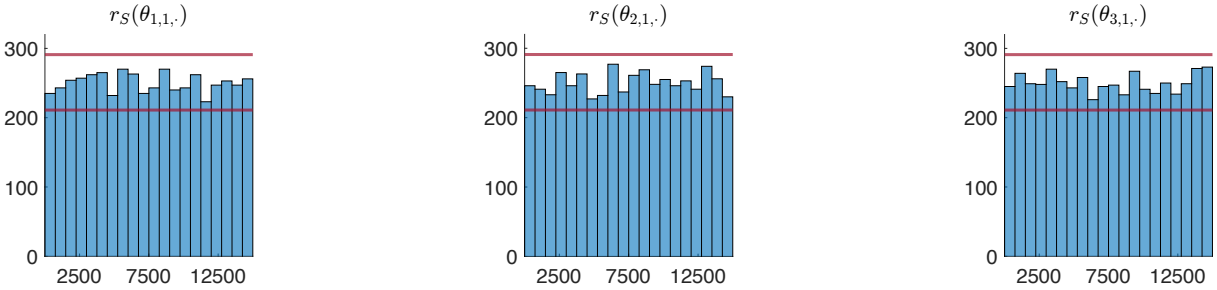

*Notes: Blue solid lines depict IRFs simulated from DGP4.*

# Appendix D Shape and Comovement

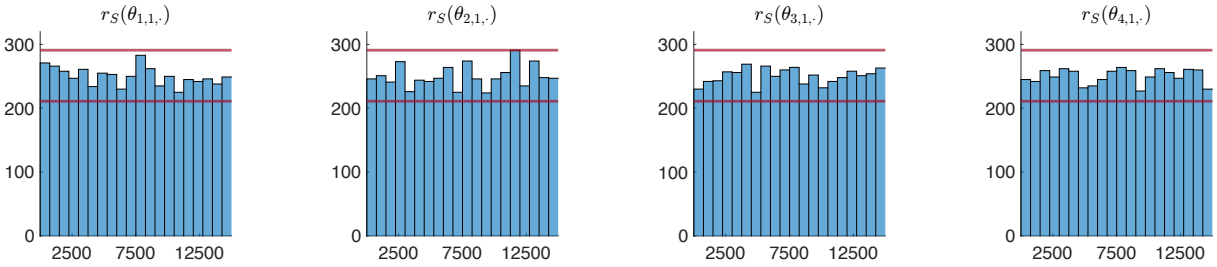
Figure D.1
Average ranks ($r_S$) of the TS vectors



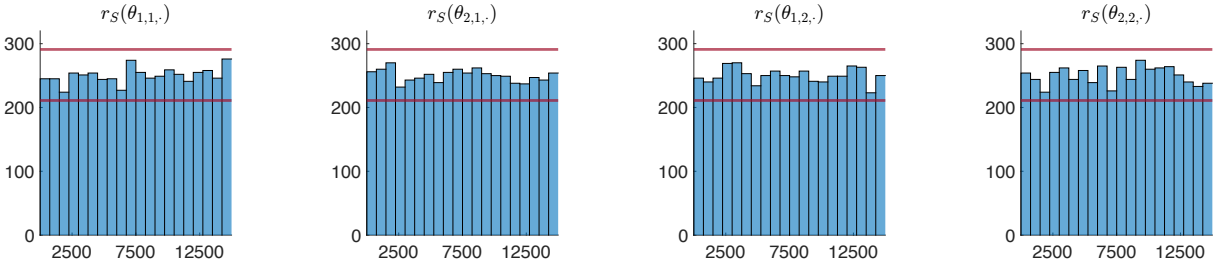

(a) DGP1




(b) DGP2



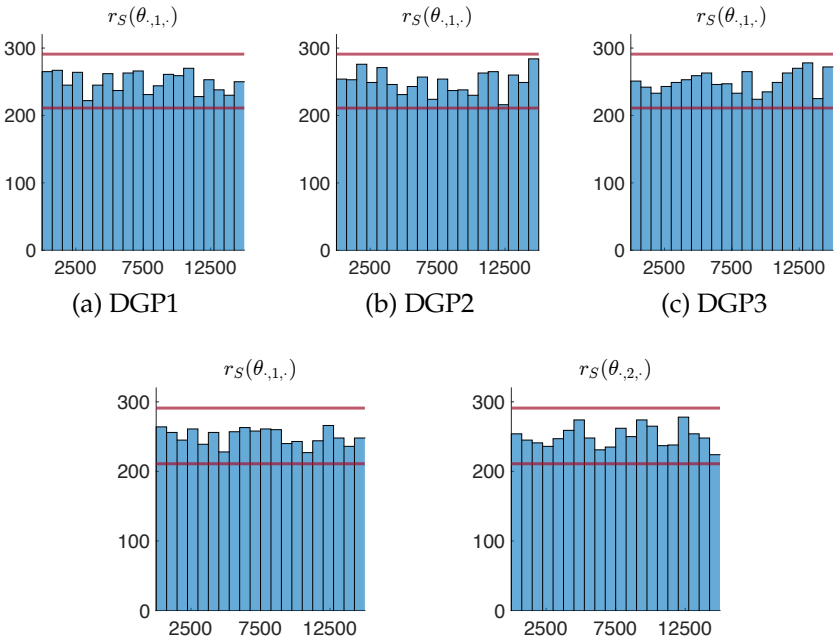

(c) DGP3




(d) DGP4

*Notes: Blue bars denote the average ranks ($r_S$). The red solid lines denote the 0.005 and 0.995 percentiles of Binomial*$(5000, 1/20)$ *distribution.*

Figure D.2
Average ranks ($r_S$) of the MV-TS vectors

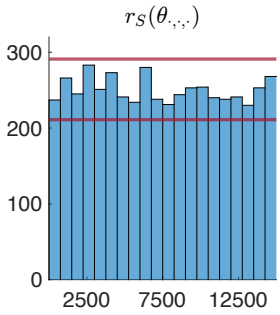



(a) DGP1 (b) DGP2 (c) DGP3

(d) DGP4

*Notes: Blue bars denote the average ranks ($r_S$). The red solid lines denote the 0.005 and 0.995 percentiles of Binomial*$(5000, 1/20)$ *distribution.*

Figure D.3
Average ranks ($r_S$) of the MV-TS-CS vectors




(a) DGP4

*Notes: Blue bars denote the average ranks ($r_S$). The red solid lines denote the 0.005 and 0.995 percentiles of Binomial*$(5000, 1/20)$ *distribution.*

# Appendix E Out-of-sample Forecasting Exercise

Table E.1
CCK18: pointwise ($\Pi^{PW}$) and time-simultaneous ($\Pi^{TS}$) coverage levels of PW error bands for United States

| Coverage level | 30% PW error band | | | 60% PW error band | | | 90% PW error band | | |
|---|---|---|---|---|---|---|---|---|---|
| | h = 2 | h = 4 | h = 8 | h = 2 | h = 4 | h = 8 | h = 2 | h = 4 | h = 8 |
| $\Pi^{PW}$ | 35,90 | 46,15 | 48,72 | 74,36 | 82,05 | 69,23 | 92,31 | 92,31 | 84,62 |
| $\Pi^{TS}$ | 10,26 | 2,56 | 0,00 | 53,85 | 38,46 | 15,38 | 89,74 | 82,05 | 61,54 |

