International Journal of Forecasting 39 (2023) 1185–1204

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Penalized estimation of panel vector autoregressive models: A panel LASSO approach

Annika Camehl *

Erasmus University Rotterdam, The Netherlands

ARTICLE INFO

Keywords:
Forecasting
Model selection
Multi-country model
Shrinkage estimation
Sparse estimation

ABSTRACT

This paper proposes LASSO estimation specific for panel vector autoregressive (PVAR) models. The penalty term allows for shrinkage for different lags, for shrinkage towards homogeneous coefficients across panel units, for penalization of lags of variables belonging to another cross-sectional unit, and for varying penalization across equations. The penalty parameters therefore build on time series and cross-sectional properties that are commonly found in PVAR models. Simulation results point towards advantages of using the proposed LASSO for PVAR models over ordinary least squares in terms of forecast accuracy. An empirical forecasting application including 20 countries supports these findings.

1. Introduction

Panel vector autoregressive (PVAR) models, such as those surveyed by Canova and Ciccarelli (2013) and Breitung (2015), allow to augment unit-specific models with lagged variables of another unit, to model covariances between the error terms of different units, and to specify unit-specific coefficient matrices.[1] PVAR models are an excellent forecasting tool. For instance, for macroeconomic applications, they can take into account additional data from the cross-sectional dimension. However, the models suffer from heavy parametrization. The high dimensionality leads to increased estimation uncertainty and, thus, decreased forecasting accuracy. To tackle this estimation challenge, Bayesian factor approaches for PVAR models compress information in common, unit- and variable-specific factors (Billio et al., 2016; Canova & Ciccarelli, 2004, 2009; Ciccarelli et al., 2016; Koop & Korobilis, 2018), Bayesian panel selection priors implement higher shrinkage across panel units (Billio et al., 2019; Koop & Korobilis, 2016; Korobilis, 2016), and restricted ordinary least squares approaches assume no dependence or homogeneity across the panel units (see Breitung, 2015; Canova & Ciccarelli, 2013).

This paper proposes a penalized estimation strategy of the LASSO type that is adapted to the nature of PVAR models. The estimation uses four penalization constraints. The first penalizes the autoregressive parameters of lags, with the penalization depending on the lags. The second penalizes parameters depending on the equation. The third penalizes the parameters in the equation of the other units in the model. The fourth penalizes heterogeneous parameters of same variables across units.

The advantages of the penalized estimation specific to PVAR models are manifold: the penalties make use of the information coming from the structure of the panel data without imposing hard restrictions, the interpretability of the results yields direct insights in which variables contribute to forecasting individual variables, the procedure allows for flexible lag selection per equation, and the estimation accounts for the correlation among idiosyncratic risks.

* Correspondence to: Department of Econometrics, Erasmus School of Economics, 3000 DR Rotterdam, The Netherlands.
E-mail address: camehl@ese.eur.nl.

[1] In the following, the term PVAR model refers to the unrestricted model which allows for dependencies and heterogeneity across units (Canova & Ciccarelli, 2004, 2013; Koop & Korobilis, 2016). In the unrestricted form the PVAR model can be seen as a large VAR model.

https://doi.org/10.1016/j.ijforecast.2022.05.007

The specified penalty parameters, regulating the amount of shrinkage and selection, build on autoregressive (AR), vector autoregressive (VAR) and PVAR characteristics. That is, the penalization constraints capture the fact that more recent lags provide more important information with respect to the dynamics than more distant ones. They can also model that lags of variables of the same unit are more important than lags of another unit. Furthermore, the penalties allow for penalization differing between equations but can also make use of homogeneity across units. A higher penalization on increasing lags is in line with the specification of the Litterman prior for VAR models (Litterman, 1986). As demonstrated by Song and Bickel (2011), Nicholson et al. (2016, 2017), Wilms et al. (2018), Messner and Pinson (2019) and Nicholson et al. (2020), including grouping structures or time series properties in the specification of the LASSO for estimating VAR models, e.g. by modeling hierarchical lag structures or encouraging similarities within or across groups, can improve forecasting accuracy compared to the LASSO penalty for VAR models introduced by Hsu et al. (2008) which is fixed for the whole system. Likewise, contributions on Bayesian selection priors for PVAR models support that accounting for the inherent panel dimension within the data can enhance forecasting performance (Koop & Korobilis, 2016; Korobilis, 2016).

The LASSO proposed in this paper allows to estimate PVAR models by regularization towards the commonly used restrictions for PVAR models, no dynamic interdependencies and homogeneity between units (see e.g., Canova & Ciccarelli, 2013). While the proposed penalty terms build on these specific restrictions assumed for PVAR models, is not necessary to restrict interdependencies across and within cross-sectional units a priori. This is especially useful for macroeconomic applications since theoretical arguments for setting restrictions on interdependencies across countries are often missing or hard to justify.

The introduced LASSO provides direct insights concerning the main contributing variables for forecasting. The selection property of the LASSO for PVAR models improves the interpretability of the model compared to Bayesian shrinkage methods and factor approaches which either produce dense output or aggregate information into factors. I show using monthly log changes in consumer prices indices and industrial production growth of five economies (Germany, France, Italy, United Kingdom, and United States) that the sparsity pattern is stable over time and can thus give insight into the most informative predictors. While the LASSO starts with the assumption of a sparse model, the researcher is flexible in choosing the maximal sparsity and density (by setting the grid values on which the penalty parameters are searched). Selecting the optimal penalty parameters via cross-validation determines the amount of sparsity in a data-driven way. With the aim of specifying the dynamic dependency structures, the researcher trades a dense solution against interpretatility when using the here proposed LASSO techniques specific to PVAR models. Alternatively, the suggested penalty parameter specification can be combined with estimators which would favor dense solutions, such as the ridge regressor, if that is more suitable for the specific data. Giannone et al. (2021) allow in their Bayesian setting for a combination of sparse and dense models finding large evidence that dense solutions are favored in several micro-, macroeconomics and financial applications. Compared to the results of Giannone et al. (2021) who find that for a cross-country model (their macro 2 example includes GDP growth rates for 90 countries) a dense model is favored, my findings provide some evidence that the dense relations stem from the static correlations while dynamic interdependencies across countries are sparse.

Due to the dimension reduction, the proposed LASSO improves point forecasts as it reduces the variance around the estimates by imposing extra shrinkage on linkages among lags and variables from other cross-sections which are expected to contribute less to the forecasting accuracy. In general, compared to Bayesian methods and factor models, penalized estimators improve forecasting performance for macroeconomic time series models as shown, among others, by Smeekes and Wijler (2018) or Wilms et al. (2018).

The introduced penalty parameters allow for a flexible lag length selection per equation. The maximal length of lags included for each variable of each cross-section can differ for each individual series. This is relevant for forecasting applications including variables which have different characteristics in terms of persistency in particular.

The penalized estimation uses a weighted sum of squared residuals as the loss function. This is an important aspect for PVAR models because the procedure allows for correlations of error terms among all cross-sectional units. Using the sum of squared residuals as the loss function, as it is done in the standard LASSO, restricts the covariance matrix to the identity matrix. Hence, the procedure imposes strict assumptions on the dependence structure between the cross-sectional units. Lee and Liu (2012), Basu and Michailidis (2015), and Davis et al. (2016) modify the loss functions in the LASSO optimization for VAR models and allow for unrestricted covariances, but they assume a fixed penalty term for the whole system. While Ngueyep and Serban (2015) propose a penalized log-likelihood scheme applying penalties for higher lags and within group or between group penalties, they still restrict the covariance matrix in their approach to a block structure by assuming no dependence across groups. I estimate the here proposed penalized methods applied to PVAR models with a coordinate descent algorithm which keeps the computational complexity at a tolerable level.[2]

Alternative models for studying cross-sectional interdependencies are large Bayesian VAR models (BVAR), global VAR models (GVAR), or factor augmented VAR models (FAVAR) (for a comparison see Canova & Ciccarelli, 2013; Feldkircher et al., 2020; Kilian & Lütkepohl, 2017). How international information enters these models differs. Large BVAR models use shrinkage priors to ensure

[2] As an example, with optimal penalty parameters estimating a PVAR model with 12800 autoregressive parameters takes around five minutes on a standard computer (Intel(R) Core(TM) i7-8750H CPU @ 2.20 GHz with 16.0 GB RAM).

estimation feasibility but these priors do not account for the panel structure in the data (Banbura et al., 2010; Carriero et al., 2015, 2009; Koop, 2013). Similar to my proposed LASSO for PVAR models these priors (e.g. the Litterman prior) impose stronger penalization on increasing lags. However, they do not differentiate between domestic and foreign lags. Thus, large BVAR models including variables of several cross-sectional units can be seen as unrestricted PVAR models. GVAR models introduce international dependencies by including weighted foreign terms in country-specific models (Crespo Cuaresma et al., 2016; Dees et al., 2007; Pesaran et al., 2009, 2004). In that way, they introduce a (weighted) homogeneity restriction on foreign autoregressive terms. Hence, a GVAR model can be specified as a restricted PVAR model. My penalty specification allows for heterogeneous foreign autoregressive terms but for a homogeneity penalty on domestic lags. Due to the aggregation GVAR models cannot provide insights on dependencies across variables dis-aggregated over all cross-sectional units as straight forward as the PVAR models estimated with the proposed LASSO. FAVAR models assume that international co-movement is driven by a small number of common factors (Bernanke et al., 2005).

