SFB 823

# Data-based priors for vector error correction models

Jan Prüser

Discussion Paper

Nr. 23/2020

# Data-Based Priors for Vector Error Correction Models*

Jan Prüser[a]

[a]TU Dortmund[†]

August 13, 2020

We propose two data-based priors for vector error correction models. Both priors lead to highly automatic approaches which require only minimal user input. An empirical investigation reveals that Bayesian vector error correction (BVEC) models equipped with our proposed priors turn out to scale well to higher dimensions and to forecast well. In addition, we find that exploiting information in the level variables has the potential for improving long-term forecasts. Thus, working with VARs in first differences may ignore valuable information. A simulation study reveals that it is beneficial, in terms of estimation accuracy, to use BVEC in the presence of cointegration. But if there is no cointegration, the proposed priors provide a sufficient amount of shrinkage so that the BVEC model has a similar estimation accuracy compared to the Bayesian vector autoregressive (BVAR) estimated in first differences.

**Keywords:** BVAR, Cointegration, Forecasting, Hierarchical Prior

**JEL classification:** C11, C32, C53

---

*I gratefully acknowledge the support of the German Research Foundation (DFG) through SFB 823, within Project A4. Moreover, I thank Boris Blagov, Rainer Schüssler, Christoph Hanck, Carsten Jentsch and Torsten Schmidt for valuable comments.

[†]Fakultät Statistik, 44221 Dortmund, Germany, e-mail: prueser@statistik.tu-dortmund.de

# 1. Introduction

BVARs have a long and successful history in macroeconomic forecasting. Some researchers estimate the BVAR with data in log-levels. They include enough lags and trust the estimation algorithm to clean up any unit roots in the errors (e.g., Doan et al. (1984), Litterman (1986), Sims and Zha (1998), Bańbura et al. (2010), Giannone et al. (2015)). More recently, BVARs estimated with data in log-differences have become more popular (e.g., Koop (2013), Carriero et al. (2016), Korobilis and Pettenuzzo (2019), Huber and Feldkircher (2019), Chan (2020) and Cross et al. (2020)). While a specification in levels can exploit any cointegration relationships between the variables, a specification in differences offers some robustness in the presence of structural breaks, see Carriero et al. (2015).

Cointegration, where two or more non-stationary (integrated) variables can form a stationary linear combination and thus are tied together in the long-run, is a powerful concept that is appealing from a forecasting standpoint. The information that the variables tend to move together in the long-run should contain valuable and exploitable information about their future states. VARs in log-differences ignore this information. The question is how important the information in the level variables is for forecasting macroeconomic variables. We address this question by providing a forecasting comparison of the BVAR in log-differences with BVEC models and BVARs in log-levels. BVEC models are rarely considered for forecasting macroeconomic variables. This is in particular the case for applications in higher dimensions. Checking for cointegration with increasing dimensions becomes burdensome and there is no automatic way to estimate BVEC models (see the discussion of the literature below). In order to fill this gap, we propose two hierarchical shrinkage priors for the long-run matrix of the BVEC model. For the first one, we propose a reduced rank prior which encourages shrinkage towards a low-rank, row-sparse and column-sparse long-run matrix. For the second one, we propose the use of the horseshoe prior which shrinks all elements of the long-run matrix towards zero. The horseshoe prior leaves the rank of the long-run matrix unrestricted. Thus, we also use the horseshoe prior with a reduced rank decomposition of the long-run matrix. The proposed approaches are highly automatic in the sense that they do not require much user input. Researchers who estimate a BVAR (e.g., for calculating impulse response functions) in first differences may be worried that possible cointegration between the level variables can bias their results. In order to address this concern, they can use one of our proposed prior distributions for the long-run matrix without any further efforts.

A simulation study reveals that BVEC models equipped with data-based priors perform well across a range of scenarios. In the presence of cointegration, BVEC models can improve estimation accuracy over BVARs in first differences. In the absence of cointegration, the data-based prior distributions are able to shrink the long-run matrix towards zero so that the estimation accuracy of the BVEC model is similar compared to the VAR estimated in first differences. Finally, BVEC models with our data-based prior turn out to be more flexible across different simulation setups than BVECs with fixed cointegration rank and BVARs in levels.

In an empirical investigation, based on several macroeconomic time series, we find that BVEC models, equipped with our proposed priors, forecast well both in terms of point and density forecasting accuracy. In particular for longer forecasting horizons, it turns out that exploiting information in the level variables has the potential for improving forecasts. Thus, working with VARs in first differences may ignore valuable information. Furthermore, the forecasting performance of BVARs in levels strongly depends on the model size and whether an expanding window or rolling window is used. In contrast, the BVEC models turn out to be more robust to such choices.

By proposing data-based priors for vector error correction models we contribute to the Bayesian cointegration literature (e.g., Villani (2000), Villani (2001), Strachan (2003), Strachan and Inder (2004), Villani (2005), Koop et al. (2009), Huber and Zörner (2019) and Hauzenberger et al. (2020)). These studies estimate the BVEC conditional on the number of cointegration relations. In practice, it may not always be straightforward to select the number of cointegration relations and to find sensible identifying restrictions for the cointegration vector or space. This problem becomes more severe in higher dimensions. Closely related to the cointegration literature is the prior proposed by Giannone et al. (2019). They use a prior based on economic theory for the long-run matrix. However, they find that it does not scale well to higher dimensions. In contrast, our proposed data-based priors are highly automatic. They do not rely on economic theory, selection of the cointegration rank, identifying restrictions and scale well to higher dimensions.

The remainder of this paper is organized as follows. Section 2 lays out and discusses the econometric framework. Section 3 provides a simulation study. Section 4 contains an overview of the data and presents the empirical findings. The last section concludes.

# 2. Econometric Framework

## 2.1. VAR

Let $\boldsymbol{y}_t = (y_{1,t}, \ldots, y_{n,t})$ be an $1 \times n$ vector of endogenous variables at time $t$. A standard VAR in levels can be written as:

$$\boldsymbol{y}_t = \boldsymbol{y}_{t-1}\boldsymbol{\phi}_1 + \cdots + \boldsymbol{y}_{t-p}\boldsymbol{\phi}_p + \boldsymbol{u}_t \tag{1}$$

where $\boldsymbol{u}_t \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$ and $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_p$ are $n \times n$ VAR coefficient matrices. Subtracting $\boldsymbol{y}_{t-1}$ on both sides of the equation and rearranging terms yields the error correction model

$$\Delta\boldsymbol{y}_t = \boldsymbol{y}_{t-1}\boldsymbol{\Pi} + \Delta\boldsymbol{y}_{t-1}\gamma_1 + \cdots + \Delta\boldsymbol{y}_{t-p+1}\gamma_{p-1} + \boldsymbol{u}_t \tag{2}$$

where $\Delta\boldsymbol{y}_t = \boldsymbol{y}_t - \boldsymbol{y}_{t-1}$, $\boldsymbol{\Pi} = (\boldsymbol{\phi}_1 + \cdots + \boldsymbol{\phi}_p) - \boldsymbol{I}_n$ and $\boldsymbol{\gamma}_l = -(\boldsymbol{\phi}_{l+1} + \cdots + \boldsymbol{\phi}_p)$, with $l = 1, \ldots, p-1$. Thus, using a VAR with variables in first differences implicitly sets the long-run matrix $\boldsymbol{\Pi} = \boldsymbol{0}$ and thereby ignores information in the level variables. We investigate whether information in the level variables is useful for forecasting macroeconomic variables. Furthermore, we propose two data-based priors for $\boldsymbol{\Pi}$. For the VAR coefficient matrices $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_{p-1}$ we employ the Minnesota prior which has led to the success of BVARs. Cross et al. (2020) compare the forecasting performance of the Minnesota prior with a range of other proposed priors in the literature and find that the Minnesota prior remains a solid choice. We provide a derivation of the Gibbs sampler for estimating the BVEC model in the online appendix.

## 2.2. Minnesota prior for the VAR coefficients

It is well known that VAR models are often overparametrized and that the most recent lags of a variable are expected to contain more information about the variable's current value than previous lags. We therefore use the Minnesota prior, which shrinks more distant lags more heavily towards zero, for the VAR coefficients

$$\gamma_l^{ij} \sim N(0, V_l^{ij}) \tag{3}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, n$, with

$$V_l^{ij} = \begin{cases} \frac{\kappa_1^2}{l^2}, & \text{if } i = j \\ \frac{\kappa_2^2}{l^2}, & \text{if } i \neq j \end{cases}. \tag{4}$$

The hyperparameters $\kappa_1$ and $\kappa_2$ control the informativeness of the prior. Giannone et al. (2015) estimate them by using a hierarchical prior for the hyperparameters. Gelman (2006) provides strong arguments for using the half-Cauchy distribution over an inverse gamma distribution for the scale parameters and Polson and Scott (2012) show that the half-Cauchy prior has excellent frequentist risk properties. Against this background, we employ the half-Cauchy prior for the hyperparameters

$$\kappa_1 \sim C^+(0, 1), \tag{5}$$

$$\kappa_2 \sim C^+(0, 1), \tag{6}$$

respectively. Note that for the VAR in levels we use the same type of prior but with mean one for the first own lag coeffcients.

## 2.3. Cointegration

If the $n$ time series in $\boldsymbol{y}_t$ are stationary, $\boldsymbol{\Pi}$ is a full rank matrix. If they all are non-stationary, integrated of order 1 or $I(1)$, but there is no cointegration, $\boldsymbol{\Pi}$ will be a zero matrix. In the latter case, no information is lost by using the VAR in first differences. In the error correction model, the focus is on the indeterminate case where $\boldsymbol{\Pi}$ is of reduced rank $0 < r < n$. It is possible to decompose $\boldsymbol{\Pi}$ into two $n \times r$ matrices $\boldsymbol{\Pi} = \boldsymbol{\beta}\boldsymbol{\alpha}'$ with $\boldsymbol{\beta}$ forming $r$ cointegrating relations, $\boldsymbol{y}_t\boldsymbol{\beta}$, or stationary linear combinations of the $I(1)$ variables in $\boldsymbol{y}_t$. In practice, it may not be straightforward to select $r$ and the uncertainty surrounding the cointegrating rank is seldom formally incorporated into the analysis.