*Notes: Pointwise ($\Pi^{MV}$) and time-simultaneous ($\Pi^{TS}$) coverage levels of 68% and 90% PW error bands for 2, 4, and 8 quarters ahead out-of-sample headline inflation forecast in United States using CCK18.*

Table E.2
CCK18: pointwise ($\Pi^{PW}$) and time-simultaneous ($\Pi^{TS}$) coverage levels of PW error bands for euro area

| Coverage level | 30% PW error band | | | 60% PW error band | | | 90% PW error band | | |
|---|---|---|---|---|---|---|---|---|---|
| | h = 2 | h = 4 | h = 8 | h = 2 | h = 4 | h = 8 | h = 2 | h = 4 | h = 8 |
| $\Pi^{PW}$ | 20,51 | 30,77 | 30,77 | 61,54 | 53,85 | 51,28 | 94,87 | 94,87 | 82,05 |
| $\Pi^{TS}$ | 10,26 | 2,56 | 0,00 | 35,90 | 12,82 | 5,13 | 87,18 | 79,49 | 48,72 |

*Notes: Pointwise ($\Pi^{PW}$) and time-simultaneous ($\Pi^{TS}$) coverage levels of 68% and 90% PW error bands for 2, 4, and 8 quarters ahead out-of-sample headline inflation forecast in the euro area using CCK18.*

Table E.3
CCK18: time-simultaneous coverage levels ($\Pi^{TS}$) and average relative widths ($\Delta^{TS}$) of TS error bands for United States

| Method | 30% TS error band | | | | | | 60% TS error band | | | | | | 90% TS error band | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | h = 2 | | h = 4 | | h = 8 | | h = 2 | | h = 4 | | h = 8 | | h = 2 | | h = 4 | | h = 8 | |
| | Π | Δ | Π | Δ | Π | Δ | Π | Δ | Π | Δ | Π | Δ | Π | Δ | Π | Δ | Π | Δ |
| Bonferroni | 69,23 | 150,53 | 79,49 | 277,81 | 61,54 | 402,48 | 89,74 | 57,36 | 82,05 | 111,04 | 66,67 | 165,82 | 92,31 | 23,69 | 84,62 | 47,61 | 79,49 | 74,29 |
| Šidák | 46,15 | 98,75 | 66,67 | 207,18 | 61,54 | 320,69 | 82,05 | 48,15 | 79,49 | 97,35 | 61,54 | 149,90 | 92,31 | 22,87 | 84,62 | 46,39 | 79,49 | 72,65 |
| sup-t (PQO) | 43,59 | 87,74 | 61,54 | 176,93 | 61,54 | 262,24 | 82,05 | 42,59 | 79,49 | 84,56 | 61,54 | 125,44 | 92,31 | 20,51 | 84,62 | 41,01 | 76,92 | 62,25 |
| min-max: absolute loss (median) | 74,36 | 160,78 | 82,05 | 393,17 | 69,23 | 554,43 | 89,74 | 98,94 | 84,62 | 219,72 | 82,05 | 289,94 | 94,87 | 67,20 | 100,00 | 120,21 | 94,87 | 149,65 |
| min-max (LQO): absolute loss (median) | 43,59 | 87,37 | 61,54 | 178,51 | 58,97 | 265,51 | 79,49 | 42,71 | 79,49 | 84,24 | 61,54 | 129,01 | 89,74 | 20,57 | 84,62 | 41,80 | 79,49 | 65,26 |
| min-max (BDR): absolute loss (mean) | 43,59 | 87,44 | 58,97 | 176,70 | 61,54 | 261,57 | 82,05 | 42,47 | 79,49 | 84,10 | 61,54 | 125,29 | 92,31 | 20,50 | 84,62 | 40,99 | 74,36 | 62,78 |
| min-max: quadratic loss (mean) | 48,72 | 109,75 | 79,49 | 250,79 | 61,54 | 394,66 | 89,74 | 59,05 | 82,05 | 131,54 | 69,23 | 205,40 | 94,87 | 33,76 | 87,18 | 73,08 | 89,74 | 109,21 |
| min-max (LQO): quadratic loss (mean) | 43,59 | 87,56 | 58,97 | 177,11 | 58,97 | 263,23 | 84,62 | 42,48 | 79,49 | 84,49 | 61,54 | 126,34 | 89,74 | 20,63 | 84,62 | 40,88 | 76,92 | 63,80 |
| min-max (BDR): quadratic loss (median) | 43,59 | 87,53 | 58,97 | 176,62 | 61,54 | 261,34 | 82,05 | 42,47 | 79,49 | 84,20 | 61,54 | 125,36 | 89,74 | 20,49 | 84,62 | 40,97 | 74,36 | 62,61 |
| min-max: angular loss (median) | 84,62 | 745,85 | 71,79 | 754,24 | 61,54 | 856,82 | 84,62 | 333,78 | 79,49 | 361,90 | 69,23 | 439,05 | 92,31 | 189,24 | 87,18 | 200,65 | 100,00 | 226,73 |
| min-max (LQO): angular loss (median) | 76,92 | 456,24 | 53,85 | 408,06 | 46,15 | 620,34 | 79,49 | 152,54 | 74,36 | 300,67 | 61,54 | 321,41 | 92,31 | 128,58 | 84,62 | 142,26 | 69,23 | 149,94 |
| min-max (BDR): angular loss (median) | 38,46 | 116,17 | 46,15 | 250,05 | 51,28 | 335,85 | 79,49 | 75,80 | 74,36 | 124,34 | 61,54 | 151,37 | 89,74 | 29,16 | 84,62 | 47,31 | 71,79 | 67,51 |
| min-max: Chebyshev loss (median) | 43,59 | 87,73 | 58,97 | 177,02 | 58,97 | 262,34 | 82,05 | 42,45 | 79,49 | 84,42 | 61,54 | 125,52 | 92,31 | 20,46 | 84,62 | 40,93 | 76,92 | 62,80 |
| min-max (LQO): Chebyshev loss (median) | 43,59 | 87,73 | 58,97 | 177,02 | 58,97 | 262,34 | 82,05 | 42,45 | 79,49 | 84,42 | 61,54 | 125,52 | 92,31 | 20,46 | 84,62 | 40,93 | 76,92 | 62,80 |
| min-max (BDR): Chebyshev loss (median) | 43,59 | 87,73 | 58,97 | 177,02 | 58,97 | 262,34 | 82,05 | 42,45 | 79,49 | 84,42 | 61,54 | 125,52 | 92,31 | 20,46 | 84,62 | 40,93 | 76,92 | 62,80 |

*Notes: Time-simultaneous coverage levels ($\Pi^{TS}$) and average relative widths ($\Delta^{TS}$) of 68% and 90% TS error bands for 2, 4, and 8 quarters ahead out-of-sample headline inflation forecast in United States using CCK18. In the table, the columns are labeled $\Pi$ and $\Delta$ for simplicity, but they represent $\Pi^{TS}$ and $\Delta^{TS}$, respectively. The 'median' or 'mean' indicated in parentheses next to each min-max estimator denotes the pointwise measure of central tendency used for the corresponding loss function.*