The results of a simulation and an empirical application show the competitiveness of the penalized estimation with penalty parameters specific to the nature of PVAR models as a frequentist alternative to Bayesian PVAR models. The simulation results in lower mean squared forecast errors, higher average logarithmic scores, and mean squared errors of the lasso techniques compared to least squares alternatives. When forecasting up to four macroeconomic variables for up to 20 countries, the proposed LASSO improves the forecasting accuracy in terms of mean squared forecast error relative to OLS, Bayesian estimation and to single country models.

Accounting for the panel dimension in the penalty terms increases the forecasting performance. A PVAR model estimated with a penalization term specific to the nature of panel data outperforms models estimated with a LASSO only including equation-specific penalties. As the latter model can be seen as an unrestricted large VAR model, this finding complements the empirical literature showing that in multi-country set-ups PVAR models improve forecasting performance compared to BVAR models (e.g., Feldkircher et al., 2020; Koop & Korobilis, 2019; Korobilis, 2016). In line with Alessi and Banbura (2009), Crespo Cuaresma et al. (2016), and Feldkircher et al. (2020) who report that GVAR models perform in general weaker than BVAR and PVAR models, my findings demonstrate that forecasting accuracy does not considerably benefit from homogeneity restrictions implemented in the penalty specification or as in GVAR models.

Differences across methods are less pronounced for density forecasts compared to point forecasts. The proposed LASSO including lag- and equation-specific penalties as well as a penalty on foreign lags performs in particular well during classified expansion periods. The relative point and density forecasting performance among models is stable with respect to the economic environment. Graphs on sparsity patterns and international connectedness visualize the richness of spillover analyses feasible with PVAR models estimated with LASSO techniques.

2. Penalized estimation for PVAR models

2.1. PVAR model

Panel vector autoregressive models include several units, such as countries, and unit-specific variables in one model. PVAR models account for interdependencies and heterogeneities across units by jointly modelling multiple variables of several units. An unrestricted PVAR model with N units and G variables per unit for $t = 1, \dots, T$ periods is given by

$$y_{it} = A_{i1}Y_{t-1} + A_{i2}Y_{t-2} + \cdots + A_{iP}Y_{t-P} + u_{it}, \tag{1}$$

where y_{it} denotes a vector of dimension $G \times 1$ for unit i with $i = 1, \dots, N$. The $Y_{t-p} = (y'_{1t-p}, \dots, y'_{Nt-p})'$ is a $NG \times 1$ vector and the coefficient matrix A_{ip} is of dimension $G \times NG$ for lags $p = 1, \dots, P$. The u_{it} have mean zero and covariance matrix Σ_{ii}. The covariance matrices across units are given by $E(u_{it}u'_{jt}) = \Sigma_{ij}$ for $i \neq j$. A penalized estimation typically requires standardization of the data and, due to the demeaning, Eq. (1) does not contain an intercept.

In compact form, the PVAR model can be written as

$$Y = BX + U, \tag{2}$$

where $Y = (Y_1, \dots, Y_T)$ with $Y_t = (y'_{1t}, \dots, y'_{Nt})'$ and the coefficient matrix $B = (B_1, \dots, B_P)$ with $B_p = (A_{1p}, \dots, A_{Np})'$ is of dimension $NG \times NGP$. The matrix $X = (X_1, \dots, X_T)$ with $X_{t-1} = (Y_{t-1}, \dots, Y_{t-P})'$ includes all lagged variables and is of dimension $NGP \times T$. The U has mean zero and covariance matrix Σ of dimension $NG \times NG$.

The unrestricted PVAR model allows for dynamic and static interdependencies as well as for heterogeneities across units. The X_{t-1} includes lagged values of every variable in each equation. The unrestricted B-matrix and the covariance matrix Σ enable unit-specific coefficients and correlations between error terms of all possible variable-unit combinations. The PVAR model has $(NG)^2P$ unknown parameters of the B-matrix and $NG(NG+1)/2$ parameters of Σ. Variables are ordered per unit meaning that the first G rows of the system model variables of unit one, while the rows $NG - G + 1$ to NG describe the variables of unit N. The large number of parameters can lead to the curse of dimensionality problem. The penalized estimation provides a solution to deal with this issue. Introducing a shrinkage penalty in the regression enables coping with situations in which $T < NGP$, can improve prediction accuracy, and produce interpretable models as discussed by Tibshirani (1996) and Hastie et al. (2015).

In the unrestricted form, as given in Eq. (2), the PVAR model can be seen as a large VAR model.[3] The difference to a VAR model including multiple time series from one unit lies in the additional information the researcher has by knowing the panel structure of the data. Based on this data structure the researcher can already form the

[3] I follow related papers such as Canova and Ciccarelli (2004), Canova and Ciccarelli (2013) and Koop and Korobilis (2016) who also refer to the unrestricted model as a PVAR model.

hypothesis that certain restrictions might apply. Commonly assumed restrictions for panel data are no static or dynamic interdependcies between cross-sectional units or homogeneity of autoregressive coefficients (see, e.g., Canova & Ciccarelli, 2013). This additional available information is the starting point for the proposed estimation procedure.

2.2. Penalty term and loss function for PVAR models

The penalized estimation for PVAR models allows for shrinkage towards homogeneous coefficients across units, for lag-specific and equation-specific shrinkage, and for an unrestricted covariance matrix. The optimization problem of the penalized estimation for PVAR models is therefore given by:

$$\underset{B}{\operatorname{argmin}} \ \frac{1}{T} tr\left[(Y - BX)'\Omega(Y - BX)\right] + \Phi(\Lambda, B, \bar{B}), \tag{3}$$

where tr denotes the trace of the matrix and Ω is the precision matrix, $\Omega = \Sigma^{-1}$. The function, $\Phi(\Lambda, B, \bar{B})$, denotes the penalty term. It depends on penalty parameters, collected in Λ, on B, the matrix of autoregressive coefficients, and on $\bar{B}$, a matrix of zeros and averaged coefficients.

The model specified so far allows for a variety of penalty functions for different penalized estimators such as LASSO, ridge regression or elastic net. The optimization problem for the LASSO is given in (4) and is solved for each element, b^{ij}_{klp}, of B:

$$\begin{aligned}\underset{b^{ij}_{klp}}{\operatorname{argmin}} \ & \frac{1}{T}\sum_{t}^{T}\sum_{i,j}^{N}\sum_{k,l}^{G}\omega^{ij}_{kl}\left(Y^{i}_{k,t} - \sum_{p}^{P} b^{ij}_{klp}X^{j}_{lp,t}\right)^2 \\ & + \gamma\sum_{i=j}^{N}\sum_{k,l}^{G}\left|b^{ij}_{kl1} - \bar{b}_{kl}\right| \\ & + \sum_{p=2}^{P}\sum_{i=j}^{N}\sum_{k,l}^{G}\lambda_k p^{\alpha}\left|b^{ij}_{klp}\right| \\ & + c\sum_{p=1}^{P}\sum_{i\neq j}^{N}\sum_{k,l}^{G}\lambda_k p^{\alpha}\left|b^{ij}_{klp}\right|,\end{aligned} \tag{4}$$

where b^{ij}_{klp} is the element in B referring to the pth lag of variable l of unit j in the equation of variable k of unit i for $p = 1, \ldots, P, i, j = 1, \ldots, N$ and $k, l = 1, \ldots, G$. The ω^{ij}_{kl} is an element of the inverse of the covariance matrix, $\Sigma^{-1} = \Omega$, corresponding to the lth variable of unit j and the kth variable of unit i. The elements $\bar{b}_{kl}$, for $k, l = 1, \ldots, G$, are variable-specific coefficients which are homogeneous over all cross-sectional units. The coefficient $\bar{b}_{kl}$ denotes the impact of the lth variable on variable k. The $Y^{i}_{k,t}$ refers to the kth variable of unit i and $X^{j}_{lp,t}$ is the element in X referring to the pth lag of variable l of unit j in t.

To allow for specific time series and cross section penalties, the penalty parameters consist of the following terms

$$\begin{array}{lll} \gamma & \text{if} \ \ i = j, \ p = 1 & i, j = 1, \ldots, N \\ \lambda_k p^{\alpha} & \text{if} \ \ i = j, \ p = 2, \ldots, P & i, j = 1, \ldots, N \ \text{and} \ k = 1, \ldots, NG \\ \lambda_k p^{\alpha} c & \text{if} \ \ i \neq j, \ p = 1, \ldots, P & i, j = 1, \ldots, N \ \text{and} \ k = 1, \ldots, NG. \end{array}$$

The penalties can be decomposed into three layers: an autoregressive penalty, a vector autoregressive penalty and a PVAR penalty. The autoregressive or time series penalty, p^{α}, captures that more recent lags provide more information than more distant ones. The penalty increases with the lag order, $p = 1, \ldots, P$, for $\alpha > 0$. This set-up closely follows the idea of the widely applied Litterman prior used for VAR models (Doan et al., 1984; Litterman, 1986). Compared to a penalty which does not vary across lags, meaning here $\alpha = 0$, it favors rather smooth structures avoiding that lags inbetween are set to zero.

The VAR penalty, λ_k, varies across equations, providing different shrinkage for the multiple time series included in the system. This flexibility across equations can consider the different nature of variables. By providing an equation-wise penalty, shrinkage can vary across equations of financial or real variables and can account for different overall persistent levels and dependence on other variables and lags.