Note that $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are not identified since any transformation with a full-rank matrix $\tilde{\boldsymbol{\alpha}}' = \boldsymbol{P}\boldsymbol{\alpha}', \tilde{\boldsymbol{\beta}} = \boldsymbol{\beta}\boldsymbol{P}^{-1}$ leaves $\boldsymbol{\Pi}$ unchanged. The traditional approach in the Bayesian cointegration literature (e.g., Villani (2001)) is to impose the normalization $\boldsymbol{\beta} = (\boldsymbol{I}_r, \boldsymbol{\beta}*)'$ conditioning on the cointegration rank, implying $r^2$ restrictions on $\beta$. Two recent examples using this normalization are Huber and Zörner (2019) and Hauzenberger et al. (2020). From a practical perspective, this identification scheme is sensitive with respect to permutations of the elements in $\boldsymbol{y}_t$, rendering the ordering of the variables an important modelling decision. In addition, as noted by, among others, Kleibergen and van Dijk (1994) and Geweke (1996), it is crucial that proper priors are used for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. An alternative approach is to place a prior on the cointegrating space, which is the only object the data is informative about, see Villani (2000). Such types of priors are studied in Strachan (2003), Strachan and Inder (2004), Villani (2005) and Koop et al. (2009).

## 2.4. Data-based Priors for the long-run Matrix

We propose two data-based priors for the vector error correction model. For the first approach, we work within a parameter-expanded framework, see Liu and Wu (1999), and assign shrinkage priors to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ to shrink the redundant columns and rows towards zero.

This means we set $r = n$ and leave $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ unidentified. The second approach involves placing a hierarchical prior directly on $\boldsymbol{\Pi}$. Both alternatives are highly automatic. They do not need economic theory, ordering of the variables, fine tuning of hyperparameters and selection of the cointegration rank.

#### 2.4.1. Reduced Rank Prior

Goh et al. (2017) show that low-rankness and row/column sparsity of $\boldsymbol{\Pi}$ can be represented as a certain row/column sparsity of $\boldsymbol{\alpha}$ and that $\boldsymbol{\beta}$ and such representations are invariant to any nonsingular transformation. Here we formulate a prior that encourages shrinkage towards a low-rank, row-sparse and column-sparse matrix $\boldsymbol{\Pi}$. We leave $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ unidentified and employ the following prior

$$\beta_{ij} \sim N(0, \lambda_{\beta,i}^2 \lambda_{\eta,j}^2) \tag{7}$$

$$\alpha_{ij} \sim N(0, \lambda_{\alpha,i}^2 \lambda_{\eta,j}^2) \tag{8}$$

for $i = 1, \dots, n$ and $j = 1, \dots, n$. The hyperparamter $\lambda_{\eta,j}$ shrinks the elements of the $j$-th column of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. If the columns of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are sparse, $\boldsymbol{\Pi}$ is of reduced rank. Thus, $\lambda_{\eta,j}$ effectively shrinks the cointegration rank. The hyperparamter $\lambda_{\beta,i}$ shrinks the $i$-th row of $\boldsymbol{\beta}$. Shrinking the $i$-th row of $\boldsymbol{\beta}$ towards zero implies that variable $y_i$ is not important for forming a stationary combination with any other set of variables. Finally, the hyperparamter $\lambda_{\alpha,i}$ shrinks the $i$-th row of $\boldsymbol{\alpha}$. If the $i$-th row of $\boldsymbol{\alpha}$ is sparse, $\Delta y_i$ does not adjust to any stationary combination of variables, i.e., it is weakly exogenous. Therefore, $\lambda_{\alpha,i}$ controls whether the information in the level variables helps explain variation in $\Delta y_i$. The selection of all three hyperparameters is important. In order to avoid fixing them at inappropriate values we estimate all three hyperparameters in a Bayesian fashion by placing half-Cauchy priors on them

$$\lambda_{\beta,i} \sim C^+(0, 1), \tag{9}$$

$$\lambda_{\alpha,i} \sim C^+(0, 1), \tag{10}$$

$$\lambda_{\eta,j} \sim C^+(0,1). \tag{11}$$

This hierarchical Bayes approach allows us to take uncertainty about the hyperparameters into account. Note that Chakraborty et al. (2020) use a related prior approach for rank reduction in Bayesian sparse multiple regression.

#### 2.4.2. Horseshoe prior

It is possible to leave the rank of $\mathbf{\Pi}$ unrestricted and place a prior directly on its elements. Here we follow Carvalho et al. (2010) and use the horseshoe prior, which is free of tuning parameters and has many appealing frequentist properties, see, e.g., Ghosh et al. (2016), Armagan et al. (2013) and van der Pas et al. (2014). The horseshoe prior takes the form

$$\Pi_{ij} \sim N(0, \tau^2 \psi_{ij}^2), \tag{12}$$

$$\tau \sim C^+(0,1), \tag{13}$$

$$\psi_{ij} \sim C^+(0,1). \tag{14}$$

The idea of this prior is that the global component $\tau^2$ shrinks all components towards zero and that the local component $\psi_{ij}^2$ prevents important coefficients from being shrunken to zero. The horseshoe prior requires absolutely no input from the researcher, while retaining its excellent shrinkage properties at the same time.

#### 2.4.3. Rank selection and reduced rank decomposition

The horseshoe prior leaves the rank of $\mathbf{\Pi}$ unrestricted. However, it is possible to combine the horseshoe prior with a reduced rank decomposition of $\mathbf{\Pi}$. Therefore, we need to select the rank of $\mathbf{\Pi}$. Based on the rank selection criterion (RSC) of Bunea et al. (2011), it is possible to determine the rank of $\mathbf{\Pi}$. For this purpose, it is useful to write the model in (2) in more compact form

$$\boldsymbol{Y}_\Delta = \boldsymbol{Y}_{-1}\boldsymbol{\beta}\boldsymbol{\alpha}' + \boldsymbol{X}\mathbf{\Gamma} + \boldsymbol{U}. \tag{15}$$

The dependent variables are stacked into a $T \times n$ matrix $\boldsymbol{Y}_\Delta$ so that its $t$-th row is $\Delta \boldsymbol{y}_t$, the $t$-th row of $\boldsymbol{Y}_{-1}$ is $\boldsymbol{y}_t$, the $t$-th row of $\boldsymbol{X}$ is $(\Delta \boldsymbol{y}_{t-1}, \ldots, \Delta \boldsymbol{y}_{t-p+1})$ and $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1', \ldots, \boldsymbol{\gamma}_{p-1}')'$.

The estimated rank $\hat{r}$ is given by the number of eigenvalues of the matrix $\tilde{\boldsymbol{Y}}_\Delta' \boldsymbol{P} \tilde{\boldsymbol{Y}}_\Delta$ which exceeds a threshold $\mu$:

$$\hat{r} = \max\{r : \lambda_r(\tilde{\boldsymbol{Y}}_\Delta' \boldsymbol{P} \tilde{\boldsymbol{Y}}_\Delta) \geq \mu\}, \tag{16}$$

with $\tilde{\boldsymbol{Y}}_\Delta = \boldsymbol{Y}_\Delta - \boldsymbol{X}\boldsymbol{\Gamma}$, $\boldsymbol{P} = \boldsymbol{Y}_{-1}(\boldsymbol{Y}_{-1}'\boldsymbol{Y}_{-1})^{-1}\boldsymbol{Y}_{-1}'$ being the projection matrix onto the column space of $\boldsymbol{Y}_{-1}$ and $\lambda_r$ denotes the $r$-th largest eigenvalue of $\tilde{\boldsymbol{Y}}_\Delta' \boldsymbol{P} \tilde{\boldsymbol{Y}}_\Delta$.

Following the recommendation of Bunea et al. (2011), the threshold $\mu$ is set equal to $\mu = 4S^2 n$ with

$$S^2 = \frac{||\tilde{\boldsymbol{Y}}_\Delta - \boldsymbol{P}\tilde{\boldsymbol{Y}}_\Delta||^2}{Tn - n^2}. \tag{17}$$

We can use a reduced rank decomposition of $\boldsymbol{\Pi}$ for any selection of the cointegration rank. Let $\boldsymbol{\Pi} = \boldsymbol{W}\boldsymbol{D}\boldsymbol{V}'$ be the singular value decomposition of $\boldsymbol{\Pi}$. Collect the $r$ largest singular values and corresponding vectors in the matrices $\boldsymbol{D}^* = \text{diag}(d_1, d_2, \ldots, d_r)$, $\boldsymbol{W}^* = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_r)$ and $\boldsymbol{V}^* = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_r)$. A rank $r < n$ restriction to $\boldsymbol{\Pi}$ is then given by $\boldsymbol{\Pi} = \boldsymbol{W}^*\boldsymbol{D}^*\boldsymbol{V}^{*\prime}$. Carriero et al. (2011) and Chakraborty et al. (2020) use the singular value decomposition for the posterior mean of the coefficient matrix for reduced rank estimation. In order to account for parameter uncertainty and uncertainty about the cointegration rank $r$, we determine $\hat{r}$, set $r = \hat{r}$ and employ the singular value decomposition for each posterior draw of $\boldsymbol{\Pi}$.

# 3. Simulation Study

In this section, we evaluate the frequentist properties of the proposed priors in a Monte Carlo study. We wish to establish that it can be beneficial, in terms of estimation accuracy, to use BVEC models in the presence of cointegration. But if no cointegration is present, the proposed priors provide sufficient shrinkage so that the BVEC has similar estimation accuracy compared to the VAR estimated in first differences. We employ the four DGPs used in Kleibergen and Paap (2002):

$$\Delta \boldsymbol{y}_t' = \begin{pmatrix} 0.1 \\ 0.1 \\ 0.1 \end{pmatrix} + \boldsymbol{\gamma}_1' \Delta \boldsymbol{y}_{t-1}' + \boldsymbol{\gamma}_2' \Delta \boldsymbol{y}_{t-2}' + \boldsymbol{\epsilon}_t,$$

$$\Delta \boldsymbol{y}_t' = \begin{pmatrix} 0.1 \\ 0.1 \\ 0.1 \end{pmatrix} + \begin{pmatrix} -0.2 \\ 0.2 \\ 0.2 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \end{pmatrix} \boldsymbol{y}_{t-1}' + \boldsymbol{\gamma}_1' \Delta \boldsymbol{y}_{t-1}' + \boldsymbol{\gamma}_2' \Delta \boldsymbol{y}_{t-2}' + \boldsymbol{\epsilon}_t,$$

$$\Delta \boldsymbol{y}_t' = \begin{pmatrix} 0.1 \\ 0.1 \\ 0.1 \end{pmatrix} + \begin{pmatrix} -0.2 & -0.2 \\ 0.2 & -0.2 \\ 0.2 & 0.2 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix} \boldsymbol{y}_{t-1}' + \boldsymbol{\gamma}_1' \Delta \boldsymbol{y}_{t-1}' + \boldsymbol{\gamma}_2' \Delta \boldsymbol{y}_{t-2}' + \boldsymbol{\epsilon}_t,$$