Table E.4
CCK18: time-simultaneous coverage levels ($\Pi^{TS}$) and average relative widths ($\Delta^{TS}$) of TS error bands for euro area

| Method | 30% TS error band | | | | | | 60% TS error band | | | | | | 90% TS error band | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | h = 2 | | h = 4 | | h = 8 | | h = 2 | | h = 4 | | h = 8 | | h = 2 | | h = 4 | | h = 8 | |
| | Π | Δ | Π | Δ | Π | Δ | Π | Δ | Π | Δ | Π | Δ | Π | Δ | Π | Δ | Π | Δ |
| Bonferroni | 46,15 | 146,03 | 43,59 | 267,14 | 56,41 | 381,51 | 61,54 | 55,32 | 79,49 | 105,20 | 82,05 | 155,41 | 97,44 | 21,92 | 100,00 | 43,50 | 97,44 | 68,16 |
| Šidák | 30,77 | 96,43 | 30,77 | 200,92 | 38,46 | 307,71 | 56,41 | 46,38 | 76,92 | 92,65 | 79,49 | 140,72 | 97,44 | 21,08 | 100,00 | 42,46 | 97,44 | 66,60 |
| sup-t (PQO) | 25,64 | 83,24 | 25,64 | 166,72 | 30,77 | 247,34 | 53,85 | 39,51 | 53,85 | 77,26 | 53,85 | 114,74 | 97,44 | 18,30 | 100,00 | 35,93 | 97,44 | 54,12 |
| min-max: absolute loss (median) | 41,03 | 153,96 | 79,49 | 370,35 | 87,18 | 513,74 | 79,49 | 94,99 | 100,00 | 203,17 | 97,44 | 254,94 | 100,00 | 66,71 | 100,00 | 106,73 | 97,44 | 124,65 |
| min-max (LQO): absolute loss (median) | 25,64 | 82,57 | 23,08 | 165,36 | 30,77 | 254,46 | 53,85 | 39,49 | 53,85 | 77,66 | 51,28 | 116,30 | 100,00 | 19,12 | 100,00 | 37,33 | 94,87 | 56,64 |
| min-max (BDR): absolute loss (mean) | 25,64 | 82,99 | 23,08 | 165,86 | 33,33 | 246,09 | 53,85 | 39,35 | 58,97 | 76,52 | 48,72 | 113,85 | 97,44 | 18,31 | 100,00 | 35,96 | 97,44 | 53,62 |
| min-max: quadratic loss (mean) | 30,77 | 104,68 | 35,90 | 236,90 | 51,28 | 374,39 | 58,97 | 55,98 | 87,18 | 123,57 | 92,31 | 190,95 | 100,00 | 33,29 | 100,00 | 69,16 | 97,44 | 99,72 |
| min-max (LQO): quadratic loss (mean) | 25,64 | 82,68 | 25,64 | 166,33 | 28,21 | 247,51 | 53,85 | 39,37 | 56,41 | 76,85 | 53,85 | 114,31 | 100,00 | 19,01 | 100,00 | 36,68 | 97,44 | 54,99 |
| min-max (BDR): quadratic loss (median) | 25,64 | 82,84 | 23,08 | 165,85 | 28,21 | 246,32 | 53,85 | 39,36 | 58,97 | 76,59 | 51,28 | 113,85 | 100,00 | 18,50 | 100,00 | 36,14 | 97,44 | 53,51 |
| min-max: angular loss (median) | 61,54 | 652,78 | 41,03 | 677,53 | 28,21 | 787,46 | 66,67 | 278,04 | 53,85 | 303,64 | 79,49 | 375,43 | 97,44 | 134,80 | 100,00 | 159,17 | 100,00 | 183,21 |
| min-max (LQO): angular loss (median) | 56,41 | 405,78 | 28,21 | 370,68 | 12,82 | 609,20 | 58,97 | 142,08 | 41,03 | 239,31 | 30,77 | 306,45 | 84,62 | 100,00 | 94,87 | 116,02 | 87,18 | 135,11 |
| min-max (BDR): angular loss (median) | 25,64 | 105,01 | 25,64 | 226,90 | 20,51 | 323,87 | 53,85 | 62,40 | 41,03 | 113,42 | 43,59 | 144,94 | 84,62 | 31,72 | 94,87 | 44,57 | 97,44 | 60,26 |
| min-max: Chebyshev loss (median) | 25,64 | 83,35 | 25,64 | 166,70 | 28,21 | 247,84 | 53,85 | 39,44 | 53,85 | 76,88 | 51,28 | 114,52 | 97,44 | 18,29 | 100,00 | 35,90 | 97,44 | 54,03 |
| min-max (LQO): Chebyshev loss (median) | 25,64 | 83,35 | 25,64 | 166,70 | 28,21 | 247,84 | 53,85 | 39,44 | 53,85 | 76,88 | 51,28 | 114,52 | 97,44 | 18,29 | 100,00 | 35,90 | 97,44 | 54,03 |
| min-max (BDR): Chebyshev loss (median) | 25,64 | 83,35 | 25,64 | 166,70 | 28,21 | 247,84 | 53,85 | 39,44 | 53,85 | 76,88 | 51,28 | 114,52 | 97,44 | 18,29 | 100,00 | 35,90 | 97,44 | 54,03 |

*Notes: Time-simultaneous coverage levels ($\Pi^{TS}$) and average relative widths ($\Delta^{TS}$) of 68% and 90% TS error bands for 2, 4, and 8 quarters ahead out-of-sample headline inflation forecast in the euro area using CCK18. In the table, the columns are labeled $\Pi$ and $\Delta$ for simplicity, but they represent $\Pi^{TS}$ and $\Delta^{TS}$, respectively. The 'median' or 'mean' indicated in parentheses next to each min-max estimator denotes the pointwise measure of central tendency used for the corresponding loss function.*

Table E.5
DP15: pointwise ($\Pi^{PW}$), multivariate ($\Pi^{MV}$), and time-simultaneous ($\Pi^{TS}$) coverage levels of PW error bands for United States

| Coverage level | Variable | 30% PW error band | | | 60% PW error band | | | 90% PW error band | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | h = 2 | h = 4 | h = 8 | h = 2 | h = 4 | h = 8 | h = 2 | h = 4 | h = 8 |
| $\Pi^{PW}$ | $y_1$ | 33,33 | 30,77 | 30,77 | 69,23 | 74,36 | 66,67 | 94,87 | 92,31 | 84,62 |
| | $y_2$ | 41,03 | 48,72 | 46,15 | 76,92 | 74,36 | 74,36 | 97,44 | 92,31 | 89,74 |
| $\Pi^{MV}$ | - | 10,26 | 17,95 | 15,38 | 51,28 | 56,41 | 56,41 | 92,31 | 87,18 | 79,49 |
| $\Pi^{TS}$ | $y_1$ | 7,69 | 2,56 | 0,00 | 48,72 | 28,21 | 17,95 | 92,31 | 84,62 | 64,10 |
| | $y_2$ | 15,38 | 5,13 | 0,00 | 58,97 | 35,90 | 15,38 | 92,31 | 84,62 | 71,79 |