The PVAR penalties, γ and c, induce shrinkage towards homogeneous coefficients and penalize lags of variables belonging to another cross-sectional unit. The penalty γ shrinks the first lag of a cross section towards a variable-specific average of all units. The penalty function penalizes the difference between an autoregressive parameter of the first lags and the group average of this parameter, $b^{ij}_{kl1} - \bar{b}_{kl}$. The group average is a homogeneous parameter across all units. With this formulation, the specified penalty function resembles ideas of the fused LASSO (Tibshirani et al., 2005) where the difference between successive lags is penalized but adopting it to the panel framework.

The cross section penalty, $c > 1$ for variables of a different unit, models that lags of variables of the same unit have a larger impact than lags of variables of another unit. The parameters γ, α and c do not vary across equations. The PVAR penalties shrink towards the commonly assumed restrictions on PVAR models, namely cross-sectional homogeneity and no dynamic interdependencies (see for an overview, Canova & Ciccarelli, 2013).

Compared to the LASSO as in Tibshirani (1996) or Hsu et al. (2008) the loss function of the optimization problem is the weighted sum of squared residuals. The weights are given by the inverse of the covariance matrix Ω. This is in line with Lee and Liu (2012) who point out that in a VAR model correlations between error terms have an impact on the estimated parameters in a restricted regression.

2.3. Penalty parameters selection and estimation of the covariance and homogenous effects

In order to solve (4) we first need to select the optimal penalty parameters, estimate the covariance matrix to obtain an estimate for Ω and estimate the homogeneous effects in $\bar{B}$.

I select the optimal penalty parameters via a rolling cross-validation technique following Song and Bickel (2011) and Nicholson et al. (2016, 2017). The penalty parameters are chosen such that they minimize one-step ahead mean squared forecast errors. This procedure

accounts for time dependence. Alternatively, Stock and Watson (2012) and Bergmeir et al. (2018) use leave-m-out cross-validation also for time series data. Like Song and Bickel (2011), I split the sample in three periods: The first period from 1 to $T_1 - 1$ is used for estimating the model, based on the second period from T_1 to $T_2 - 1$ different penalty parameters are evaluated, and the third period from T_2 to the end of the sample is later used for forecast evaluation of the lasso. The model is estimated in a rolling scheme taking the observations from t to $T_1 + t - 1$ for $t = 1, \ldots, T_2 - T_1$. For each t the out-of-sample forecast accuracy for a specific penalty parameter λ_k is measured by the one-step ahead mean squared forecast error (MSFE) for variable k, $k = 1, \ldots, NG$.

The search for the optimal λ_k is done over a grid of penalty parameter values whereby at the maximal value all coefficients equal zero while γ, a and c are fixed. This is repeated for different γ, a and c combinations. The choice of the length (number of grid points) and density of the grid faces the trade-off between estimation time and performance improvements. In practice, a grid with a wide range of values but with larger distance between the values can be chosen first to select the rough value range. After that a grid with a smaller range and distance between values can be used. The optimal combination of all penalty parameters is chosen such that the average MSFEs are minimized.

The cross section penalty, c, separates variables from the same unit and those from a different unit. However, all variables from a different unit are treated in a similar way as c is fixed for the whole model. That is done to simplify the selection of the penalty parameter. A c varying across different units complicates the determination of the optimal penalty parameters and increases the computational time. For some empirical application a more flexible c might be appropriate. Such a flexibility can be build in by grouping units and having different c parameters for sub-groups of units. The same argumentation holds for γ. Following the idea of global VAR models, the c parameter can also be modelled depending on exogenous connectivity measures such as trade weights for countries.

The covariance matrix is estimated using a graphical LASSO (GLASSO) following Friedman et al. (2008). To obtain an estimator for the covariance matrix the Gaussian penalized log-likelihood

$$\log\det(\Omega) - tr(S\Omega) - \rho\,\|\Omega\| \tag{5}$$

is maximized with respect to the non-negative definite inverse of the covariance matrix $\Omega = \Sigma^{-1}$. The matrix S is the empirical covariance, $tr(S\Omega)$ is the trace of $S\Omega$ and $\|\Omega\|$ is the sum of the absolute values of each element of Ω. No initial estimation is needed since the estimator uses the empirical covariance of the data Y as input. For $\rho > 0$ the regression is penalized, while for $\rho = 0$ the classical maximum likelihood estimator is obtained. The details of the GLASSO are in the Supplementary Appendix. As pointed out by Banerjee et al. (2008), $\hat{\Sigma}$ is invertible even in the case when the number of variables is larger than the number of observations due to the regularization using $\rho > 0$. This estimator of the covariance matrix is plugged into Eq. (4).

An alternative way to estimate the covariance matrix is to use a joint likelihood approach (Basu & Michailidis, 2015; Davis et al., 2016; Lee & Liu, 2012; Ngueyep & Serban, 2015) or the least squares estimator $\hat{\Sigma} = \frac{1}{T-kk}(Y - \hat{B}X)(Y - \hat{B}X)'$, where kk is the number of degrees of freedom (Tibshirani, 1996). In contrast to the GLASSO estimation, this approach can lead to problems for the invertability of the covariance matrix in large systems.

The homogeneous coefficients are calculated with an aggregated estimator. The aggregated estimator uses data averaged over the cross section. With the averaged data a VAR model is estimated with ordinary least squares. The homogeneous coefficients are updated during the estimation. That means, for each LASSO estimate on a specific set of observations the homogeneous coefficients are calculated on the same set of observations.

The optimization problem of the lasso for PVAR models is solved using a coordinate descent algorithm as proposed in Friedman et al. (2007, 2010). This iterative algorithm updates for iteration *iter* from B_{iter} to B_{iter+1} by a univariate minimization over a single b^{ij}_{klp}. It iterates over all elements in B till convergence is reached. The coordinate descent algorithm can be used since the non differentiable part of the optimization problem is separable. Convexity and separability of the problem ensure the existence of a global solution. The optimization algorithm and the derivation of the lasso estimator are described in detail in Appendices A and B.

Additionally, the maximum lag length needs to be set. The maximum lag length can be chosen either via information criteria or according to the frequency of the data. As the regularized estimator will set irrelevant lags to zero, a rather high lag length can be selected. However, the maximum lag order has an impact on the computational time needed.

The forecast performance of the estimators is evaluated for the period T_2 to T by iterative point forecasts based on rolling window forecasts with fixed penalty parameters determined for the period one to $T_2 - 1$. As point forecasts do not capture the model uncertainty, I additionally report density forecasts which summarize the estimated forecast distribution capturing uncertainty around the estimated autoregressive parameters, the error distribution, or the lag length selection. While in the Bayesian approach density forecasts are straight forward to generate, obtaining the predictive density for LASSO estimators is not trivial and no common way exist in the literature. I construct the predictive densities based on bootstrapping in-sample residuals (following Garcia et al., 2017). That way no distributional assumption on the predictive density has to be made a-priori. For each time period in each rolling window a randomly drawn observation from the in-sample residuals is added to each h-step ahead point forecast. By repeating that 10,000 times a empirical predictive density for each model is obtained.

2.4. Model specification and benchmark models

I consider four variants of the estimator: a LASSO with equation-specific shrinkage by λ_k, lag-specific shrinkage by α, higher shrinkage for lags of another unit by c and

weighted sum of squared residuals, a LASSO with the same penalty parameters where the covariance matrix is set to the identity matrix, a LASSO with equation-specific shrinkage by λ_k, lag-specific shrinkage by α, higher shrinkage for lags of another unit by c, shrinkage towards homogeneous coefficients by γ, and with weighted sum of squared residuals, and a LASSO with equation-specific shrinkage by λ_k where the covariance matrix is set to the identity matrix.

As a general benchmark model, the PVAR model is estimated with ordinary least squares. While this can serve as a benchmark for small models, OLS is infeasible for larger models for which $T < NGp$. As a widely used Bayesian alternative, I implement the cross-sectional shrinkage approach proposed by Canova and Ciccarelli (2004, 2009). It groups coefficients due to factorizing but it does not use possible sparsity for dimension reduction (see Korobilis, 2016). The Supplementary Appendix 3.1 gives implementation details on the Bayesian method. In addition, a block-diagonal system ordering the variables in unit blocks is estimated with OLS, thus, assuming no dynamic interdependencies between units. Such an a priori assumption can be hard to justify theoretically especially for macroeconomic applications. Furthermore, I compare the performance to separate VAR models for each unit estimated with OLS and with a LASSO with equation-specific shrinkage.

3. Simulation studies

3.1. Simulation set-ups

This section evaluates the finite sample performance of the penalized estimators for PVAR models. I consider five Monte Carlo simulations, each with sample sizes $T = 100$ and $T = 500$ for number of units $N = 5$ and number of variables per unit $G = 2$. All DGPs are generated from stationary PVAR models. The idiosyncratic errors, U_t, for all data generating processes (DGP) are assumed to be normally distributed with mean zero and covariance matrix Σ with 0.02 on the main diagonal and 0.01 on the off-diagonal.

DGP1. The observations Y_t are generated from a PVAR(4) model where each B_p, for $p = 1, \ldots, 4$, has a block-diagonal structure. To ensure stationarity the block-diagonal elements are low-triangular. For having some heterogeneity in the coefficients each non-zero coefficient is equal to $(-0.6)^{(p-1)}$ plus a number randomly drawn in the interval $[-0.1, 0.1]$. The coefficients are getting smaller with more distant lags. This scenario is motivated by a sparse model where lags of an unit have no effect on variables of a different unit.