$$\Delta \boldsymbol{y}_t' = \begin{pmatrix} 0.1 \\ 0.1 \\ 0.1 \end{pmatrix} + \begin{pmatrix} -0.2 & -0.2 & -0.2 \\ 0.2 & -0.2 & -0.2 \\ 0.2 & 0.2 & -0.2 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \boldsymbol{y}_{t-1}' + \boldsymbol{\gamma}_1' \Delta \boldsymbol{y}_{t-1}' + \boldsymbol{\gamma}_2' \Delta \boldsymbol{y}_{t-2}' + \boldsymbol{\epsilon}_t,$$

where $\boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \boldsymbol{I})$. The four DGPs, denoted by DGP1, DGP2, DGP3 and DGP4, contain zero, one, two and three cointegrating relations, respectively. Setting $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_2 = \mathbf{0}$ and using the sample size $T = 100$ gives the exact simulation setups used in Kleibergen and Paap (2002). For our purposes we extend the simulation setups by setting $\gamma_1^{ij} = 0.2$ if $i = j$ and $\gamma_1^{ij} = 0.1$ if $i \neq j$ and similarly $\gamma_2^{ij} = -0.1$ if $i = j$ and $\gamma_2^{ij} = -0.01$ if $i \neq j$. We denote the four DGPs with $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_2 = \mathbf{0}$ by DGP1a, DGP2a, DGP3a and DGP4a and the four DGPs with $\boldsymbol{\gamma}_1 \neq \boldsymbol{\gamma}_2 \neq \mathbf{0}$ by DGP1b, DGP2b, DGP3b and DGP4b, respectively. Finally, we also consider a sample size of $T = 200$. Both considered sample sizes are typical in macroeconomic applications.

We compare the estimation accuracy of the BVAR both in levels and first differences with the BVEC model. The BVEC model is estimated with the reduced rank prior (BVECrr), with the horseshoe prior (BVEChh) and with the horsehoe prior combined with the reduced rank based on the SVD decomposition (BVEChhrs). Finally, we also estimate the BVEC with a fixed rank of $r = 1$, $r = 2$ and $r = 3$. In this case, we leave $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ unidentified and place fairly uninformative independent $N(0, 1)$ priors on their elements.[1] All models use the Minnesota prior described in equation (3) and use $p = 5$ (five lags for the BVAR in level and four lags for the BVAR in differences), as in our empirical investigation. In order to be able to compare the estimation accuracy, we transform the coefficient estimates of the BVAR in differences and the BVEC into the level coefficients $\boldsymbol{\phi}$ in equation (1). We do the same for the true coefficients to calculate the mean absolute error[2] (MAE) using $S = 500$ simulations

$$MAE = \frac{1}{S \times n^2 \times p} \sum_{s=1}^{S} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{l=1}^{p} |\phi_l^{ij} - \hat{\phi}_l^{ij}(s)|. \tag{18}$$

The MAE for the all DGPs is summarized in the top half of Table 1. The simulation study shows that in case of cointegration the BVAR in first differences generates the least precise estimates. In this case, estimation accuracy can be improved by using BVEC models. In case of no cointegration relations, the data-based priors are able to shrink the long-run matrix towards zero so that the BVEC models have similar estimation accuracy compared to the BVAR estimated in first differences. The BVAR in levels performs best when $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_2 = \mathbf{0}$ and no cointegration is present. In the other cases it provides the worst estimation accuracy among all models. The reason is that the Minnesota prior of the BVAR in levels shrinks the coefficients towards a random walk. In cases of cointegration and when $\boldsymbol{\gamma}_1 \neq \boldsymbol{\gamma}_2 \neq \mathbf{0}$, BVEC leads to improved estimation accuracy compared to the BVAR in levels. Moreover, the BVEC models improve much more than the BVAR in first differences or BVAR in levels in terms of relative MAE for $T = 100$ versus $T = 200$. Overall, the BVEC models equipped with data-based priors turn out to be safe options as they are doing well across all simulation setups. They increase estimation accuracy in case of cointegration, but if no cointegration is present the data-based priors provide a sufficient amount of shrinkage so that the BVEC model yields similar estimation accuracy

[1] We also considered $N(0, 5)$ priors, which led to very similar results.

[2] We also have calculated the mean squared error which leads to the same conclusions.

Table 1: Simulation Results

| Model | DGP1a | DGP2a | DGP3a | DGP4a | DGP1b | DGP2b | DGP3b | DGP4b |
|---|---|---|---|---|---|---|---|---|
| $T = 100$ | *MAE* | | | | | | | |
| BVARdiff | 0.0181 | 0.0442 | 0.0656 | 0.0705 | 0.0493 | 0.0606 | 0.0745 | 0.0777 |
| BVARlevel | 0.0062 | 0.0251 | 0.0322 | 0.0338 | 0.0662 | 0.0623 | 0.0599 | 0.0548 |
| BVEChh | 0.0192 | 0.0299 | 0.0255 | 0.0270 | 0.0502 | 0.0540 | 0.0512 | 0.0498 |
| BVEChhrs | 0.0193 | 0.0288 | 0.0257 | 0.0271 | 0.0501 | 0.0540 | 0.0521 | 0.0499 |
| BVECrr | 0.0197 | 0.0234 | 0.0253 | 0.0263 | 0.0503 | 0.0511 | 0.0515 | 0.0488 |
| BVECr=1 | 0.0211 | 0.0228 | 0.0339 | 0.0379 | 0.0506 | 0.0498 | 0.0554 | 0.0541 |
| BVECr=2 | 0.0228 | 0.0236 | 0.0240 | 0.0277 | 0.0515 | 0.0508 | 0.0507 | 0.0490 |
| BVECr=3 | 0.0239 | 0.0239 | 0.0242 | 0.0250 | 0.0521 | 0.0512 | 0.0512 | 0.0488 |
| $T = 200$ | *MAE* | | | | | | | |
| BVARdiff | 0.0124 | 0.0429 | 0.0635 | 0.0680 | 0.0401 | 0.0541 | 0.0692 | 0.0702 |
| BVARlevel | 0.0032 | 0.0207 | 0.0273 | 0.0277 | 0.0605 | 0.0511 | 0.0490 | 0.0448 |
| BVEChh | 0.0129 | 0.0165 | 0.0164 | 0.0176 | 0.0404 | 0.0414 | 0.0405 | 0.0393 |
| BVEChhrs | 0.0128 | 0.0165 | 0.0164 | 0.0177 | 0.0404 | 0.0414 | 0.0405 | 0.0393 |
| BVECrr | 0.0133 | 0.0151 | 0.0163 | 0.0169 | 0.0405 | 0.0406 | 0.0403 | 0.0386 |
| BVECr=1 | 0.0139 | 0.0147 | 0.0295 | 0.0337 | 0.0406 | 0.0401 | 0.0462 | 0.0462 |
| BVECr=2 | 0.0148 | 0.0152 | 0.0162 | 0.0206 | 0.0408 | 0.0406 | 0.0402 | 0.0396 |
| BVECr=3 | 0.0152 | 0.0164 | 0.0165 | 0.0167 | 0.0410 | 0.0410 | 0.0405 | 0.0386 |
| $T = 100$ | *MSFE* | | | | | | | |
| BVARlevel | 1.0280 | 0.9758 | 0.8851 | 0.8690 | 1.0308 | 1.0049 | 0.9159 | 0.8447 |
| BVEChh | 1.0210 | 0.9710 | 0.8759 | 0.8638 | 1.0076 | 0.9896 | 0.9111 | 0.8454 |
| BVEChhrs | 1.0287 | 0.9661 | 0.8803 | 0.8784 | 1.0235 | 0.9939 | 0.9180 | 0.8511 |
| BVECrr | 1.0240 | 0.9512 | 0.8665 | 0.8642 | 1.0064 | 0.9671 | 0.9056 | 0.8358 |
| BVECr=1 | 1.0226 | 0.9476 | 0.9039 | 0.9093 | 1.0078 | 0.9495 | 0.9517 | 0.8970 |
| BVECr=2 | 1.0499 | 0.9539 | 0.8640 | 0.8820 | 1.0250 | 0.9675 | 0.8965 | 0.8548 |
| BVECr=3 | 1.0673 | 0.9647 | 0.8669 | 0.8612 | 1.0346 | 0.9827 | 0.8999 | 0.8336 |
| $T = 200$ | *MSFE* | | | | | | | |
| BVARlevel | 1.0157 | 0.9526 | 0.8944 | 0.8532 | 1.0467 | 0.9906 | 0.9199 | 0.8508 |
| BVEChh | 1.0132 | 0.9398 | 0.8830 | 0.8497 | 1.0108 | 0.9778 | 0.9094 | 0.8507 |
| BVEChhrs | 1.0149 | 0.9455 | 0.8817 | 0.8496 | 1.0153 | 0.9944 | 0.9110 | 0.8502 |
| BVECrr | 1.0092 | 0.9283 | 0.8820 | 0.8477 | 1.0112 | 0.9653 | 0.9090 | 0.8491 |
| BVECr=1 | 1.0168 | 0.9210 | 0.9096 | 0.9184 | 1.0052 | 0.9579 | 0.9551 | 0.9206 |
| BVECr=2 | 1.0248 | 0.9330 | 0.8807 | 0.8648 | 1.0141 | 0.9683 | 0.9085 | 0.8812 |
| BVECr=3 | 1.0365 | 0.9364 | 0.8842 | 0.8457 | 1.0172 | 0.9724 | 0.9087 | 0.8455 |

The top half of the table shows the MAE as defined in equation (18). The bottom half of the table shows the MSFE as defined in equation (19) relative to the BVAR in first differences. We consider the BVAR in first differences (BVARdiff), the BVAR in levels (BVARlevel), the BVEC model with horseshoe prior (BVEChh), with horseshoe prior combined with rank selection (BVEChhrs), with reduced rank prior (BVECrr) and with fixed rank (BVECr=x).

compared to the BVAR estimated in first differences.

Comparing the BVEC models with data-based priors with the BVEC models with fixed rank reveals that it is beneficial in terms of estimation accuracy if the selected rank is equal to the true cointegration rank. However, there is a high risk, in terms of estimation accuracy, to select the wrong rank. This is in particular the case if the selected rank is smaller than the true rank. The BVEC models with data-based priors do better if the true rank is equal to zero but do not do much worse if the true rank is larger than zero. In particular, the reduced rank prior does a good job in shrinking the rank so that estimation accuracy is only slightly worse than the BVEC which fixes the rank equal to the true rank.