*Notes: Pointwise ($\Pi^{PW}$), multivariate ($\Pi^{MV}$), and time-simultaneous ($\Pi^{TS}$) coverage levels of 68% and 90% PW error bands for 2, 4, and 8 quarters ahead out-of-sample headline inflation and real GDP growth rate forecast in United States using DP15.*

Table E.6
DP15: pointwise ($\Pi^{PW}$), multivariate ($\Pi^{MV}$), and time-simultaneous ($\Pi^{TS}$) coverage levels of PW error bands for euro area

| Coverage level | Variable | 30% PW error band | | | 60% PW error band | | | 90% PW error band | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | h = 2 | h = 4 | h = 8 | h = 2 | h = 4 | h = 8 | h = 2 | h = 4 | h = 8 |
| $\Pi^{PW}$ | $y_1$ | 28,21 | 25,64 | 23,08 | 53,85 | 58,97 | 56,41 | 84,62 | 79,49 | 82,05 |
| | $y_2$ | 33,33 | 48,72 | 43,59 | 66,67 | 89,74 | 87,18 | 100,00 | 94,87 | 89,74 |
| $\Pi^{MV}$ | - | 15,38 | 17,95 | 7,69 | 41,03 | 53,85 | 51,28 | 84,62 | 76,92 | 74,36 |
| $\Pi^{TS}$ | $y_1$ | 10,26 | 0,00 | 0,00 | 35,90 | 17,95 | 2,56 | 74,36 | 56,41 | 38,46 |
| | $y_2$ | 10,26 | 0,00 | 0,00 | 48,72 | 43,59 | 35,90 | 94,87 | 89,74 | 79,49 |

*Notes: Pointwise ($\Pi^{PW}$), multivariate ($\Pi^{MV}$), and time-simultaneous ($\Pi^{TS}$) coverage levels of 68% and 90% PW error bands for 2, 4, and 8 quarters ahead out-of-sample headline inflation and real GDP growth rate forecast in the euro area using DP15.*

## Table E.7
## DP15: time-simultaneous coverage levels ($\Pi^{TS}$) and average relative widths ($\Delta^{TS}$) of TS error bands for United States