DGP2. The observations Y_t are generated from a PVAR(4) model where each B_p, for $p = 1, \ldots, 4$, has a triangular structure. The non-zero coefficients in the lower triangular part of B_1 are 0.6 for variables of the same unit and 0.4 for variables from another unit plus a number randomly drawn in the interval $[-0.1, 0.1]$. For $p = 2, 3, 4$ the non-zero coefficients equal $(-0.6)^{(p-1)}$ times the corresponding element in B_1. This structure models a sparse system with heterogeneous coefficients and smaller coefficients for lags belonging to another unit and for distant lags.

DGP3. Similar to DGP2 the data are generated from a PVAR(4) model where each B_p, for $p = 1, \ldots, 4$, has a triangular structure. In contrast to DGP3 all non-zero coefficients in B_1 are 0.6 plus a number randomly drawn in the interval $[-0.1, 0.1]$. For $p = 2, 3, 4$ the non-zero coefficients equal $(-0.6)^{(p-1)}$ times the corresponding element in B_1. This case presents a sparse model with heterogeneous coefficients and smaller coefficients for distant lags. There is no distinction between lags of the same unit and of a different unit.

DGP4. The data are generated from a PVAR(4) model as specified in DGP2 but no randomly drawn number is added to the coefficients. This case presents a sparse model with homogeneous coefficients and smaller coefficients for lags belonging to another unit and for distant lags.

DGP5. The data are generated from a PVAR(1) model with element j, m of B_1 given by $B(j, m) = (-1)^{|j-m|} 0.4^{|j-m|+1}$. This case represents a dense DGP. None of the coefficients are equal to zero but several are close to zero. The sizes of the coefficients decrease exponentially from the main diagonal. Thus, coefficients belonging to another unit are smaller than those of the same unit.

The coefficient matrices of the five DGPs are shown in Fig. 1. The lag length of estimated PVAR models is set to the true lag length. The sample is split into initialization, penalty parameter selection and forecast evaluation according to $T_2 = T - 0.1T$ and $T_1 = T_2 - 0.15T$. For the three simulation set-ups the penalty parameters are determined via cross-validation. The grid for λ_k consists of five values between 0.01 and $(1/NGp)\lambda_k^{max} = \max(\max(XY'))$. The grid for c is [1.2, 1.4, 1.6], for α [0.2, 0.4, 0.6], for γ [0.05, 0.2], and for ρ [0, 0.1, 0.2, 0.3, 0.4].

3.2. Performance criteria

The performance of the estimators is evaluated along mean squared forecast errors (MSFE), average logarithmic scores (ALS), the correct sparsity patterns, the fraction of relevant variables included, the number of variables included and mean squared errors (MSE) (similar to e.g. Kock & Callot, 2015; Ren & Zhang, 2010; Tibshirani, 1996). The h-step ahead MSFE for variable k of unit i for one Monte Carlo replication is calculated as

$$MSFE_k^i = \frac{1}{T - h - T_2 - 1} \sum_{t=T_2}^{T-h_{max}} (\hat{Y}_{k,t+h}^i - Y_{k,t+h}^i)^2$$

where $\hat{Y}_{k,t+h}^i$ denotes the iteratively estimated h-step ahead forecast for variable k of unit i for t with $t = T_2, \ldots, T-1$ and $h = 1, \ldots, h_{max}$, $h_{max} = 12$. To evaluate density forecasts, the $h-$step ALS for variable k of unit i are defined as (Amisano & Giacomini, 2007)

$$ALS_k^i = \frac{1}{T - h - T_2 - 1} \sum_{t=T_2}^{T-h_{max}} \log \hat{f}_t(Y_{k,t+h}^i) \quad (6)$$

where $\hat{f}_t$ denotes the calculated predictive density.

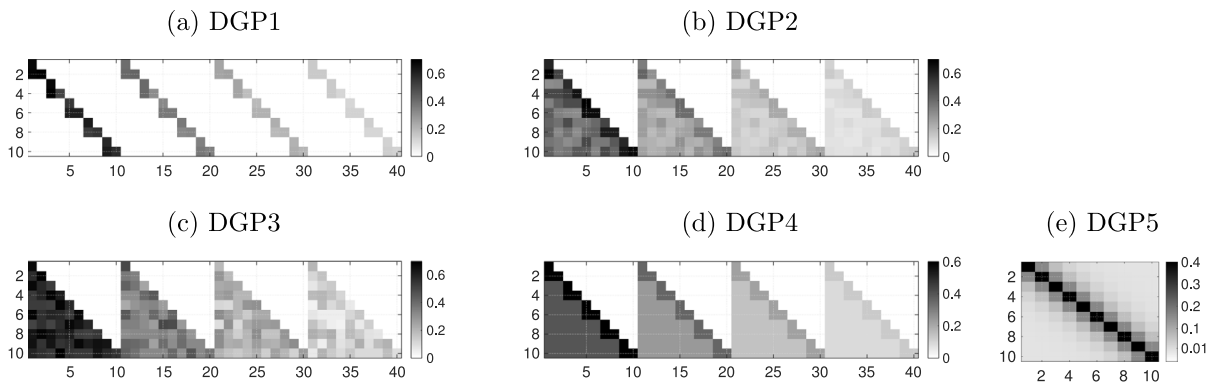

Fig. 1. Coefficient matrices of the five DGPs. NOTE: The figure shows autoregressive coefficients of the simulated DGPs. Zero coefficients are colored in white and non-zero coefficients in gray shades whereby negative values are multiplied by −1 and the darkest shade is given to the highest coefficient value.

The correct sparsity pattern calculates how often the evaluated procedure takes the correct decision whether to include or exclude a variable. The fraction of relevant variables included counts the number of true relevant variables included in the models relative to the number of all true non-zero coefficients. The number of variables included reports the average number of variables included in the model. It evaluates the dimension reduction done by the estimator. The MSE of the parameter estimates for one Monte Carlo replication is calculated as

$$MSE = \frac{1}{NG^2p}\sum_{p=1}^{P}\sum_{i,j}^{N}\sum_{k,l}^{G}(\hat{b}_{klp}^{ij} - b_{klp}^{ij})^2$$

where $\hat{b}_{klp}^{ij}$ is the estimate of the true parameter b_{klp}^{ij}.

3.3. Simulation results

Tables 1–3 contain the evaluation of the various estimation procedures along the six performance criteria for the five DGPs for $T = 100$. The performance criteria are averages over 1000 Monte Carlo replications. Overall, the simulation studies provide supporting evidence that using LASSO techniques for estimating PVAR models is beneficial in terms of lower mean squared forecast errors, higher average logarithmic scores, and reduced mean squared errors relative to a PVAR model estimated with OLS.

The usage of the selection methods leads to a sizable reduction in mean squared forecast errors compared to OLS for all simulations for the presented one-, two-, and six-steps ahead forecasts, Table 1. The largest improvement is found for horizon one. The performance compared among the LASSO alternatives is similar.Using the cross-sectional shrinkage approach leads to smaller relative gains compared to the LASSO techniques. This also holds for the dense DGP despite the fact that the penalization techniques favor sparse and the factor approach dense solutions. Similarly, the increase in forecast performance of the LASSO techniques is larger than the improvement of a block-diagonal model estimated with OLS or a single unit VAR model estimated with OLS for sparse DGPs. For the dense DGP5, single-country models perform well as the DGP closely re-samples such a structure with exponentially decreasing coefficients from the main diagonal. The overall performance gain of the LASSO alternatives compared to *OLS* is smaller for DGP5 as the number of unknown parameters is considerably smaller compared to the sparse DGPs. In general, the findings are in line with the results of Smeekes and Wijler (2018), showing that the forecasting performance of penalized estimators is relatively robust against a variety of DGP specifications including dense set-ups.

Similar to the point forecasts, the predictive densities of the LASSO techniques for PVAR models have larger ALSs than the PVAR model estimated with OLS, see Table 2. Due to the extremely small values of the predictive densities for the PVAR model estimated with OLS the performance is not presented in relative terms to avoid numerical issues. The observed value has almost zero support in the constructed predictive density and hence, the log scores are set to minus infinity. Usually, the prediction is very close to zero, the variance of the in-sample residuals is small, while the observed value deviates from zero more than the maximal in-sample residual. This gives an indication that bootstrapping the empirical predictive densities based on the in-sample residuals might underestimate the uncertainty out-of-sample and thus generating empirical predictive densities with a very low variance. While the performance of the various LASSO techniques applied to PVAR models is similar, average log scores are highest for the cross-sectional shrinkage approach. Both the block-diagonal model and the single unit model estimated by OLS also slightly outperform the LASSO alternatives in terms of density forecasting accuracy. Table A-1 in Supplementary Appendix 2.1 shows that the performance gains relative to a PVAR model estimated with LASSO including equation-specific, lag-specific and foreign cross-section penalties, λ_k, α, c, and weighted sum of squared residuals are only significant at the 10% level for some of the one-step ahead forecasts of the cross-sectional shrinkage approach but mainly significant at the 5% level for the block diagonal model. The differences among the

Table 1
Mean squared forecast errors relative to OLS.