So far we have focused on estimation accuracy. Next we consider the out-of-sample forecasting performance measured by the mean squared forecasting error (MSFE) averaged over all variables and all $S = 500$ simulations

$$MSFE = \frac{1}{S \times n} \sum_{s=1}^{S} \sum_{i=1}^{n} (y_{i,T+1}^{(s)} - \hat{y}_{i,T+1}^{(s)})^2. \tag{19}$$

The results for the MSFE can be found in the bottom half of Table 1. We report the MSFE of each model by dividing it by the MSFE of the BVAR in first differences. Thus, values lower than one indicate better forecasting performance than the BVAR in first differences. Plausibly, we find, in case of no cointegration, that the BVAR in first differences provides the best forecasting performance. Nevertheless, the BVEC models with data-based priors are not much worse in this case. In the presence of cointegration the BVEC models and the BVAR in levels outperform the BVAR in first differences. These results are in line with the results for the estimation accuracy. Thus, we find that higher estimation accuracy translates into better out-of-sample forecasting performance. Comparing the BVEC models with data-based priors with the BVEC models with fixed cointegration rank reveals that fixing the cointegration at the true rank improves the forecasting performance. However, selecting the cointegration rank lower than the true rank is harmful for the forecasting performance. In contrast, the BVECs with data-based priors perform well for all DGPs and only perform slightly worse than the BVECs with fixed rank equal to the true rank.

Finally, we wish to emphasize the benefit of the shrinkage priors in a situation where additional variables are added to the system which may not be cointegrated with any other variable. In the online appendix we provide results for this case. In particular, we consider adding one variable to the DGPs which is not cointegrated with the other variables. In this setup, the BVEC models with data-based priors improve relative to the BVEC models with fixed rank. This illustrates the advantage of the data-based priors. In order to save space, we do not report the forecasting results for the BVEC models with fixed rank for macroeconomic time series, but instead note that they are in line with our simulation results. In particular, the data-based priors turn out to be important as a protection against overfitting which translates into poor out-of-sample forecasts.

# 4. Empirical Investigation

## 4.1. Data

We consider 15 macroeconomic time series, documented in Table 2, which are typical choices in the BVAR literature. All data are sourced from the Federal Reserve Bank of St. Louis economic database. The quarterly data cover the period from 1959-Q1 to 2019-Q4. We use a logarithmic transformation for all variables. We assume that all variables in log-differences are approximately stationary. This is a standard assumption in the BVAR literature and thus allows us to stay in line with the variable transformations (model specifications) used in the BVAR literature. For the estimation, all variables are standardized and for the forecasting evaluation we undo the standardization of the variables.

Table 2: Macroeconomic Time series

| Name | ID |
|---|---|
| Real Gross Domestic Product | GDPC1 |
| Real Personal Consumption Expenditures | PCECC96 |
| Real Gross Private Domestic Investment | GPDIC1 |
| Consumer Price Index for All Urban Consumers: All Items | CPIAUCSL |
| Civilian Employment | CE16OV |
| Gross Domestic Product: Chain-type Price Index | GDPCTPI |
| Gross Private Domestic Investment: Chain-type Price Index | GPDICTPI |
| Average Hourly Earnings of Production and Nonsupervisory Employees | CES2000000008 |
| Nonfarm Business Sector: Real Output Per Hour of All Persons | OPHNFB |
| Real M2 Money Stock | M2REAL |
| Private Residential Fixed Investment | PRFI |
| Industrial Production Index | INDPRO |
| Industrial Production: Final Products | IPFINAL |
| All Employees, Service-Providing | SRVPRD |
| Personal Consumption Expenditures: Chain-type Price Index | PCECTPI |

All times series have been downloaded from the FRED database.

## 4.2. Forecasting setup

We consider both an expanding window and a rolling window (using 100 periods) to evaluate the forecasts from 1984-Q2 to 2019-Q4. The rolling window approach allows us to investigate the importance of taking into account structural breaks in the BVEC models. In order to investigate how well the models and different priors scale to higher dimensions, we estimate BVARs with 4 variables, BVARs with 10 variables and BVARs with 15 variables. The four variable VAR includes real gross domestic product (GDP), real consumption, real investment and consumer prices (CPI), the 10 variable VAR includes the first 10 variables in Table 2 and the 15 variables VAR includes all variables in Table 2. Combining the expanding window approach and rolling window approach with the four, ten and fifteen variable VAR gives us six different scenarios which we use for our forecasting comparison. Using six different scenarios allows us to compare the robustness of the different BVARs. For real GDP, real consumption and CPI we compute point forecasts $\hat{y}_{i,t+h}$ as well as density forecasts $p(y_{i,t+h}|\boldsymbol{Y}_{1:t})$ for the horizons $h = 1, \ldots, 40$. As argued by Giannone et al. (2019), the accuracy of long-term forecasts is of direct importance in many VAR applications, including the estimation of impulse response functions, obtained as the difference between conditional and unconditional forecasts. Furthermore, the analysis of long-run forecasts appears to be a particularly useful device to detect spurious deterministic overfitting, a type of model misspecification that can affect all other aspects and applications of VARs, for detailed discussion, see Giannone et al. (2019). The forecasts are evaluated at $t = t_0, \ldots, T-h$. Let $y^o_{i,t+h}$ denote the actual value of the variable $y_{i,t+h}$. The mean squared forecasting error (MSFE) is calculated by

$$\text{MSFE} = \frac{\sum_{t=t_0}^{T-h} (y^o_{i,t+h} - \hat{y}_{i,t+h})^2}{T-h-t_0+1} \tag{20}$$

and the average log predictive likelihood (ALPL) is calculated as

$$\text{ALPL} = \frac{1}{T-h-t_0+1} \sum_{t=t_0}^{T-h} \log p(y_{i,t+h} = y^o_{i,t+h} | \boldsymbol{Y}_{1:t}). \tag{21}$$

In our forecasting evaluation we compare the BVAR in log-levels, the BVAR in log-first differences and the BVEC models with each other. We consider the BVEC model with the horseshoe prior, the BVEC model with the horseshoe prior combined with rank selection and the BVEC model with the reduced rank prior. Consistent with quarterly data, we use $p = 5$ (five lags for the BVAR in levels and four lags for the BVAR in differences). As a simple benchmark we use an AR(1) model.

## 4.3. Forecasting Results

We start with the discussion of the short-run forecasting performance followed by a discussion about the forecasting results for all 40 horizons. In particular, we first focus on the more conventional forecasting horizons of one quarter, one year and two years. These can be found in Figures A.1 to A.6. The first three figures show the MSFE, relative to the AR(1) benchmark model, for GDP, Consumption and CPI, respectively. The last three show the ALPL, relative to the AR(1) benchmark model, for GDP, Consumption and CPI, respectively. Furthermore, each figure contains the results for all six forecasting setups. In most cases the different BVAR models outperform the benchmark AR(1) model. This is in particular the case for one and two years ahead and for density forecasts. In addition, we find that the BVEC equipped with our different priors forecasts well. This finding tends to hold across the different model sizes. Furthermore, in several cases the forecasting performance improves with a higher model size. This shows the BVEC models with the data-based priors scale well to higher dimensions. Overall, they can deliver the best forecasting performance, but where not, they do not go too far wrong. This stresses the potential importance of using information in the level variables. However, the BVAR in levels tends to forecast worse at short horizons compared to the BVEC models.

The best forecasting performance of the BVARs depends on the predicted variable, forecasting horizons and whether a rolling or an expanding window is used. For the predicted variable GDP, no single best model forecasts emerges. Nevertheless, it is always a BVEC model which provides the best point or density forecasts for all horizons. For one quarter ahead, the BVEC model with the horseshoe prior, four variables and an expanding window provides the best point forecasts. For one year ahead, the BVEC model with the reduced rank prior, 15 variables and an expanding window approach provides the best point forecasts. Finally, for two years ahead, the BVEC model with the horseshoe prior, 10 variables and a rolling window provides the best point forecasts. In terms of density forecasts, we find that the BVEC model with the horseshoe prior, four variables and a rolling window for one quarter ahead does best. The BVEC model with the reduced rank prior, 15 variables and a expanding window for one year ahead provides the best density forecasts. Finally, the BVEC with the reduced rank prior, 15 variables and a rolling window provides the best density forecasts for two years ahead.

For consumption, we obtain a clearer picture. We find that the BVEC with the horseshoe prior and 15 variables using a rolling window provides the best point and density forecasts. Only for one quarter ahead, the BVEC model with four variables provides slightly better density forecasts. Finally, for CPI it turns out that the BVAR in log-differences tends to provide the best point as well as density forecasts. The forecasting performance of the BVEC models is in many cases only slightly worse or equal. For one quarter ahead, the BVEC models with the horseshoe prior and the BVAR in log-differences with 15 variables and a rolling window provide the best point forecasts. The best point forecasts for one year and two years are provided by the BVAR in log-differences using 10 variables and a rolling window. In terms of density forecasts, the BVAR in log-differences does best for one quarter and one year ahead. Finally, for two years ahead the BVAR in log-differences with 15 variables and an expanding window and the BVEC model with the horseshoe prior, 15 variables and a rolling approach provide the best density forecasts for CPI.

Figures A.7 to A.12 show the out-of-sample forecasting results for all 40 horizons. The first three figures depict the MSFE for GDP, Consumption and CPI, respectively. The last three figures show the ALPL for GDP, Consumption and CPI, respectively. Furthermore, each figure contains the results for all six forecasting setups. These figures reveal that the BVEC models tend to outperform the BVAR in differences as well as the AR(1) benchmark model for all variables. These findings hold for all six scenarios and for point as well as density forecasts at forecasting horizons higher than two years. Thus, in particular for forecasting at long horizons, the information from the level variables turn out to be valuable. The forecasting performance of the BVAR in levels strongly depends on the model size and whether an expanding window size or rolling window approach is used. Although they do best in some cases, they perform quite poorly in other cases. In contrast, the BVEC models turns out to be more robust and appears to be a safe choice.

# 5. Conclusion

We have proposed two different priors for the BVEC models. These priors are free of tuning parameters, do not require economic theory, and do not require the selection of the cointegration rank. We provide a simulation study which illustrates that it is beneficial, in terms of estimation accuracy, to use a BVEC model in the presence of cointegration. In the presence of no cointegration, the proposed priors sufficiently shrink the coefficients, so that the BVEC has similar estimation accuracy compared to the BVAR estimated in first differences. Furthermore, BVEC models with our data-based prior turn out to be more flexible across different simulation setups than BVEC models with fixed rank and BVARs in levels. A comparison of the point as well as the density forecasting performance of BVARs in log-differences with BVEC models and BVARs in log-levels reveals that it can be beneficial to use information from the level variables for out-of-sample forecasting. Furthermore, we find that BVEC models scale well to higher dimensions and deliver precise forecasts.