| Method | Variable | 30% TS error band | | | | | | 60% TS error band | | | | | | 90% TS error band | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | h = 2 | | h = 4 | | h = 8 | | h = 2 | | h = 4 | | h = 8 | | h = 2 | | h = 4 | | h = 8 | |
| | | $\Pi$ | $\Delta$ | $\Pi$ | $\Delta$ | $\Pi$ | $\Delta$ | $\Pi$ | $\Delta$ | $\Pi$ | $\Delta$ | $\Pi$ | $\Delta$ | $\Pi$ | $\Delta$ | $\Pi$ | $\Delta$ | $\Pi$ | $\Delta$ |
| Bonferroni | $y_1$ | 56,41 | 155,54 | 74,36 | 300,36 | 64,10 | 478,17 | 74,36 | 61,76 | 84,62 | 125,42 | 69,23 | 202,71 | 97,44 | 26,52 | 89,74 | 57,01 | 87,18 | 95,93 |
| | $y_2$ | 66,67 | 145,34 | 74,36 | 264,79 | 74,36 | 377,86 | 84,62 | 54,59 | 84,62 | 103,72 | 82,05 | 154,25 | 97,44 | 21,35 | 94,87 | 42,43 | 84,62 | 66,35 |
| Šidák | $y_1$ | 66,67 | 145,34 | 74,36 | 264,79 | 74,36 | 377,86 | 84,62 | 54,59 | 84,62 | 103,72 | 82,05 | 154,25 | 97,44 | 21,35 | 94,87 | 42,43 | 84,62 | 66,35 |
| | $y_2$ | 35,90 | 100,98 | 51,28 | 221,37 | 61,54 | 369,93 | 71,79 | 51,20 | 82,05 | 109,33 | 69,23 | 181,32 | 97,44 | 25,46 | 89,74 | 55,35 | 87,18 | 94,09 |
| sup-t (PQO) | $y_1$ | 35,90 | 100,98 | 51,28 | 221,37 | 61,54 | 369,93 | 71,79 | 51,20 | 82,05 | 109,33 | 69,23 | 181,32 | 97,44 | 25,46 | 89,74 | 55,35 | 87,18 | 94,09 |
| | $y_2$ | 48,72 | 95,86 | 58,97 | 198,18 | 61,54 | 305,32 | 82,05 | 45,94 | 79,49 | 91,67 | 79,49 | 139,51 | 97,44 | 20,60 | 94,87 | 41,29 | 84,62 | 64,51 |
| min-max: absolute loss (median) | $y_1$ | 48,72 | 95,86 | 58,97 | 198,18 | 61,54 | 305,32 | 82,05 | 45,94 | 79,49 | 91,67 | 79,49 | 139,51 | 97,44 | 20,60 | 94,87 | 41,29 | 84,62 | 64,51 |
| | $y_2$ | 33,33 | 83,62 | 43,59 | 166,87 | 38,46 | 255,61 | 66,67 | 41,30 | 79,49 | 83,77 | 64,10 | 131,12 | 97,44 | 21,19 | 89,74 | 44,31 | 82,05 | 72,13 |
| min-max (LQO): absolute loss (median) | $y_1$ | 33,33 | 83,62 | 43,59 | 166,87 | 38,46 | 255,61 | 66,67 | 41,30 | 79,49 | 83,77 | 64,10 | 131,12 | 97,44 | 21,19 | 89,74 | 44,31 | 82,05 | 72,13 |
| | $y_2$ | 46,15 | 88,31 | 56,41 | 178,73 | 56,41 | 269,55 | 79,49 | 42,12 | 79,49 | 83,34 | 79,49 | 123,84 | 94,87 | 19,05 | 94,87 | 37,89 | 84,62 | 57,56 |
| min-max (BDR): absolute loss (mean) | $y_1$ | 46,15 | 88,31 | 56,41 | 178,73 | 56,41 | 269,55 | 79,49 | 42,12 | 79,49 | 83,34 | 79,49 | 123,84 | 94,87 | 19,05 | 94,87 | 37,89 | 84,62 | 57,56 |
| | $y_2$ | 48,72 | 149,42 | 79,49 | 332,39 | 66,67 | 458,90 | 89,74 | 94,10 | 89,74 | 188,08 | 76,92 | 252,02 | 97,44 | 64,55 | 97,44 | 106,94 | 89,74 | 138,85 |
| min-max: quadratic loss (mean) | $y_1$ | 48,72 | 149,42 | 79,49 | 332,39 | 66,67 | 458,90 | 89,74 | 94,10 | 89,74 | 188,08 | 76,92 | 252,02 | 97,44 | 64,55 | 97,44 | 106,94 | 89,74 | 138,85 |
| | $y_2$ | 66,67 | 154,87 | 84,62 | 360,98 | 84,62 | 495,28 | 92,31 | 93,42 | 94,87 | 192,81 | 84,62 | 249,33 | 100,00 | 60,75 | 94,87 | 97,00 | 84,62 | 114,48 |
| min-max (LQO): quadratic loss (mean) | $y_1$ | 66,67 | 154,87 | 84,62 | 360,98 | 84,62 | 495,28 | 92,31 | 93,42 | 94,87 | 192,81 | 84,62 | 249,33 | 100,00 | 60,75 | 94,87 | 97,00 | 84,62 | 114,48 |
| | $y_2$ | 33,33 | 82,57 | 48,72 | 166,32 | 43,59 | 258,15 | 71,79 | 41,54 | 71,79 | 84,48 | 66,67 | 134,61 | 94,87 | 21,85 | 89,74 | 45,61 | 79,49 | 75,63 |
| min-max (BDR): quadratic loss (median) | $y_1$ | 33,33 | 82,57 | 48,72 | 166,32 | 43,59 | 258,15 | 71,79 | 41,54 | 71,79 | 84,48 | 66,67 | 134,61 | 94,87 | 21,85 | 89,74 | 45,61 | 79,49 | 75,63 |
| | $y_2$ | 48,72 | 87,62 | 51,28 | 179,32 | 56,41 | 275,40 | 76,92 | 41,95 | 79,49 | 83,72 | 76,92 | 128,25 | 97,44 | 19,57 | 94,87 | 39,27 | 84,62 | 61,64 |
| min-max: angular loss (median) | $y_1$ | 48,72 | 87,62 | 51,28 | 179,32 | 56,41 | 275,40 | 76,92 | 41,95 | 79,49 | 83,72 | 76,92 | 128,25 | 97,44 | 19,57 | 94,87 | 39,27 | 84,62 | 61,64 |
| | $y_2$ | 33,33 | 82,53 | 48,72 | 164,73 | 41,03 | 251,72 | 69,23 | 40,91 | 79,49 | 83,24 | 66,67 | 129,39 | 94,87 | 20,82 | 89,74 | 43,97 | 84,62 | 70,67 |
| min-max (LQO): angular loss (median) | $y_1$ | 33,33 | 82,53 | 48,72 | 164,73 | 41,03 | 251,72 | 69,23 | 40,91 | 79,49 | 83,24 | 66,67 | 129,39 | 94,87 | 20,82 | 89,74 | 43,97 | 84,62 | 70,67 |
| | $y_2$ | 48,72 | 87,54 | 53,85 | 177,19 | 56,41 | 266,36 | 79,49 | 41,65 | 79,49 | 82,39 | 76,92 | 122,65 | 97,44 | 18,92 | 94,87 | 37,40 | 84,62 | 56,82 |
| min-max (BDR): angular loss (median) | $y_1$ | 48,72 | 87,54 | 53,85 | 177,19 | 56,41 | 266,36 | 79,49 | 41,65 | 79,49 | 82,39 | 76,92 | 122,65 | 97,44 | 18,92 | 94,87 | 37,40 | 84,62 | 56,82 |
| | $y_2$ | 35,90 | 103,76 | 56,41 | 228,38 | 64,10 | 351,47 | 74,36 | 56,67 | 87,18 | 125,31 | 69,23 | 191,30 | 97,44 | 34,96 | 92,31 | 71,93 | 87,18 | 105,89 |
| min-max: Chebyshev loss (median) | $y_1$ | 35,90 | 103,76 | 56,41 | 228,38 | 64,10 | 351,47 | 74,36 | 56,67 | 87,18 | 125,31 | 69,23 | 191,30 | 97,44 | 34,96 | 92,31 | 71,93 | 87,18 | 105,89 |
| | $y_2$ | 53,85 | 108,99 | 69,23 | 246,32 | 71,79 | 377,35 | 84,62 | 57,45 | 89,74 | 124,68 | 84,62 | 187,87 | 100,00 | 31,45 | 94,87 | 64,29 | 84,62 | 92,01 |
| min-max (LQO): Chebyshev loss (median) | $y_1$ | 53,85 | 108,99 | 69,23 | 246,32 | 71,79 | 377,35 | 84,62 | 57,45 | 89,74 | 124,68 | 84,62 | 187,87 | 100,00 | 31,45 | 94,87 | 64,29 | 84,62 | 92,01 |
| | $y_2$ | 30,77 | 82,80 | 46,15 | 166,15 | 43,59 | 253,66 | 71,79 | 41,15 | 79,49 | 83,32 | 66,67 | 130,48 | 94,87 | 21,39 | 89,74 | 44,57 | 82,05 | 73,92 |
| min-max (BDR): Chebyshev loss (median) | $y_1$ | 30,77 | 82,80 | 46,15 | 166,15 | 43,59 | 253,66 | 71,79 | 41,15 | 79,49 | 83,32 | 66,67 | 130,48 | 94,87 | 21,39 | 89,74 | 44,57 | 82,05 | 73,92 |
| | $y_2$ | 51,28 | 87,66 | 53,85 | 177,88 | 58,97 | 268,78 | 79,49 | 42,07 | 79,49 | 83,07 | 74,36 | 124,10 | 97,44 | 19,27 | 94,87 | 37,94 | 84,62 | 58,05 |

*Notes: Time-simultaneous coverage levels ($\Pi^{TS}$) and average relative widths ($\Delta^{TS}$) of 68% and 90% TS error bands for 2, 4, and 8 quarters ahead out-of-sample headline inflation and real GDP growth rate forecast in United States using DP15. In the table, the columns are labeled $\Pi$ and $\Delta$ for simplicity, but they represent $\Pi^{TS}$ and $\Delta^{TS}$, respectively. The 'median' or 'mean' indicated in parentheses next to each min-max estimator denotes the pointwise measure of central tendency used for the corresponding loss function.*

Table E.8
DP15: time-simultaneous coverage levels ($\Pi^{TS}$) and average relative widths ($\Delta^{TS}$) of TS error bands for euro area