	LASSO techniques						Single unit	
	λ_k, c, α	λ_k, c, α $\Sigma = I$	$\lambda_k, c, \alpha, \gamma$ *(AG)*	λ_k $\Sigma = I$	*CC*	*bl-diag OLS*	*OLS*	*LASSO* λ_k
One-step ahead mean squared forecast errors								
DGP1	0.49***	0.48***	0.50***	0.48***	0.56***	0.51***	0.52***	0.48***
DGP2	0.49***	0.48***	0.51***	0.48***	0.56***	0.51***	0.51***	0.48***
DGP3	0.47***	0.47***	0.50***	0.47***	0.54***	0.50***	0.51***	0.47***
DGP4	0.49***	0.48***	0.50***	0.48***	0.56***	0.51***	0.51***	0.48***
DGP5	0.90***	0.89***	0.91***	0.89***	0.96***	0.87***	0.88***	0.87***
Two-steps ahead mean squared forecast errors								
DGP1	0.58**	0.58**	0.58**	0.58**	0.64**	0.60**	0.60**	0.58**
DGP2	0.57**	0.57**	0.58**	0.57**	0.63**	0.59**	0.60**	0.57**
DGP3	0.57**	0.57**	0.57**	0.57**	0.62**	0.59**	0.60**	0.56**
DGP4	0.57**	0.57**	0.57**	0.57**	0.63**	0.59**	0.60**	0.57**
DGP5	0.98*	0.98*	0.98	0.98*	1.01	0.98*	0.98*	0.98*
Six-steps ahead mean squared forecast errors								
DGP1	0.78	0.78	0.78	0.78	0.93	0.78	0.78	0.78
DGP2	0.75	0.75	1.73	0.75	0.92	0.76	0.76	0.75
DGP3	0.74	0.74	0.74	0.74	1.26	0.74	0.74	0.74
DGP4	0.75	0.76	0.76	0.75	0.92	0.76	0.76	0.75
DGP5	1.00	1.00	1.00	1.00	1.01	1.00	1.00	1.00

NOTE: MSFEs are relative to a PVAR model estimated by OLS and average over all t, all units and variables and over 1000 Monte Carlo replications. Estimators (in columns): LASSO with penalties λ_k, α, c and weighted sum of squared residuals, LASSO with the same penalty parameters where the covariance matrix is set to the identity matrix, LASSO with penalties $\lambda_k, \alpha, c, \gamma$ with aggregated estimator for homogeneous coefficients and with weighted sum of squared residuals, LASSO λ_k where the covariance matrix is set to the identity matrix, cross-sectional shrinkage approach, block-diagonal system ordering the variables in unit blocks estimated with OLS, separate VAR models for each unit estimated with OLS and with a LASSO with equation-specific shrinkage.

*** indicates 1% significance level of Diebold Mariano test.

** indicates 5% significance level of Diebold Mariano test.

* indicates 10% significance level of Diebold Mariano test.

Table 2
Average logarithmic scores.

	LASSO techniques						Single unit		
	λ_k, c, α	λ_k, c, α $\Sigma = I$	$\lambda_k, c, \alpha, \gamma$ *(AG)*	λ_k $\Sigma = I$	*CC*	*bl-diag OLS*	*OLS*	*LASSO* λ_k	*PVAR OLS*
One-step ahead average logarithmic scores									
DGP1	−1.83	−1.78	−1.74	−1.86	−1.64	−1.66	−1.72	−Inf	−Inf
DGP2	−1.83	−1.81	−1.75	−1.85	−1.64	−1.66	−1.72	−Inf	−Inf
DGP3	−1.75	−Inf	−1.74	−1.92	−1.62	−1.62	−1.68	−Inf	−Inf
DGP4	−1.84	−1.81	−1.74	−1.85	−1.64	−1.66	−1.72	−Inf	−Inf
DGP5	−1.63	−1.61	−1.62	−1.62	−1.62	−1.55	−1.57	−Inf	−1.84
Two-steps ahead average logarithmic scores									
DGP1	−1.68	−1.69	−1.74	−1.75	−1.62	−1.61	−1.67	−Inf	−Inf
DGP2	−1.73	−1.73	−1.74	−1.76	−1.62	−1.61	−1.67	−Inf	−Inf
DGP3	−1.78	−1.86	−1.75	−Inf	−1.60	−1.62	−1.69	−Inf	−Inf
DGP4	−1.74	−1.73	−1.74	−1.77	−1.62	−1.61	−1.67	−Inf	−Inf
DGP5	−1.53	−1.53	−1.53	−1.54	−1.52	−1.51	−1.52	−Inf	−1.59
Six-steps ahead average logarithmic scores									
DGP1	−1.70	−1.69	−1.74	−1.73	−1.82	−1.59	−1.63	−Inf	−Inf
DGP2	−1.69	−1.69	−Inf	−1.72	−1.88	−1.59	−1.63	−Inf	−Inf
DGP3	−1.67	−1.67	−1.74	−1.70	−Inf	−1.59	−1.63	−Inf	−Inf
DGP4	−1.69	−1.68	−1.73	−1.72	−1.89	−1.59	−1.63	−Inf	−Inf
DGP5	−1.55	−1.55	−1.53	−1.55	−1.51	−1.52	−1.53	−Inf	−1.58

NOTE: ALSs are averaged over 1000 Monte Carlo replications and calculated on predictive densities obtained from 10,000 bootstrapped forecasts (randomly drawn observation from the in-sample residuals added to the h-step ahead point forecast). Estimators (in columns): LASSO with penalties λ_k, α, c and weighted sum of squared residuals, LASSO with the same penalty parameters where the covariance matrix is set to the identity matrix, LASSO with penalties $\lambda_k, \alpha, c, \gamma$ with aggregated estimator for homogeneous coefficients and with weighted sum of squared residuals, LASSO λ_k where the covariance matrix is set to the identity matrix, cross-sectional shrinkage approach, block-diagonal system ordering the variables in unit blocks estimated with OLS, separate VAR models for each unit estimated with OLS, with a LASSO with equation-specific shrinkage, and a PVAR estimated with OLS. The log scores are set to minus infinity if the observed value has almost zero support in the constructed predictive density.

Table 3
Performance evaluation of estimators.

	LASSO techniques						Single unit		
	λ_k, c, α	λ_k, c, α $\Sigma = I$	$\lambda_k, c, \alpha, \gamma$ *(AG)*	λ_k $\Sigma = I$	*CC*	*bl-diag* *OLS*	*OLS*	*LASSO* λ_k	*PVAR* *OLS*
Correct sparsity pattern									
DGP1	0.21	0.19	0.17	0.22	0.85	0.05	0.05	0.13	0.85
DGP2	0.53	0.54	0.52	0.54	0.45	0.45	0.45	0.53	0.45
DGP3	0.53	0.54	0.52	0.54	0.45	0.45	0.45	0.53	0.45
DGP4	0.53	0.54	0.52	0.54	0.45	0.45	0.45	0.53	0.45
DGP5	0.63	0.74	0.53	0.71	0.00	0.80	0.80	0.92	0.00
Fraction of relevant variables included									
DGP1	0.20	0.12	0.33	0.11	1.00	1.00	1.00	0.24	–
DGP2	0.11	0.07	0.13	0.09	1.00	0.27	0.27	0.06	–
DGP3	0.10	0.07	0.13	0.08	1.00	0.27	0.27	0.06	–
DGP4	0.11	0.07	0.12	0.09	1.00	0.27	0.27	0.06	–
DGP5	0.37	0.26	0.47	0.29	1.00	0.20	0.20	0.08	1.00
Number of variables included									
DGP1	46.35	30.91	47.82	42.84	400.00	80.00	80.00	19.06	400.00
DGP2	40.13	27.15	42.99	35.37	400.00	80.00	80.00	18.33	400.00
DGP3	34.56	24.32	43.01	28.80	400.00	80.00	80.00	18.16	400.00
DGP4	40.42	27.21	42.76	35.31	400.00	80.00	80.00	18.39	400.00
DGP5	36.65	25.86	46.58	29.48	100.00	20.00	20.00	8.33	100.00
Mean squared error									
DGP1	0.02	0.02	0.02	0.02	0.03	0.02	0.03	0.02	0.08
DGP2	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.12
DGP3	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.16
DGP4	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.12
DGP5	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.02	0.05

NOTE: All measures are averaged over 1000 Monte Carlo replications. Estimators (in columns): LASSO with penalties λ_k, α, c and weighted sum of squared residuals, LASSO with the same penalty parameters where the covariance matrix is set to the identity matrix, LASSO with penalties $\lambda_k, \alpha, c, \gamma$ with aggregated estimator for homogeneous coefficients and with weighted sum of squared residuals, LASSO λ_k where the covariance matrix is set to the identity matrix, cross-sectional shrinkage approach, block-diagonal system ordering the variables in unit blocks estimated with OLS, separate VAR models for each unit estimated with OLS, with a LASSO with equation-specific shrinkage, and a PVAR estimated with OLS.

LASSO techniques are mostly insignificant. Overall, the performance differences across forecasting horizons as well as across the first four DGPs are small for most of the models.

The LASSO alternative with λ_k, c and α includes true relevant and discards irrelevant variables in the same range as the other LASSO techniques, shown in Table 3. The number of detection of the correct sparsity pattern as well as the fraction of relevant variables varies among the five DGPs. Compared to DGP1, the LASSO techniques find the correct sparsity pattern more often for the last four DGPs with the best performance for DGP5. The cross-sectional shrinkage approach condenses the information into factors. Thus, the correct sparsity pattern, fraction of variables included, and number of variables included of this approach are the same as for the unrestricted PVAR model estimated by OLS.