## References

Armagan, A. Dunson, D., Bajwa, W., Lee, J., and Strawn, N. (2013). Posterior consistency in linear models under shrinkage priors. *Biometrika*, 100(4):1011–1018.

Bańbura, M., Giannone, D., and Reichlin, L. (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics*, 25(1):71–92.

Bunea, F., She, Y., and Wegkamp, M. (2011). Optimal selection of reduced rank estimators of high-dimensional matrices. *The Annals of Statistics*, 39(2):1282–1309.

Carriero, A., Clark, T., and Marcellino, M. (2015). Bayesian VARs: Specification choices and forecast accuracy. *Journal of Applied Econometrics*, 30(1):46–73.

Carriero, A., Clark, T., and Marcellino, M. (2016). Common drifting volatility in large Bayesian VARs. *Journal of Business and Economic Statistics*, 34(3):375–390.

Carriero, A., Kapetanios, G., and Marcellino, M. (2011). Forecasting large datasets with Bayesian reduced rank multivariate models. *Journal of Applied Econometrics*, 26(5):735–761.

Carvalho, C., Polson, N., and Scott, J. (2010). The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480.

Chakraborty, A., Bhattacharya, A., and Mallick, B. (2020). Bayesian sparse multiple regression for simultaneous rank reduction and variable selection. *Biometrika*, 107(1):205–221.

Chan, J. (2020). Large Bayesian VARs: A flexible kronecker error covariance structure. *Journal of Business and Economic Statistics*, 38(1):68–79.

Cross, J., Hou, C., and Poon, A. (2020). Macroeconomic forecasting with large Bayesian VARs: Global-local priors and the illusion of sparsity. *International Journal of Forecasting*, 36(3):899–915.

Doan, T., Litterman, R., and Sims, C. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews*, 3(1):1–100.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3):515–534.

Geweke, J. (1996). Bayesian reduced rank regression in econometrics. *Journal of Econometrics*, 75(1):121–146.

Ghosh, P., Tang, X., Gosh, M., and Chakrabarti, A. (2016). Asymptotic properties of Bayes risk of a general class of shrinkage priors in multiple hypothesis testing under sparsity. *Bayesian Analysis*, 11(3):753–796.

Giannone, D., Lenza, M., and Primiceri, G. (2015). Prior selection for vector autoregressions. *Review of Economics and Statistics*, 97(2):436–451.

Giannone, D., Lenza, M., and Primiceri, G. (2019). Priors for the long run. *Journal of the American Statistical Association*, 114(526):565–580.

Goh, G., Dey, D., and Chen, K. (2017). Bayesian sparse reduced rank multivariate regression. *Journal of Multivariate Analysis*, 157:14–28.

Hauzenberger, N., Huber, F., Pfarrhofer, M., and Zörner, T. (2020). Stochastic model specification in Markov switching vector error correction models. *Studies in Nonlinear Dynamics and Econometrics*.

Huber, F. and Feldkircher, M. (2019). Adaptive shrinkage in Bayesian vector autoregressive models. *Journal of Business and Economic Statistics*, 37(1):27–39.

Huber, F. and Zörner, T. (2019). Threshold cointegration in international exchange rates: A Bayesian approach. *International Journal of Forecasting*, 35(2):458–473.

Kleibergen, F. and Paap, R. (2002). Priors, posteriors and bayes factors for a bayesian analysis of cointegration. *Journal of Econometrics*, 111(2):223–249.

Kleibergen, F. and van Dijk, H. (1994). On the shape of the likelihood/posterior in cointegration models. *Econometric Theory*, 10(3-4):514–551.

Koop, G. (2013). Forecasting with medium and large Bayesian VARs. *Journal of Applied Econometrics*, 28(2):177–203.

Koop, G., León-González, R., and Strachan, R. (2009). Efficient posterior simulation for cointegrated models with priors on the cointegration space. *Econometric Reviews*, 29(2):224–242.

Korobilis, D. and Pettenuzzo, D. (2019). Adaptive hierarchical priors for high-dimensional Vector Autoregressions. *Journal of Econometrics*, 212(1):241–271.

Litterman, R. (1986). Forecasting with bayesian vector autoregressions: five years of experience. *Journal of Business and Economic Statistics*, 4(1):25–38.

Liu, J. and Wu, Y. (1999). Parameter expansion for data augmentation. *Journal of the American Statistical Association*, 94(448):1264–1274.

Polson, N. and Scott, J. (2012). On the half-Cauchy prior for a global scale parameter. *Bayesian Analysis*, 7(4):887–902.

Sims, C. and Zha, T. (1998). Bayesian methods for dynamic multivariate models. *International Economic Review*, 39(4):949–968.

Strachan, R. (2003). Valid bayesian estimation of the cointegrating error correction model. *Journal of Business and Economic Statistics*, 21(1):185–195.

Strachan, R. and Inder, B. (2004). Bayesian analysis of the error correction model. *Journal of Econometrics*, 123(2):307–325.

van der Pas, S., Kleijn, B., and van der Vaart, A. (2014). The horseshoe estimator: Posterior concentration around nearly black vectors. *Electronic Journal of Statistics*, 8(2):2585–2618.

Villani, M. (2000). *Aspects of Bayesian Cointegration*. PhD thesis, Stockholm University.

Villani, M. (2001). Bayesian prediction with cointegrated vector autoregressions. *International Journal of Forecasting*, 17(4):585–605.

Villani, M. (2005). Bayesian reference analysis of cointegration. *Econometric Theory*, 21(2):326–357.

# Appendix A. Figures

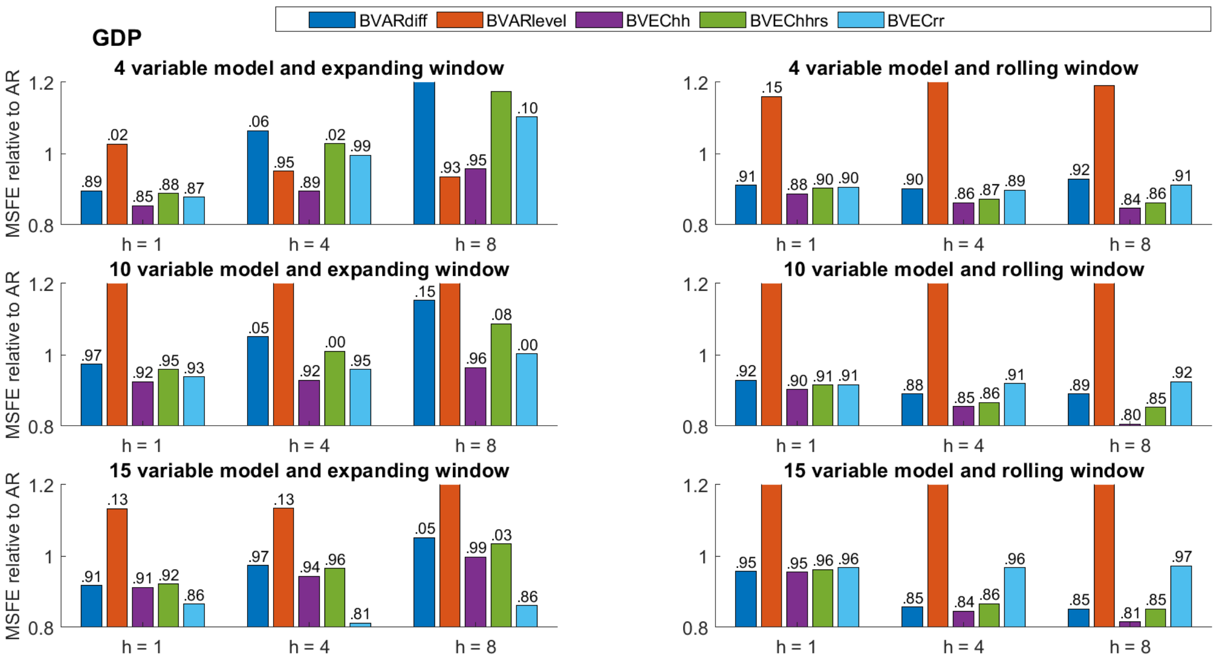


Figure A.1: Short-run MSFEs for GDP. The MSFE of each model is divided by the MSFE of an AR(1) benchmark model. We consider the BVAR estimated in first differences (BVARdiff), the BVAR estimated in levels (BVARlevel), the BVEC model with horseshoe prior (BVEChh), the BVEC model with horseshoe prior combined with rank selection (BVEChhrs) and the BVEC model with reduced rank prior (BVECrr).

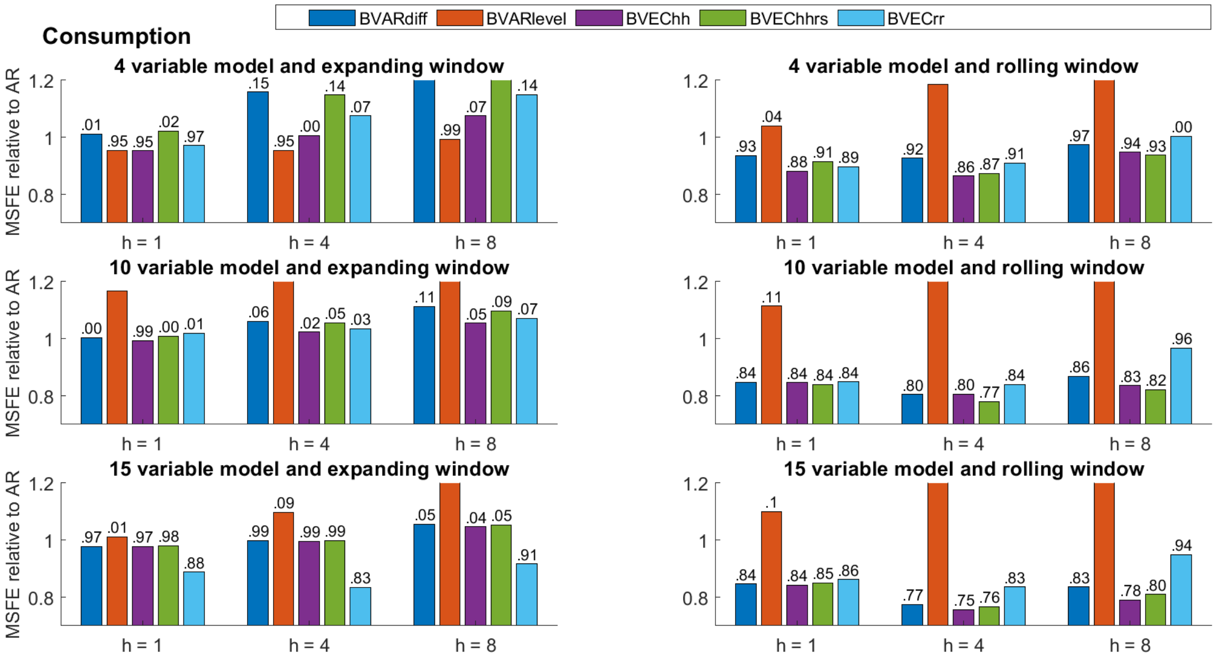


Figure A.2: Short-run MSFEs for Consumption. The MSFE of each model is divided by the MSFE of an AR(1) benchmark model. We consider the BVAR estimated in first differences (BVARdiff), the BVAR estimated in levels (BVARlevel), the BVEC model with horseshoe prior (BVEChh), the BVEC model with horseshoe prior combined with rank selection (BVEChhrs) and the BVEC model with reduced rank prior (BVECrr).