| Method | Variable | 30% TS error band | | | | | | 60% TS error band | | | | | | 90% TS error band | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | h = 2 | | h = 4 | | h = 8 | | h = 2 | | h = 4 | | h = 8 | | h = 2 | | h = 4 | | h = 8 | |
| | | Π | Δ | Π | Δ | Π | Δ | Π | Δ | Π | Δ | Π | Δ | Π | Δ | Π | Δ | Π | Δ |
| Bonferroni | $y_1$ | 41,03 | 144,90 | 43,59 | 260,79 | 48,72 | 363,98 | 53,85 | 53,80 | 56,41 | 100,12 | 56,41 | 144,38 | 84,62 | 20,75 | 87,18 | 40,37 | 97,44 | 59,84 |
| | $y_2$ | 51,28 | 147,72 | 82,05 | 267,46 | 79,49 | 375,11 | 87,18 | 55,60 | 89,74 | 104,40 | 82,05 | 153,08 | 97,44 | 22,16 | 92,31 | 43,29 | 84,62 | 67,25 |
| Šidák | $y_1$ | 51,28 | 147,72 | 82,05 | 267,46 | 79,49 | 375,11 | 87,18 | 55,60 | 89,74 | 104,40 | 82,05 | 153,08 | 97,44 | 22,16 | 92,31 | 43,29 | 84,62 | 67,25 |
| | $y_2$ | 33,33 | 96,43 | 30,77 | 197,48 | 33,33 | 297,31 | 51,28 | 45,54 | 51,28 | 88,78 | 53,85 | 131,44 | 84,62 | 20,04 | 87,18 | 39,09 | 94,87 | 58,92 |
| sup-t (PQO) | $y_1$ | 33,33 | 96,43 | 30,77 | 197,48 | 33,33 | 297,31 | 51,28 | 45,54 | 51,28 | 88,78 | 53,85 | 131,44 | 84,62 | 20,04 | 87,18 | 39,09 | 94,87 | 58,92 |
| | $y_2$ | 41,03 | 98,51 | 64,10 | 200,65 | 76,92 | 304,13 | 82,05 | 46,58 | 89,74 | 91,94 | 82,05 | 139,18 | 97,44 | 21,19 | 92,31 | 42,09 | 84,62 | 66,03 |
| min-max: absolute loss (median) | $y_1$ | 41,03 | 98,51 | 64,10 | 200,65 | 76,92 | 304,13 | 82,05 | 46,58 | 89,74 | 91,94 | 82,05 | 139,18 | 97,44 | 21,19 | 92,31 | 42,09 | 84,62 | 66,03 |
| | $y_2$ | 33,33 | 90,31 | 30,77 | 178,52 | 25,64 | 263,82 | 51,28 | 42,53 | 46,15 | 80,65 | 46,15 | 117,51 | 84,62 | 18,92 | 84,62 | 35,88 | 92,31 | 53,41 |
| min-max (LQO): absolute loss (median) | $y_1$ | 33,33 | 90,31 | 30,77 | 178,52 | 25,64 | 263,82 | 51,28 | 42,53 | 46,15 | 80,65 | 46,15 | 117,51 | 84,62 | 18,92 | 84,62 | 35,88 | 92,31 | 53,41 |
| | $y_2$ | 33,33 | 82,25 | 48,72 | 163,07 | 64,10 | 240,99 | 71,79 | 39,09 | 87,18 | 75,51 | 79,49 | 110,99 | 97,44 | 18,29 | 92,31 | 34,54 | 82,05 | 52,29 |
| min-max (BDR): absolute loss (mean) | $y_1$ | 33,33 | 82,25 | 48,72 | 163,07 | 64,10 | 240,99 | 71,79 | 39,09 | 87,18 | 75,51 | 79,49 | 110,99 | 97,44 | 18,29 | 92,31 | 34,54 | 82,05 | 52,29 |
| | $y_2$ | 43,59 | 160,65 | 58,97 | 366,07 | 66,67 | 482,56 | 71,79 | 93,98 | 84,62 | 185,42 | 92,31 | 234,51 | 100,00 | 59,18 | 100,00 | 91,82 | 100,00 | 106,94 |
| min-max: quadratic loss (mean) | $y_1$ | 43,59 | 160,65 | 58,97 | 366,07 | 66,67 | 482,56 | 71,79 | 93,98 | 84,62 | 185,42 | 92,31 | 234,51 | 100,00 | 59,18 | 100,00 | 91,82 | 100,00 | 106,94 |
| | $y_2$ | 51,28 | 146,70 | 89,74 | 321,75 | 82,05 | 419,08 | 94,87 | 91,62 | 92,31 | 172,42 | 84,62 | 209,86 | 100,00 | 60,17 | 94,87 | 87,13 | 84,62 | 98,04 |
| min-max (LQO): quadratic loss (mean) | $y_1$ | 51,28 | 146,70 | 89,74 | 321,75 | 82,05 | 419,08 | 94,87 | 91,62 | 92,31 | 172,42 | 84,62 | 209,86 | 100,00 | 60,17 | 94,87 | 87,13 | 84,62 | 98,04 |
| | $y_2$ | 33,33 | 90,88 | 28,21 | 178,60 | 28,21 | 270,80 | 51,28 | 42,44 | 43,59 | 81,73 | 43,59 | 122,06 | 84,62 | 19,41 | 84,62 | 37,67 | 87,18 | 56,77 |
| min-max (BDR): quadratic loss (median) | $y_1$ | 33,33 | 90,88 | 28,21 | 178,60 | 28,21 | 270,80 | 51,28 | 42,44 | 43,59 | 81,73 | 43,59 | 122,06 | 84,62 | 19,41 | 84,62 | 37,67 | 87,18 | 56,77 |
| | $y_2$ | 38,46 | 81,58 | 56,41 | 162,52 | 74,36 | 240,22 | 71,79 | 39,09 | 92,31 | 78,92 | 82,05 | 114,18 | 97,44 | 18,52 | 94,87 | 38,42 | 84,62 | 56,04 |
| min-max: angular loss (median) | $y_1$ | 38,46 | 81,58 | 56,41 | 162,52 | 74,36 | 240,22 | 71,79 | 39,09 | 92,31 | 78,92 | 82,05 | 114,18 | 97,44 | 18,52 | 94,87 | 38,42 | 84,62 | 56,04 |
| | $y_2$ | 30,77 | 89,79 | 28,21 | 176,80 | 25,64 | 261,49 | 51,28 | 42,19 | 48,72 | 79,70 | 46,15 | 116,91 | 84,62 | 18,59 | 87,18 | 35,63 | 89,74 | 52,66 |
| min-max (LQO): angular loss (median) | $y_1$ | 30,77 | 89,79 | 28,21 | 176,80 | 25,64 | 261,49 | 51,28 | 42,19 | 48,72 | 79,70 | 46,15 | 116,91 | 84,62 | 18,59 | 87,18 | 35,63 | 89,74 | 52,66 |
| | $y_2$ | 38,46 | 81,57 | 51,28 | 160,23 | 74,36 | 233,94 | 71,79 | 38,41 | 87,18 | 74,40 | 82,05 | 109,01 | 97,44 | 17,84 | 94,87 | 35,00 | 84,62 | 50,07 |
| min-max (BDR): angular loss (median) | $y_1$ | 38,46 | 81,57 | 51,28 | 160,23 | 74,36 | 233,94 | 71,79 | 38,41 | 87,18 | 74,40 | 82,05 | 109,01 | 97,44 | 17,84 | 94,87 | 35,00 | 84,62 | 50,07 |
| | $y_2$ | 33,33 | 111,99 | 38,46 | 245,10 | 46,15 | 371,88 | 53,85 | 57,96 | 64,10 | 121,47 | 79,49 | 180,02 | 89,74 | 30,89 | 97,44 | 62,10 | 100,00 | 86,12 |
| min-max: Chebyshev loss (median) | $y_1$ | 33,33 | 111,99 | 38,46 | 245,10 | 46,15 | 371,88 | 53,85 | 57,96 | 64,10 | 121,47 | 79,49 | 180,02 | 89,74 | 30,89 | 97,44 | 62,10 | 100,00 | 86,12 |
| | $y_2$ | 43,59 | 102,16 | 82,05 | 222,97 | 79,49 | 334,61 | 79,49 | 54,97 | 92,31 | 115,77 | 84,62 | 168,24 | 97,44 | 33,34 | 94,87 | 62,22 | 84,62 | 81,39 |
| min-max (LQO): Chebyshev loss (median) | $y_1$ | 43,59 | 102,16 | 82,05 | 222,97 | 79,49 | 334,61 | 79,49 | 54,97 | 92,31 | 115,77 | 84,62 | 168,24 | 97,44 | 33,34 | 94,87 | 62,22 | 84,62 | 81,39 |
| | $y_2$ | 30,77 | 89,43 | 28,21 | 178,71 | 25,64 | 263,86 | 51,28 | 41,91 | 46,15 | 80,09 | 46,15 | 118,46 | 82,05 | 18,77 | 84,62 | 36,99 | 84,62 | 54,57 |
| min-max (BDR): Chebyshev loss (median) | $y_1$ | 30,77 | 89,43 | 28,21 | 178,71 | 25,64 | 263,86 | 51,28 | 41,91 | 46,15 | 80,09 | 46,15 | 118,46 | 82,05 | 18,77 | 84,62 | 36,99 | 84,62 | 54,57 |
| | $y_2$ | 38,46 | 81,79 | 58,97 | 161,10 | 74,36 | 238,14 | 74,36 | 39,02 | 89,74 | 76,42 | 82,05 | 112,03 | 97,44 | 18,87 | 94,87 | 37,72 | 84,62 | 54,66 |