The LASSO techniques clearly reduce the dimension of the models indicated by the number of variables included. Compared to the benchmark OLS, all estimators reveal lower mean squared errors in all simulations. The weaker performance of the PVAR model estimated with OLS in terms of MSE reflects the problem of overfitting.

Results for $T = 500$ and $N = 5, G = 5$ are presented in the Supplementary Appendix. As expected, the performance among the LASSO techniques and the least squares estimators become align for $T = 500$ while for $N = 5, G = 5$ the outperformance of the LASSO techniques with respect to the least squares alternatives is pronounced.

4. Forecasting and international spillovers with PVAR models

PVAR models are well suited as forecasting models for macroeconomic time series since they allow for international interdependencies and commonalities. Several studies report good forecasting performance of models that account for international dependencies while forecasting national and international key macroeconomic variables (e.g., Bjørnland et al., 2017; Crespo Cuaresma et al., 2016; Pesaran et al., 2009). However, for accurate point forecasts the over parameterization of PVAR models needs to be addressed. The dimensionality reduction implemented with the proposed LASSO for PVAR models leads to efficiency gains in the estimation and, thus, to smaller forecasting errors.

Furthermore, due to their characteristics PVAR models allow for rich spillover analyses taking into account multiple countries and macroeconomic variables. Recent literature stresses the existence of global financial connectedness and international co-movement of business cycles (e.g., Demirer et al., 2018; Diebold & Yilmaz, 2014, 2015; Greenwood-Nimmo & Nguyen, 2015; Louzis, 2015).

I evaluate the forecast performance of the penalized estimators for PVAR models and analyse international macroeconomic spillovers for two sets of variables and countries. The first PVAR model includes monthly log changes in the harmonized index of consumer prices (CPI) and industrial production growth (IP) for five countries: Germany (DE), France (FR), Italy (IT), the United Kingdom (UK), and the United States (US). The second model includes additionally real effective exchange rates (REER) and unemployment rates (UN) for 20 countries.[4] The time series are seasonally adjusted, de-meaned and standardized. The data are from the OECD and cover the period from 2001:1 to 2016:6. The models with two variables and five countries includes six lags, the ones with four variables and 20 countries two lags. I show in Supplementary Appendix 3.3. that the results are robust to choosing twelve lags.

I split the sample into initialization, penalty parameter selection and forecast evaluation according to $T_1 = T_2 - 30$ and $T_2 = T - 60$. Thus, the out-of-sample forecasting period runs from 2011:7 to 2016:6. The iterated forecasts are calculated by $\hat{Y}_{t+h} = \hat{B}\hat{X}_{t+h-1}$ for $h = 1, \ldots, 12$ using a rolling window. The results of Marcellino et al. (2006) motivate the choice of performing iterated rather than direct forecasts. The forecasts are evaluated by mean squared forecast errors.

I quantify international spillovers with the connectedness concept of Diebold and Yilmaz (2014) which is based on the generalized variance decomposition. Pairwise connectedness of a variable k of country j to variable l of country i for the first horizon is calculated as

$$C_{ij}^{lk} = \frac{(\sigma_{jj}^{kk})^{-1}(e_i^{l\prime} A_0^{-1} \Sigma e_j^k)^2}{e_i^{l\prime} A_0^{-1} \Sigma A_0^{-1\prime} e_i^l} \tag{7}$$

where e_i^l is a selection vector of zeros and one element equal to one at the position of variable l of country i. The variance of variable k of country j is labelled as σ_{jj}^{kk}. The matrix A_0 measures contemporaneous relations, therefore it holds that $A_0^{-1} = chol(\Sigma)$. The connectedness measures is normalized by

$$\tilde{C}_{ij}^{lk} = \frac{C_{ij}^{lk}}{\sum_{k=1}^{G}\sum_{j=1}^{N} C_{ij}^{lk}}. \tag{8}$$

$\tilde{C}_{ij}^{lk}$ measures the (normalized) fraction of the one-step ahead forecast error variance of variable l of country i explained by a shock to variable k of country j.

Optimal penalty parameters for the LASSO techniques are selected via the cross-validation procedure. The grid of λ_k consists of twelve values between 0.01 and $(1/T)\lambda_k^{max} = \max(\max(XY'))$. The grids for α, c, γ and ρ have a length of four. For each equation the optimal penalty parameters minimize the MSFEs. See Appendix C and the Supplementary Appendix for details on the penalty selection.

[4] The 20 countries are Austria (AT), Belgium (BE), Germany (DE), Denmark (DK), Spain (ES), Finland (FI), France (FR), Greece (GR), Ireland (IE), Italy (IT), Luxembourg (LU), Netherlands (NL), Norway (NO), Poland (PL), Portugal (PT), Sweden (SE), Slovenia (SI), Slovakia (SK), United Kingdom (UK), and United States (US).

In addition to the alternative models described in 2.4, I consider a LASSO with weighted sum of squared residuals and equation-specific shrinkage by λ_k and two Bayesian alternative estimation procedures for PVAR models, the Bayesian selection prior of Koop and Korobilis (2016) called stochastic search specification selection, *SSSS*, and the Bayesian dynamic learning approach of Koop and Korobilis (2019), *KK*. The use of the selection prior is limited since it relies on a Markov Chain Monte Carlo algorithm not suitable for large systems. Furthermore, I compare the forecasting performance to a Bayesian GVAR model with a stochastic search variable selection prior (SSVS) as specified in Crespo Cuaresma et al. (2016). The Supplementary Appendix 3.1 gives implementation details on the Bayesian methods and the GVAR model.

4.1. Application results

Table 4 presents variable- and country-specific one-step ahead mean squared forecast errors relative to a PVAR model estimated with LASSO including equation-specific, lag-specific and foreign cross-section penalties, λ_k, α, c, and weighted sum of squared residuals. The stars in the tables indicate the significance levels of Diebold Mariano tests. Variable-specific MSFEs for each country for one-step, two-steps and six-steps forecasts are given in Fig. 2 for the application including CPI and IP of five countries.

The LASSO alternatives with λ_k, c and α (with estimated covariance or $\Sigma = I$) produce on average the lowest one-step ahead mean squared forecast errors among all LASSO procedures. The slight improvement of the LASSO version with $\Sigma = I$ can be explained by the reduced uncertainty due to setting the covariance matrix equal to the identity matrix. For the smaller application (with two variables and five countries), the use of LASSO techniques for PVAR models improves forecast performance relative to OLS. On average, the statistically significant loss in forecast performance when using OLS to estimate a PVAR model relative to the LASSO with λ_k, c and α is 1.76. OLS is not feasible for the application with four variables of 20 countries since the number of parameters per equation exceeds the number of observations.

Accounting for the time series and cross-sectional characteristics in the penalty terms leads to gains in the forecast accuracy. The LASSO techniques which also include penalty parameters c and α outperform on average the LASSO variants with λ_k. The latter one can be interpreted as a large VAR model, thus, an unrestricted PVAR model. This result is in line with empirical evidence in the Bayesian literature showing that, in general, PVAR models outperform unrestricted large BVAR models. For example, Korobilis (2016) reports that PVAR models increase the forecasting performance relative to BVAR models when forecasting 10-y bond yield spreads of ten countries. Forecasting inflation for ten countries, Koop and Korobilis (2019) show that PVAR models estimated with their proposed Bayesian dynamic learning approach increases forecasting performance on

Table 4
One-step ahead mean squared forecast errors.

	LASSO techniques				Bayesian					Single unit		
	$\lambda_k, c, \alpha, \Sigma = I$	$\lambda_k, c, \alpha, \gamma$	λ_k	λ_k $\Sigma = I$	*SSSS*	*CC*	*KK*	*GVAR*	*bl-dg OLS*	*OLS*	λ_k	*PVAR OLS*
	Application with two variables of five countries											
	Variable-specific mean squared forecast errors											
CPI	1.00	1.06**	1.01	1.02	2.55***	0.99	1.34	1.17***	1.06	1.11*	1.03	1.76***
IP	0.96***	1.83***	1.05**	1.09**	4.03***	1.05	1.24	1.18*	0.98	1.12	0.98	1.77***
	Country-specific mean squared forecast errors											
DE	0.99	1.53***	1.07**	1.06	2.09***	0.96	1.17	1.02	1.03	1.17**	1.02	1.87***
FR	1.01	1.51***	1.00	1.02	3.77***	1.10	2.38***	1.22**	1.15*	1.28***	1.07	1.64***
IT	1.00	1.91***	1.02*	1.01	3.96***	1.14	0.76	1.50***	1.16	1.31	1.17*	1.38**
UK	0.98***	1.24**	1.01	1.03	3.78***	1.14***	0.94	1.37***	0.94	0.96	1.00	1.58***
US	0.94**	1.05	1.07**	1.15**	2.86***	0.77**	1.21	0.75***	0.82***	0.85**	0.77***	2.34***
	Mean squared forecast errors averaged over countries and variables											
Av	0.98**	1.45***	1.03***	1.05***	3.29***	1.02	1.29***	1.17***	1.02	1.11*	1.01	1.76***
Av_b	1.00	1.77***	1.05***	1.09***	3.49***	1.13	1.22***	1.23**	1.06	1.21	1.10	2.06***
Av_r	0.97**	1.22***	1.02	1.03	3.26***	0.97	1.44***	1.13**	1.02	1.07	0.94	1.62***
	Application with four variables of 20 countries											
	Variable-specific mean squared forecast errors											
CPI	0.99	1.27***	1.06*	0.99			1.28	1.03	1.27***	1.00	1.01	
IP	0.99	1.34***	1.02	1.01			1.31	1.08**	1.13**	1.01	0.99	
REER	0.87***	1.08	1.27***	0.88***			1.00***	0.95	0.98	0.89**	0.86***	
UN	1.05**	1.11***	1.03***	1.02			2.18**	1.44***	1.53***	1.02	1.01	
	Mean squared forecast errors averaged over countries and variables											
Av	0.97***	1.20***	1.10***	0.98**			1.44	1.12	1.22***	0.98***	0.97***	
Av_b	0.99	1.20***	1.10***	0.99			1.59	1.12	1.25***	0.98*	0.98*	
Av_r	0.96***	1.26**	1.10***	0.97***			1.48	1.13	1.24**	0.98***	0.97***	