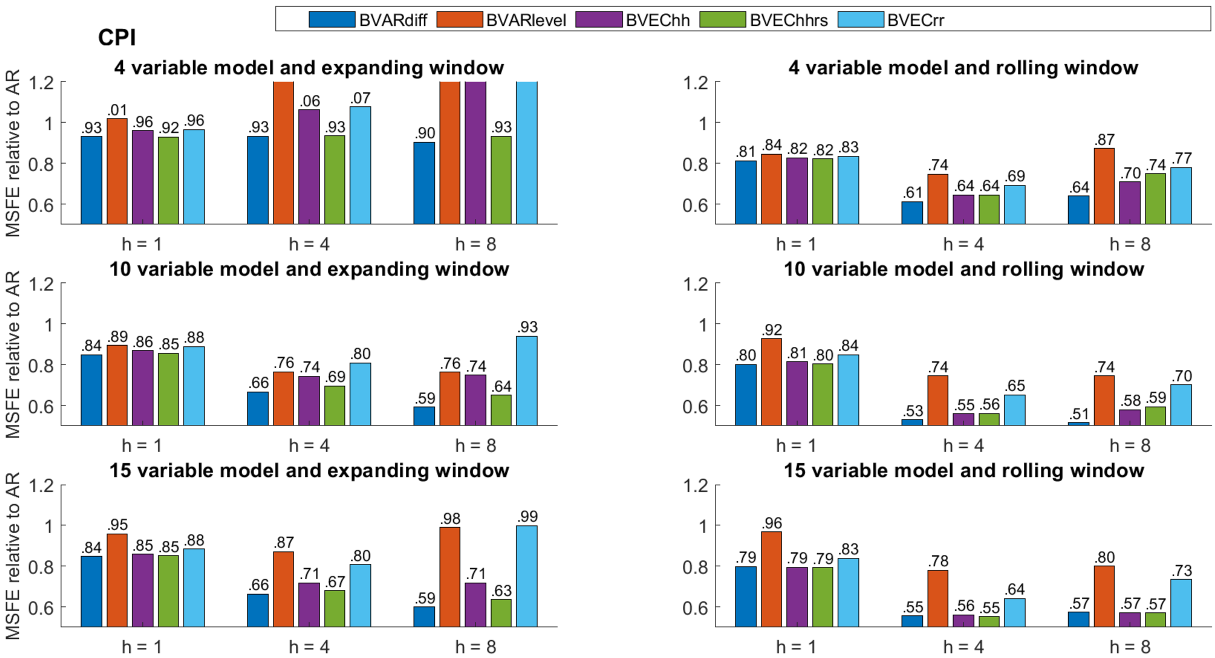


Figure A.3: Short-run MSFEs for CPI. The MSFE of each model is divided by the MSFE of an AR(1) benchmark model. We consider the BVAR estimated in first differences (BVARdiff), the BVAR estimated in levels (BVARlevel), the BVEC model with horseshoe prior (BVEChh), the BVEC model with horseshoe prior combined with rank selection (BVEChhrs) and the BVEC model with reduced rank prior (BVECrr).

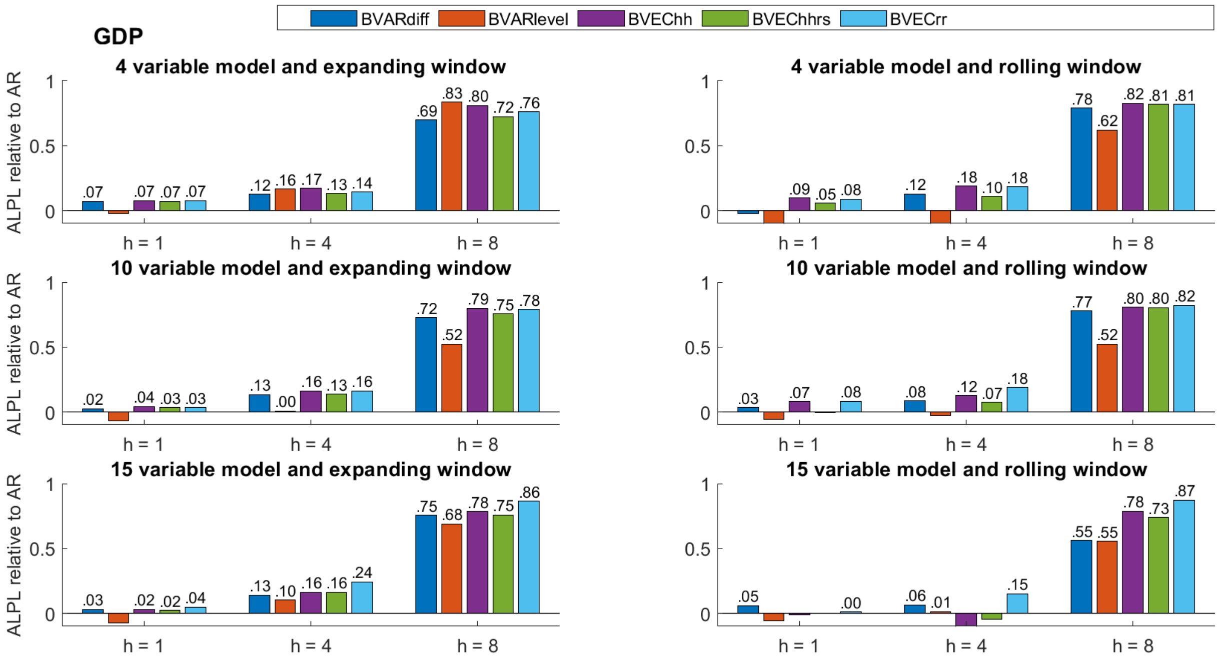


Figure A.4: Short-run ALPLs for GDP. The ALPL of each model is subtracted by the ALPL of an AR(1) benchmark model. We consider the BVAR estimated in first differences (BVARdiff), the BVAR estimated in levels (BVARlevel), the BVEC model with horseshoe prior (BVEChh), the BVEC model with horseshoe prior combined with rank selection (BVEChhrs) and the BVEC model with reduced rank prior (BVECrr).

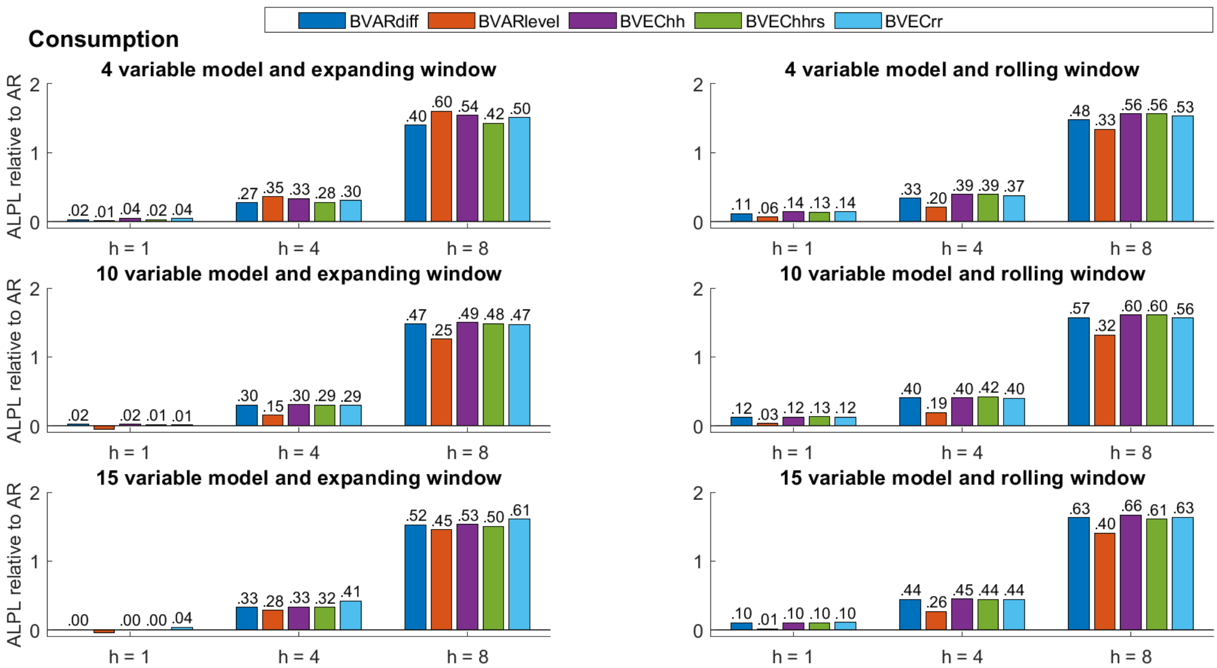


Figure A.5: Short-run ALPLs for Consumption. The ALPL of each model is subtracted by the ALPL of an AR(1) benchmark model. We consider the BVAR estimated in first differences (BVARdiff), the BVAR estimated in levels (BVARlevel), the BVEC model with horseshoe prior (BVEChh), the BVEC model with horseshoe prior combined with rank selection (BVEChhrs) and the BVEC model with reduced rank prior (BVECrr).