*Notes: Time-simultaneous coverage levels ($\Pi^{TS}$) and average relative widths ($\Delta^{TS}$) of 68% and 90% TS error bands for 2, 4, and 8 quarters ahead out-of-sample headline inflation and real GDP growth rate forecast in the euro area using DP15. In the table, the columns are labeled $\Pi$ and $\Delta$ for simplicity, but they represent $\Pi^{TS}$ and $\Delta^{TS}$, respectively. The 'median' or 'mean' indicated in parentheses next to each min-max estimator denotes the pointwise measure of central tendency used for the corresponding loss function.*

Table E.9
DP15: multivariate coverage levels ($\Pi^{MV}$) and average relative widths ($\Delta^{MV}$) of MV error bands for United States

| Method | 30% TS error band | | | | | | 60% TS error band | | | | | | 90% TS error band | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | h=1:2 | | h=1:4 | | h=1:8 | | h=1:2 | | h=1:4 | | h=1:8 | | h=1:2 | | h=1:4 | | h=1:8 | |
| | Π | Δ | Π | Δ | Π | Δ | Π | Δ | Π | Δ | Π | Δ | Π | Δ | Π | Δ | Π | Δ |
| Bonferroni | 61,54 | 151,23 | 64,10 | 152,88 | 64,10 | 155,49 | 82,05 | 58,31 | 82,05 | 59,44 | 74,36 | 61,30 | 97,44 | 24,01 | 92,31 | 24,84 | 84,62 | 26,00 |
| Šidák | 43,59 | 99,01 | 46,15 | 99,98 | 43,59 | 101,60 | 79,49 | 48,79 | 74,36 | 49,69 | 74,36 | 51,16 | 97,44 | 23,12 | 92,31 | 23,99 | 84,62 | 25,13 |
| sup-t (PQO) | 43,59 | 93,83 | 46,15 | 93,49 | 43,59 | 93,75 | 79,49 | 45,72 | 74,36 | 46,27 | 71,79 | 46,93 | 97,44 | 21,93 | 92,31 | 22,45 | 84,62 | 23,52 |
| min-max: absolute loss (median) | 61,54 | 157,21 | 64,10 | 157,21 | 53,85 | 157,74 | 89,74 | 96,70 | 87,18 | 97,39 | 79,49 | 98,17 | 100,00 | 61,93 | 94,87 | 61,77 | 89,74 | 62,74 |
| min-max (LQO): absolute loss (median) | 51,28 | 89,52 | 51,28 | 90,40 | 46,15 | 91,39 | 76,92 | 48,67 | 74,36 | 48,74 | 76,92 | 49,83 | 94,87 | 28,12 | 92,31 | 27,89 | 87,18 | 28,27 |
| min-max (BDR): absolute loss (mean) | 48,72 | 91,26 | 51,28 | 91,21 | 46,15 | 92,14 | 79,49 | 46,95 | 76,92 | 47,24 | 74,36 | 47,94 | 94,87 | 23,33 | 92,31 | 23,61 | 84,62 | 24,20 |
| min-max: quadratic loss (mean) | 58,97 | 110,48 | 53,85 | 110,42 | 48,72 | 110,99 | 84,62 | 62,42 | 79,49 | 62,63 | 76,92 | 63,73 | 97,44 | 37,42 | 94,87 | 37,31 | 89,74 | 37,74 |
| min-max (LQO): quadratic loss (mean) | 51,28 | 89,85 | 51,28 | 90,29 | 46,15 | 90,76 | 76,92 | 48,53 | 74,36 | 48,61 | 76,92 | 49,52 | 97,44 | 29,38 | 92,31 | 28,62 | 89,74 | 28,61 |
| min-max (BDR): quadratic loss (median) | 51,28 | 90,23 | 51,28 | 90,18 | 46,15 | 91,24 | 76,92 | 47,79 | 74,36 | 48,11 | 76,92 | 48,74 | 94,87 | 26,35 | 92,31 | 26,10 | 87,18 | 26,28 |
| min-max: angular loss (median) | 100,00 | 1142,00 | 94,87 | 1241,78 | 92,31 | 1364,39 | 100,00 | 521,43 | 94,87 | 558,19 | 94,87 | 621,17 | 100,00 | 218,71 | 97,44 | 223,34 | 94,87 | 249,88 |
| min-max (LQO): angular loss (median) | 94,87 | 599,16 | 92,31 | 597,71 | 84,62 | 611,16 | 94,87 | 210,42 | 92,31 | 207,47 | 84,62 | 213,48 | 94,87 | 50,81 | 92,31 | 48,61 | 84,62 | 49,52 |
| min-max (BDR): angular loss (median) | 43,59 | 93,16 | 43,59 | 92,60 | 43,59 | 93,13 | 82,05 | 45,58 | 74,36 | 46,12 | 71,79 | 46,61 | 94,87 | 21,75 | 92,31 | 22,11 | 84,62 | 23,31 |
| min-max: Chebyshev loss (median) | 46,15 | 93,64 | 46,15 | 93,25 | 46,15 | 93,74 | 79,49 | 45,89 | 76,92 | 46,39 | 74,36 | 47,36 | 97,44 | 21,49 | 92,31 | 22,04 | 84,62 | 23,12 |
| min-max (LQO): Chebyshev loss (median) | 46,15 | 93,64 | 46,15 | 93,25 | 46,15 | 93,74 | 79,49 | 45,89 | 76,92 | 46,39 | 74,36 | 47,36 | 97,44 | 21,49 | 92,31 | 22,04 | 84,62 | 23,12 |
| min-max (BDR): Chebyshev loss (median) | 46,15 | 93,64 | 46,15 | 93,25 | 46,15 | 93,74 | 79,49 | 45,89 | 76,92 | 46,39 | 74,36 | 47,36 | 97,44 | 21,49 | 92,31 | 22,04 | 84,62 | 23,12 |