NOTE: The forecast period is from 2011:7 to 2016:6. MSFEs are relative to LASSO with penalties λ_k, α, c and weighted sum of squared residuals and are averaged over all t. Subscript b indicates boom periods, r recession. Estimators (in columns): LASSO with the same penalty parameters where the covariance matrix is set to the identity matrix, LASSO with penalties $\lambda_k, \alpha, c, \gamma$ with aggregated estimator for homogeneous coefficients and with weighted sum of squared residuals, LASSO with weighted sum of squared residuals and λ_k, LASSO λ_k where the covariance matrix is set to the identity matrix, stochastic search specification selection, cross-sectional shrinkage approach, Bayesian dynamic learning approach, Bayesian GVAR with SSVS prior, block-diagonal system ordering the variables in unit blocks estimated with OLS, separate VAR models for each unit estimated with OLS and with a LASSO with equation-specific shrinkage, PVAR estimated with OLS.

*** indicates 1% significance level of Diebold Mariano test.

** indicates 5% significance level of Diebold Mariano test.

* indicates 10% significance level of Diebold Mariano test.

average compared to a BVAR and FAVAR model. Feldkircher et al. (2020) demonstrate that forecast accuracy improves for G7 countries when using PVAR models estimated with a coefficient-pooling prior compared to BVAR models. Likewise, Koop and Korobilis (2016) demonstrate that an unrestricted PVAR model (i.e. a large BVAR) performs weaker than a model estimated with the stochastic search specification selection prior when comparing in-sample fit measures for ten countries. Table A-13 in Supplementary Appendix 3.3.1 gives a comprehensive overview of the empirical findings in the literature.

The LASSO with λ_k, c and α outperforms on average Bayesian methods for estimating PVAR models, block-diagonal VAR models estimated with OLS and - for the smaller application - single unit VAR models. While compressing information into factors as done by the approach of Canova and Ciccarelli (2009), *CC*, yields comparable forecasting accuracy to the LASSO, the Bayesian selection prior of Koop and Korobilis (2016), *SSSS*, performs considerably worse. The performance of *SSSS* might be improvable depending on the setting of the hyperparameters. Accounting for the specified factor structure and time-variation in the error covariance matrix as done in the Bayesian dynamic learning approach, *KK*, does not lead to an increase in forecasting performance. Hence, the empirical literature showing that those PVAR models estimated with the considered Bayesian methods outperform BVAR models (such as Feldkircher et al., 2020; Koop & Korobilis, 2016, 2019; Korobilis, 2016) indicates that my proposed LASSO specific to PVAR models should be preferable to BVAR models when forecasting in multi-country settings.

Imposing homogeneity restrictions does not pay off in terms of forecasting accuracy. The performance of the LASSO allowing also for shrinkage towards homogeneity is weaker. Especially for this specification allowing γ varying over groups might be beneficial. However, this LASSO variant is already the computationally most demanding one. Similarly, the homogeneity restrictions as implemented in GVAR models do not lead to considerable improvements. The forecasting performance of the Bayesian GVAR model is on average weaker than of the PVAR model estimated with the benchmark LASSO. The performance difference is significant at the 1% level for

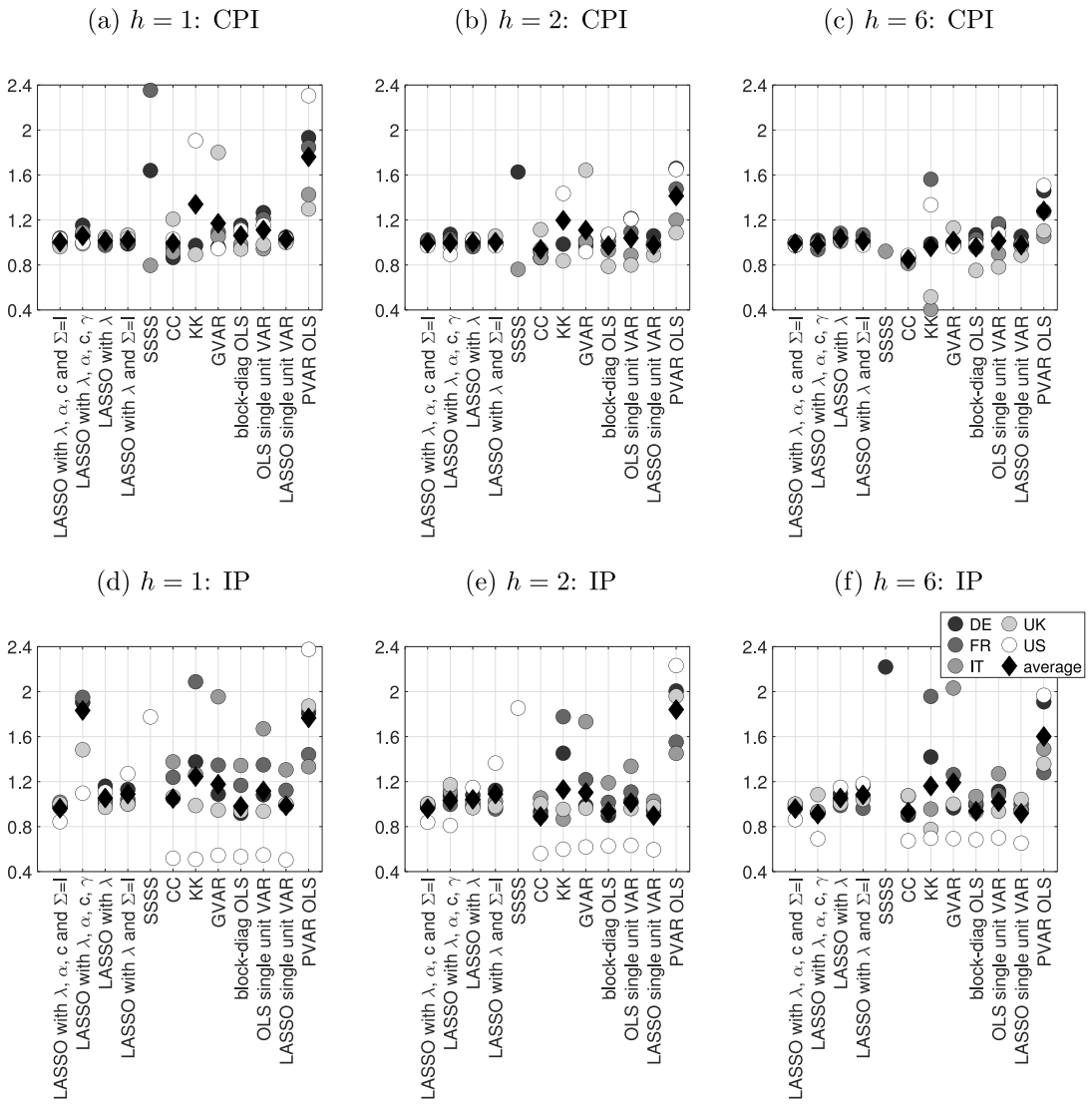

Fig. 2. One-step, two-steps, and six-steps ahead mean squared forecast errors. NOTE: The figure shows one-step, two-steps and six-steps ahead MSFEs relative to a PVAR model estimated with LASSO including equation-specific, lag-specific, foreign cross-section penalties, λ_k, α, c, and weighted sum of squared residuals for the application with two variables of five countries. Coloured circles denote variable-specific MSFEs for each country (averaged over the forecast period from 2011:7 to 2016:6) and diamonds the average across all countries.

the application with two variables of five countries (average relative one-step ahead MSFE of 1.17).

These findings are in line with the literature providing evidence that GVAR models perform in general weaker than BVAR and PVAR models. For example, Alessi and Banbura (2009) show that forecasts of key macroeconomic variables for 33 countries obtained with BVAR models are more accurate than forecasts of GVAR and dynamic factor models. Giannone and Reichlin (2009), reporting the results of Alessi and Banbura (2009), discuss that while BVAR models outperform GVAR models on average, disaggregated results for countries and variables also show some variation in the performance. Using data for 36 countries on several macroeconomic variables, Crespo Cuaresma et al. (2016) demonstrate that Bayesian GVAR models can only outperform BVAR models in terms of forecasting accuracy in some cases depending on the prior specification chosen and on the aggregation level. Feldkircher et al. (2020) provide evidence that GVAR models loose on average forecasting accuracy compared to PVAR and BVAR models. They forecast average inflation, however, more accurate than BVAR models but less precise than PVAR models.