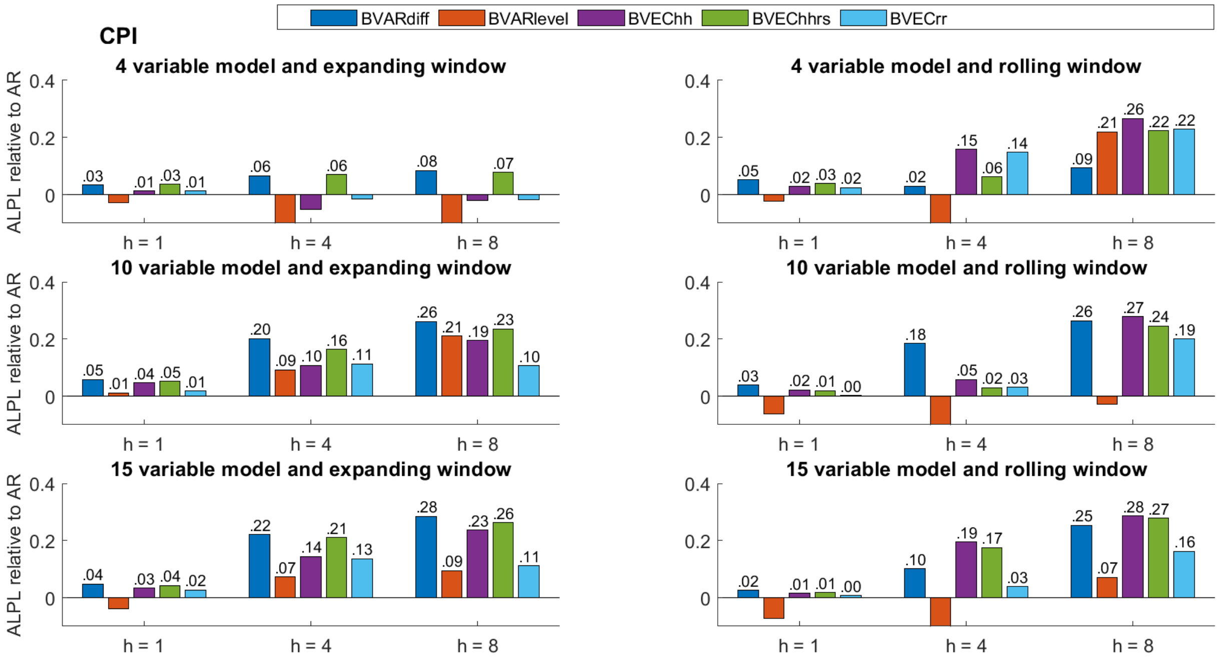


Figure A.6: Short-run ALPLs for CPI. The ALPL of each model is subtracted by the ALPL of an AR(1) benchmark model. We consider the BVAR estimated in first differences (BVARdiff), the BVAR estimated in levels (BVARlevel), the BVEC model with horseshoe prior (BVEChh), the BVEC model with horseshoe prior combined with rank selection (BVEChhrs) and the BVEC model with reduced rank prior (BVECrr).

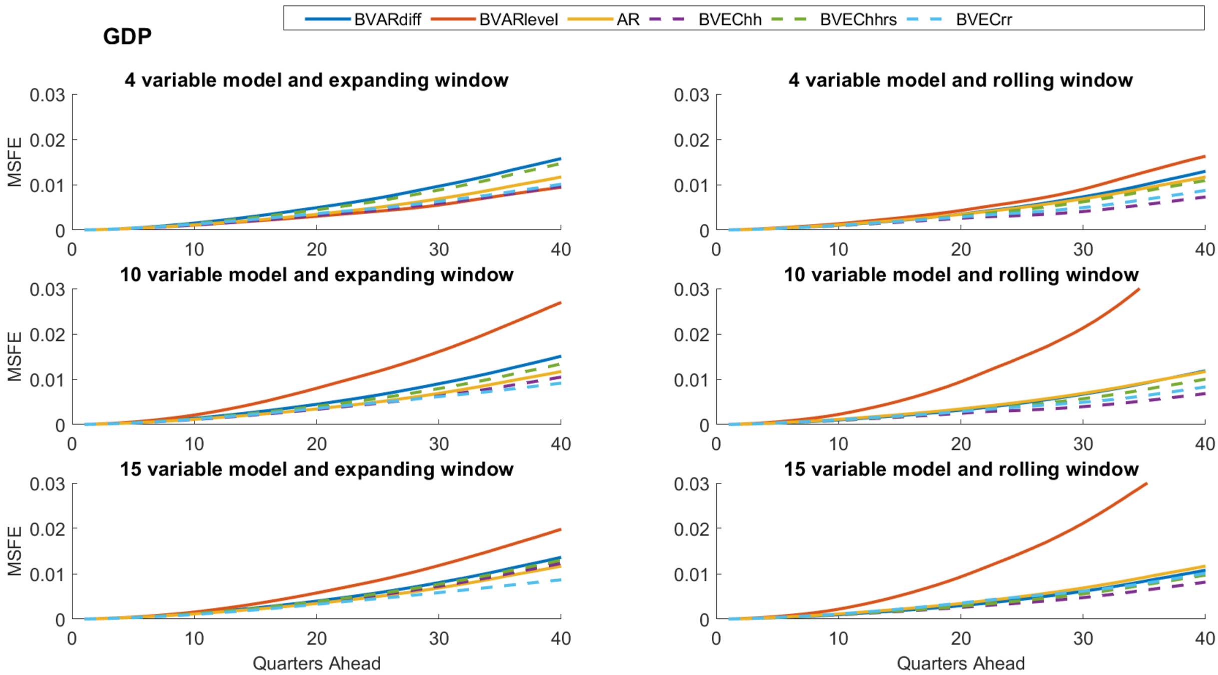


Figure A.7: Long-run MSFEs for GDP. We consider the BVAR estimated in first differences (BVARdiff), the BVAR estimated in levels (BVARlevel), the BVEC model with horseshoe prior (BVEChh), the BVEC model with horseshoe prior combined with rank selection (BVEChhrs), the BVEC model with reduced rank prior (BVECrr) and a AR(1) benchmark model.

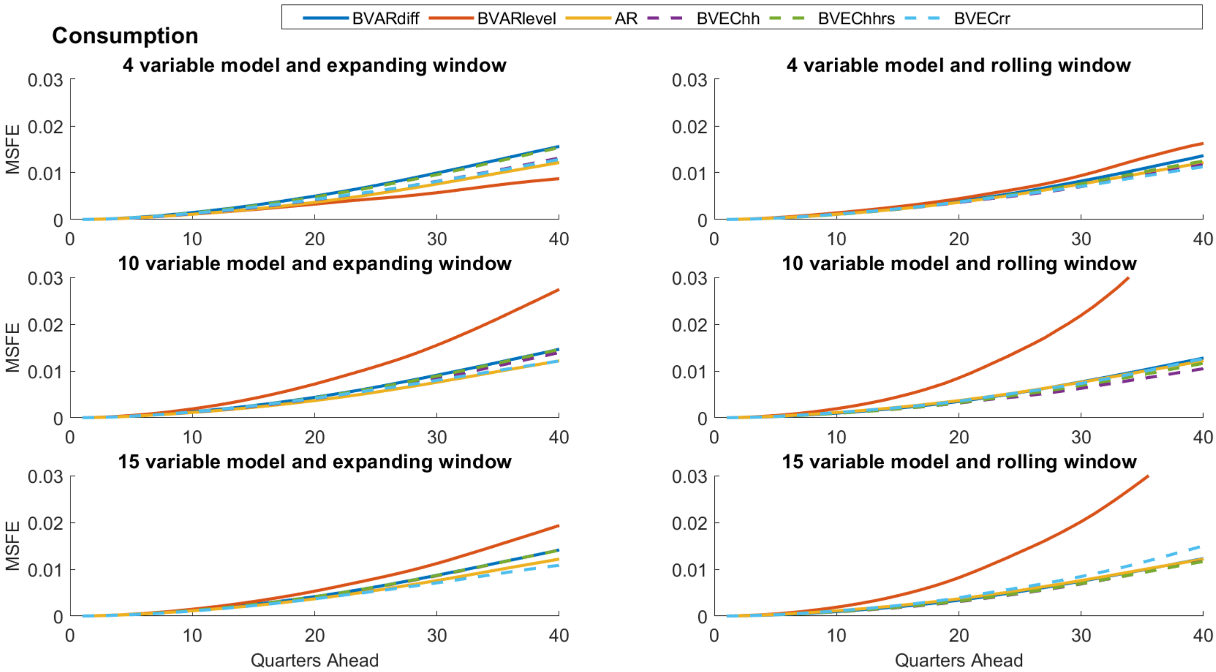


Figure A.8: Long-run MSFEs for Consumption. We consider the BVAR estimated in first differences (BVARdiff), the BVAR estimated in levels (BVARlevel), the BVEC model with horseshoe prior (BVEChh), the BVEC model with horseshoe prior combined with rank selection (BVEChhrs), the BVEC model with reduced rank prior (BVECrr) and a AR(1) benchmark model.

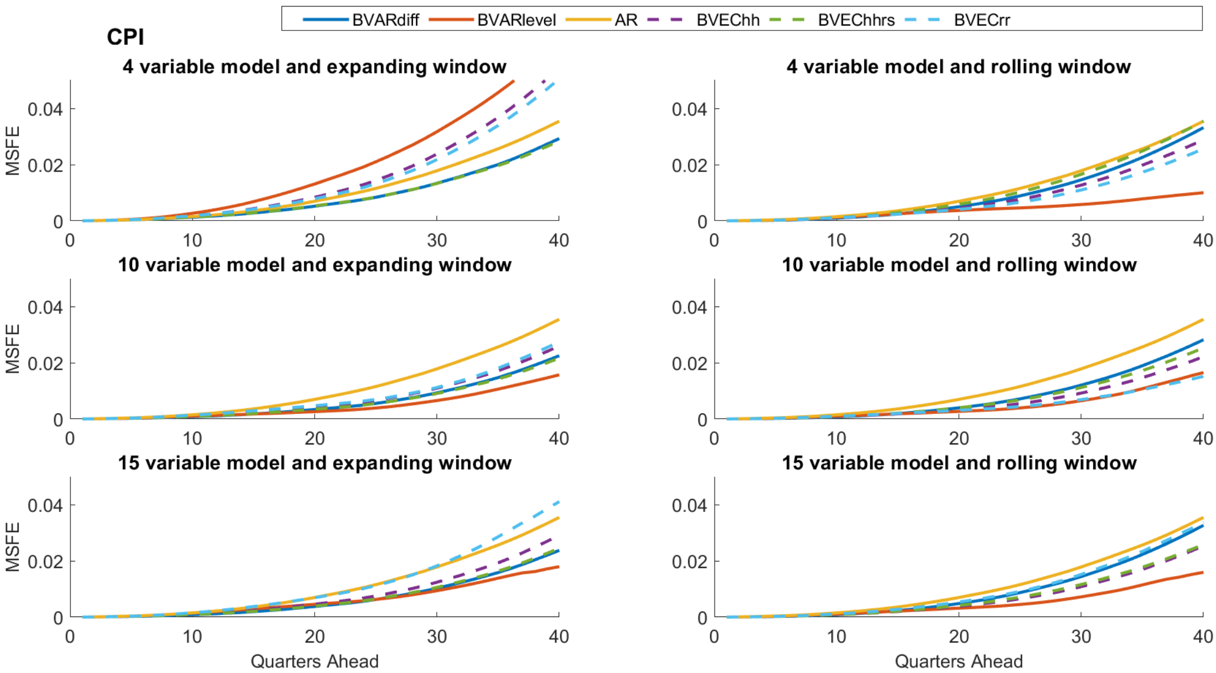


Figure A.9: Long-run MSFEs for CPI. We consider the BVAR estimated in first differences (BVARdiff), the BVAR estimated in levels (BVARlevel), the BVEC model with horseshoe prior (BVEChh), the BVEC model with horseshoe prior combined with rank selection (BVEChhrs), the BVEC model with reduced rank prior (BVECrr) and a AR(1) benchmark model.