*Notes: Multivariate coverage levels ($\Pi^{MV}$) and average relative widths ($\Delta^{MV}$) of 68% and 90% MV error bands for 2, 4, and 8 quarters ahead out-of-sample headline inflation and real GDP growth rate forecast in United States using DP15. In the table, the columns are labeled $\Pi$ and $\Delta$ for simplicity, but they represent $\Pi^{MV}$ and $\Delta^{MV}$, respectively. The 'median' or 'mean' indicated in parentheses next to each min-max estimator denotes the pointwise measure of central tendency used for the corresponding loss function.*

Table E.10
DP15: multivariate coverage levels ($\Pi^{MV}$) and average relative widths ($\Delta^{MV}$) of MV error bands for euro area

| Method | 30% TS error band | | | | | | 60% TS error band | | | | | | 90% TS error band | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | h=1:2 | | h=1:4 | | h=1:8 | | h=1:2 | | h=1:4 | | h=1:8 | | h=1:2 | | h=1:4 | | h=1:8 | |
| | Π | Δ | Π | Δ | Π | Δ | Π | Δ | Π | Δ | Π | Δ | Π | Δ | Π | Δ | Π | Δ |
| Bonferroni | 48,72 | 147,32 | 58,97 | 146,10 | 53,85 | 147,08 | 66,67 | 54,91 | 71,79 | 54,88 | 64,10 | 55,44 | 89,74 | 21,30 | 87,18 | 21,57 | 82,05 | 22,20 |
| Šidák | 33,33 | 97,34 | 51,28 | 97,29 | 41,03 | 97,48 | 61,54 | 46,34 | 69,23 | 46,22 | 56,41 | 46,54 | 89,74 | 20,58 | 87,18 | 20,85 | 82,05 | 21,34 |
| sup-t (PQO) | 33,33 | 92,49 | 46,15 | 92,63 | 38,46 | 91,40 | 58,97 | 44,06 | 69,23 | 43,88 | 56,41 | 43,63 | 89,74 | 19,79 | 84,62 | 19,85 | 82,05 | 20,01 |
| min-max: absolute loss (median) | 53,85 | 151,73 | 66,67 | 150,88 | 56,41 | 149,08 | 84,62 | 93,02 | 92,31 | 92,15 | 89,74 | 92,32 | 100,00 | 58,15 | 94,87 | 57,21 | 89,74 | 56,29 |
| min-max (LQO): absolute loss (median) | 33,33 | 86,97 | 38,46 | 85,45 | 28,21 | 83,95 | 64,10 | 45,08 | 79,49 | 45,92 | 82,05 | 46,55 | 100,00 | 28,61 | 94,87 | 30,64 | 89,74 | 32,44 |
| min-max (BDR): absolute loss (mean) | 35,90 | 87,33 | 38,46 | 87,19 | 33,33 | 85,85 | 64,10 | 43,89 | 69,23 | 44,09 | 71,79 | 44,15 | 97,44 | 20,81 | 92,31 | 21,82 | 89,74 | 22,10 |
| min-max: quadratic loss (mean) | 43,59 | 106,96 | 53,85 | 105,63 | 38,46 | 103,69 | 69,23 | 58,92 | 84,62 | 59,41 | 82,05 | 59,70 | 100,00 | 37,25 | 94,87 | 38,28 | 89,74 | 39,85 |
| min-max (LQO): quadratic loss (mean) | 30,77 | 87,03 | 38,46 | 85,61 | 28,21 | 83,57 | 64,10 | 45,16 | 79,49 | 45,95 | 82,05 | 46,79 | 100,00 | 30,28 | 94,87 | 32,49 | 89,74 | 34,65 |
| min-max (BDR): quadratic loss (median) | 33,33 | 86,61 | 38,46 | 86,14 | 28,21 | 84,15 | 64,10 | 44,56 | 74,36 | 44,94 | 76,92 | 45,54 | 100,00 | 24,98 | 94,87 | 25,85 | 89,74 | 27,42 |
| min-max: angular loss (median) | 100,00 | 917,05 | 97,44 | 913,63 | 92,31 | 999,70 | 100,00 | 434,76 | 97,44 | 481,09 | 92,31 | 547,61 | 100,00 | 177,54 | 97,44 | 201,37 | 92,31 | 248,71 |
| min-max (LQO): angular loss (median) | 94,87 | 537,70 | 82,05 | 521,81 | 87,18 | 535,36 | 94,87 | 188,71 | 82,05 | 182,92 | 87,18 | 187,21 | 94,87 | 44,77 | 82,05 | 43,73 | 87,18 | 44,66 |
| min-max (BDR): angular loss (median) | 28,21 | 91,71 | 43,59 | 91,43 | 41,03 | 90,53 | 61,54 | 43,76 | 69,23 | 43,25 | 56,41 | 43,22 | 89,74 | 19,54 | 84,62 | 19,62 | 82,05 | 19,87 |
| min-max: Chebyshev loss (median) | 35,90 | 91,82 | 51,28 | 91,73 | 38,46 | 91,00 | 64,10 | 43,88 | 69,23 | 43,61 | 56,41 | 43,18 | 89,74 | 19,98 | 84,62 | 19,75 | 82,05 | 19,92 |
| min-max (LQO): Chebyshev loss (median) | 35,90 | 91,82 | 51,28 | 91,73 | 38,46 | 91,00 | 64,10 | 43,88 | 69,23 | 43,61 | 56,41 | 43,18 | 89,74 | 19,98 | 84,62 | 19,75 | 82,05 | 19,92 |
| min-max (BDR): Chebyshev loss (median) | 35,90 | 91,82 | 51,28 | 91,73 | 38,46 | 91,00 | 64,10 | 43,88 | 69,23 | 43,61 | 56,41 | 43,18 | 89,74 | 19,98 | 84,62 | 19,75 | 82,05 | 19,92 |

*Notes: Multivariate coverage levels ($\Pi^{MV}$) and average relative widths ($\Delta^{MV}$) of 68% and 90% MV error bands for 2, 4, and 8 quarters ahead out-of-sample headline inflation and real GDP growth rate forecast in the euro area using DP15. In the table, the columns are labeled $\Pi$ and $\Delta$ for simplicity, but they represent $\Pi^{MV}$ and $\Delta^{MV}$, respectively. The 'median' or 'mean' indicated in parentheses next to each min-max estimator denotes the pointwise measure of central tendency used for the corresponding loss function.*

**Acknowledgements**

I am grateful to Michael Binder for his continued guidance. I also thank Stefan Girstmair, Christian Mücke, an anonymous referee, and seminar participants at the European Central Bank, Deutsche Bundesbank, King's College & Bank of England's 2024 Workshop in Empirical Macroeconomics, University of East Anglia's 2nd Time Series Workshop, Bank of Albania's 17th South-Eastern European Economic Research Workshop, and Goethe University Frankfurt for helpful comments and suggestions. The routines for computing the joint inference methods discussed in this paper are implemented as a standalone toolbox "JFerence" in MATLAB available here.

**Andrian Yambolov**
Goethe University Frankfurt, Frankfurt am Main, Germany; email: yambolov.andrian@gmail.com



PDF ISBN 978-92-899-7409-7 ISSN 1725-2806 doi:10.2866/4530120 QB-01-25-182-EN-N