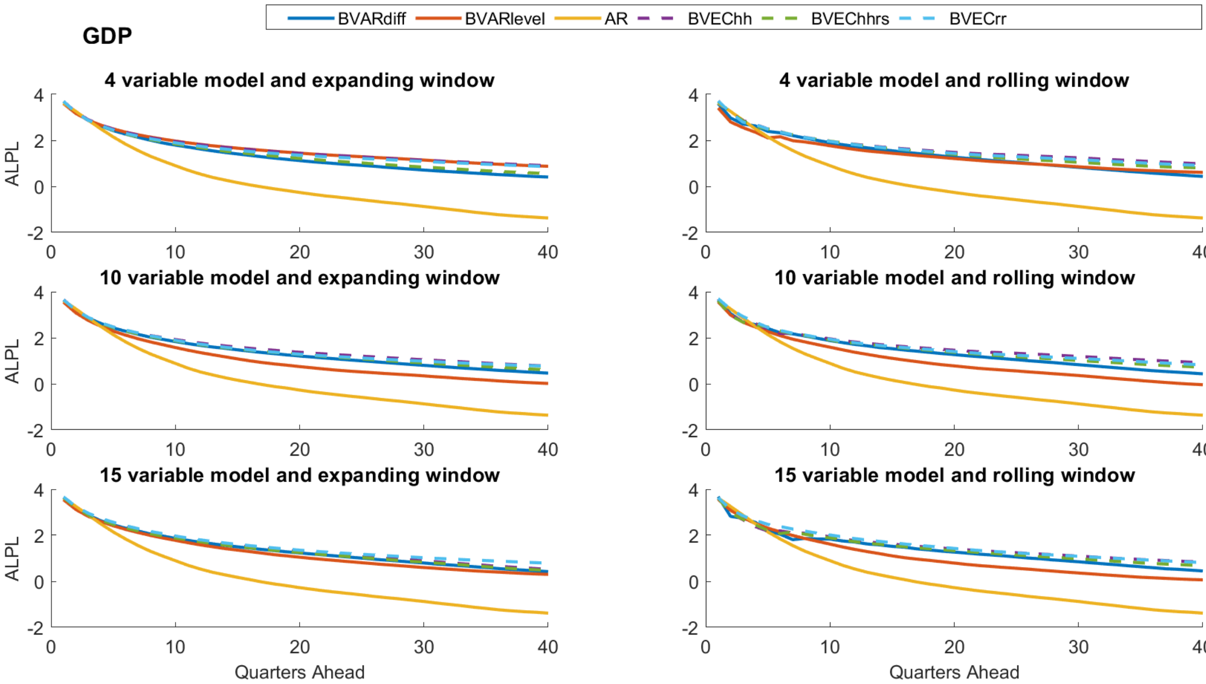


Figure A.10: Long-run ALPLs for GDP. We consider the BVAR estimated in first differences (BVARdiff), the BVAR estimated in levels (BVARlevel), the BVEC model with horseshoe prior (BVEChh), the BVEC model with horseshoe prior combined with rank selection (BVEChhrs), the BVEC model with reduced rank prior (BVECrr) and a AR(1) benchmark model.

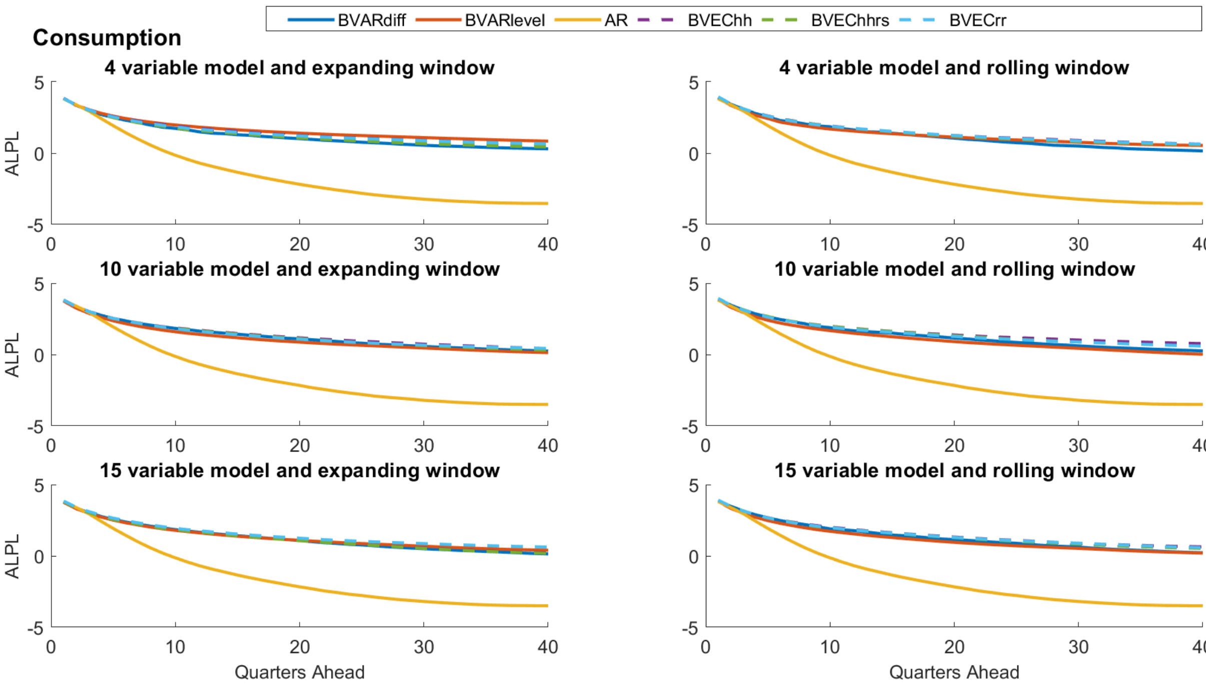


Figure A.11: Long-run ALPLs for Consumption. We consider the BVAR estimated in first differences (BVARdiff), the BVAR estimated in levels (BVARlevel), the BVEC model with horseshoe prior (BVEChh), the BVEC model with horseshoe prior combined with rank selection (BVEChhrs), the BVEC model with reduced rank prior (BVECrr) and a AR(1) benchmark model.

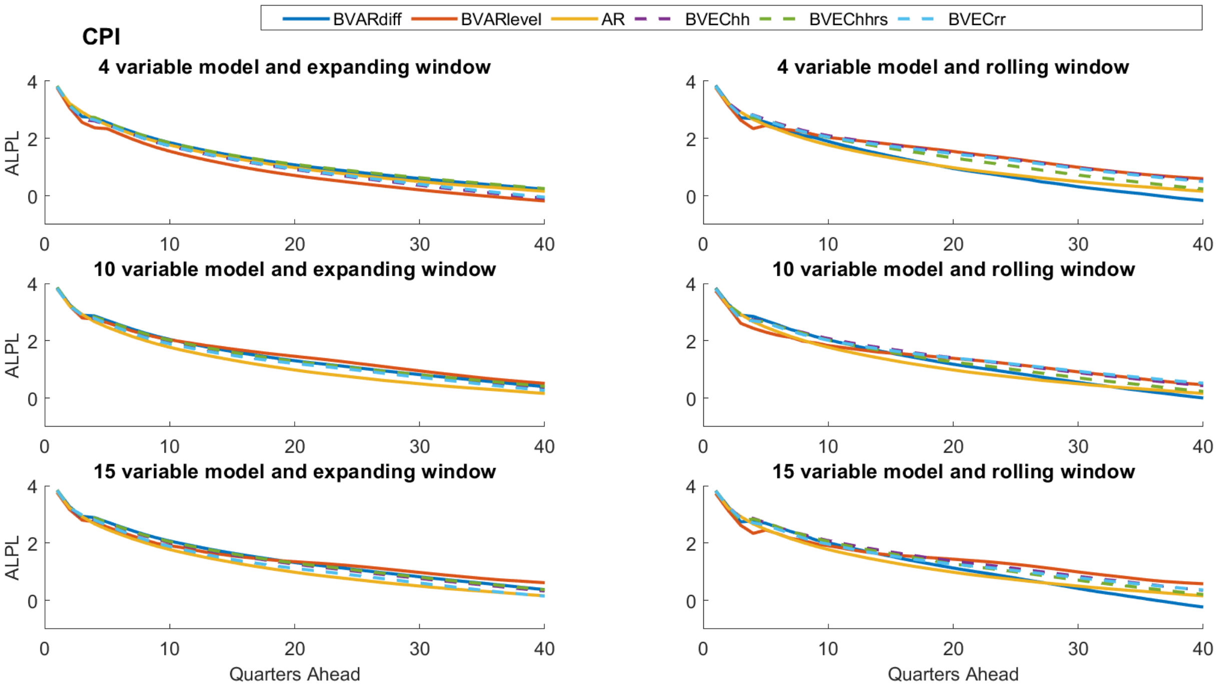


Figure A.12: Long-run ALPLs for CPI. We consider the BVAR estimated in first differences (BVARdiff), the BVAR estimated in levels (BVARlevel), the BVEC model with horseshoe prior (BVEChh), the BVEC model with horseshoe prior combined with rank selection (BVEChhrs), the BVEC model with reduced rank prior (BVECrr) and a AR(1) benchmark model.