For the smaller application, I find the largest gain in variable-specific forecast performance for one-step ahead industrial production growth forecasts for the LASSO variant with with λ_k, c, α and $\Sigma = I$ (first column). While several alternative models have more accurate forecasts

for REER, they have considerably worse forecasts as the benchmark LASSO for UN. The mean squared forecast errors are particularly low for the US for selection methods compared to OLS. Variables of other countries have a low impact on US variables, thus, including these variables does not seem to improve the forecasts for the US. Similarly, while the benchmark model outperforms the GVAR in four out of five countries, the GVAR model has the smallest MSFE of all models for the US.

Fig. 2 provides a more detailed picture of the performance across countries and variables. It shows MSFEs for one-step, two-steps and six-steps ahead forecasts errors relative to a PVAR model estimated with LASSO including equation-specific, lag-specific, foreign cross-section penalties, λ_k, α, c, and weighted sum of squared residuals for the application with two variables of five countries. The graphs denote variable-specific MSFEs for each country (averaged over the forecast period from 2011:7 to 2016:6) by a coloured circle and denote the average across all countries with a diamond. The forecast performance across countries varies less for LASSO techniques compared to models estimated with OLS (block-diagonal OLS, single VAR models estimated with OLS and PVAR estimated with OLS) and with the Bayesian dynamic learning approach or GVAR models. The variation across countries for all estimators is higher for IP than CPI forecasts. Forecasts of US industrial production growth seem not to improve when international information is added. MSFEs for US industrial production growth are especially low for models estimated with LASSO techniques which implement higher shrinkage on foreign variables and models which do not include an international dimension. A LASSO with a penalty parameter which only varies across equations cannot account for this fact. In general, the relative performance of the estimators is comparable among different forecast horizons. Further results for additional forecast horizons, given in the Supplementary Appendix 3.3, support the good forecast accuracy of LASSO techniques for PVAR models compared to alternative methods.

The results provide some evidence that the use of multi-country models estimated with dimension-reduction techniques compared to models which exclude international dependencies is beneficial to improve forecast performance. MSFEs of models accounting for interdependencies across countries are lower than MSFEs of block-diagonal VAR models estimated with OLS and in some cases lower than MSEFs of single unit VARs estimated with OLS or LASSO.

The LASSO with penalties λ_k, c and α and weighted sum of squared residuals performs in particular well during expansion periods. The values Av_b and Av_r in Table 4 show one-step ahead MSFEs divided into measures for boom and recession periods, denoted by subscript b and r, respectively. Observations are classified as belonging to boom periods if the country's industrial production growth rate is larger or equal to zero, while periods with negative growth rates are labeled as recession periods. While this is an adhoc division, it still provides interesting insight regarding the forecasting performance depending on the state of the economy. The benchmark LASSO significantly outperforms five of the competing models measured by average MSFEs. The *CC* approach has an advantage in recession periods - although the performance difference is not statistically significant. Overall, the relative performances among models are stable with respect to the economic environment. Changes in the performance difference between the LASSO with penalties λ_k, α, c in weighted sum of squared residuals and the respective competitor are similar for both states of the economy. Typically, only the size of the relative difference and the significance level changes, depending on the state of the economy. The findings are comparable for multi-step ahead forecasts, as presented in Supplementary Appendix 3.3.

Turning to density forecasts, Table 5 shows one-step ahead average logarithmic scores relative to the benchmark LASSO with penalties λ_k, c and α and weighted sum of squared residuals. The stars in the table indicate the significance levels of Amisano and Giacomini (2007) tests. For the application with two variables of five countries, differences across methods are less pronounced when comparing density forecasts. The LASSO with λ_k, c and α clearly outperforms an PVAR model estimated with OLS, *SSSS* (both significantly at 1% level), the Bayesian GVAR model, and the single country model estimated with LASSO (the "-" denotes an absolute ALS equal to minus infinity).[5] Similar to the point forecast results, the LASSO penalizing heterogeneity is the weakest method among the LASSO techniques applied to PVAR models. ALSs for inflation density forecasts are comparable among most of the models (for eight models the differences are not statistically significant). The density forecasting performance including four variables of 20 countries differs substantially across variables. REER sticks out as almost all alternative estimators outperform the benchmark LASSO (with penalties λ_k, c and α and weighted sum of squared residuals). The ALSs for REER of the benchmark LASSO are small (large negative numbers) driven by close to zero point forecasts, small in-sample variation, and observed values which are non-zero. The point forecasts evaluation reflects this weak performance for REER as well (six methods outperform the benchmark). For CPI, IP and UN none of the estimators improves the density forecasting performance significantly with respect to the LASSO with penalties λ_k, c and α and weighted sum of squared residuals. The average measures reflect the weak performance for REER.

Performance gains for density forecasts using the LASSO with penalties λ_k, α, c and weighted sum of squared residuals are on average larger in boom than recession periods. That might be explained by less uncertainty in boom periods. The standard deviations of logarithmic scores of the benchmark model over the

[5] Relative ALSs for all models are calculated on predictive densities obtained from 10,000 bootstrapped forecasts (randomly drawn observation from the in-sample residuals added to the h-step ahead point forecast). Commonly, density forecasts for the Bayesian methods are simulated from the posterior distributions. To ensure comparability, I calculated the ALSs for all methods based on the bootstrap procedure.

Table 5
One-step ahead average logarithmic scores.

	LASSO techniques				Bayesian					Single unit		
	$\lambda_k, c, \alpha, \Sigma = I$	$\lambda_k, c, \alpha, \gamma$	λ_k	λ_k $\Sigma = I$	*SSSS*	*CC*	*KK*	*GVAR*	*bl-dg OLS*	*OLS*	λ_k	*PVAR OLS*
	Application with two variables of five countries											
	Variable-specific average logarithmic scores											
CPI	−0.01	−0.12	0.03	0.00	−0.26	0.00	−0.08	–	0.15	0.11	–	−4.4**
IP	0.06*	−0.87***	−0.09	−0.24	−0.63***	−0.07	0.01	−0.05	0.06	−0.04	−0.07	−2.98***
	Country-specific average logarithmic scores											
DE	0.00	−0.66***	−0.07	−0.03	−0.38***	0.01	−0.04	–	−0.01	−0.04	−13.31**	−4.28**
FR	−0.02	−0.61***	0.04	0.04	−0.64***	−0.03	−0.05	–	−0.11**	−0.12***	–	−0.98*
IT	0.02	0.01	0.00	−0.05	0.02	−0.05	−0.32	–	0.45*	0.28	–	−5.62
UK	−0.01	−0.63***	0.03	0.05	−0.73***	−0.29	−0.19**	–	0.08	−0.09	–	−2.72*
US	0.14*	−0.58***	−0.16	−0.62	−0.5***	0.19*	0.17	−0.12	0.12	0.15	−0.18	−4.85***
	Average logarithmic scores averaged over countries and variables											
Av	0.02	−0.5***	−0.03	−0.12	−0.44***	−0.03	−0.08	–	0.11	0.03	–	−3.69***
Av_b	0.02	−0.32***	−0.06	−0.14	−0.27***	0.02	0.00	0.00	0.05	0.03	–	−1.56***
Av_r	0.01	−0.17**	0.02	0.02	−0.18**	−0.06	0.03	–	0.06	0.00	–	−2.13**
	Application with four variables of 20 countries											
	Variable-specific average logarithmic scores											
CPI	0.49	−0.40	−2.59***	0.55			0.29	–	0.60	0.23	–	
IP	0.00	−1.15***	0.05	−0.02			−0.10	−0.24***	−0.05	−0.09	−0.25***	
REER	1.03***	50.10***	–	8.88***			0.91***	1.00***	0.91***	1.00***	37.41	
UN	–	−0.50***	−0.19**	−0.02			−0.45***	–	0.03	0.14	–	
	Average logarithmic scores averaged over countries and variables											
Av	–	0.90*	–	1.86***			1.66***	–	1.88***	1.81***	–	
Av_b	–	0.25	–	0.81***			0.70***	–	0.79***	0.78***	–	
Av_r	–	0.60*	–	1.05***			0.96***	–	1.09***	1.03***	–	

NOTE: The forecast period is from 2011:7 to 2016:6. Relative ALSs are calculated on predictive densities obtained from 10,000 bootstrapped forecasts (randomly drawn observation from the in-sample residuals added to the h-step ahead point forecast). ALSs are relative to LASSO with penalties λ_k, α, c and weighted sum of squared residuals and are averaged over all t. Subscript b indicates boom periods, r recession. Estimators (in columns): LASSO with the same penalty parameters where the covariance matrix is set to the identity matrix, LASSO with penalties $\lambda_k, \alpha, c, \gamma$ with aggregated estimator for homogeneous coefficients and with weighted sum of squared residuals, LASSO with weighted sum of squared residuals and λ_k, LASSO λ_k where the covariance matrix is set to the identity matrix, stochastic search specification selection, cross-sectional shrinkage approach, Bayesian dynamic learning approach, Bayesian GVAR with SSVS prior, block-diagonal system ordering the variables in unit blocks estimated with OLS, separate VAR models for each unit estimated with OLS and with a LASSO with equation-specific shrinkage, PVAR estimated with OLS. The "–" denote scores which are zero (log score equals minus infinity).

*** indicates 1% significance level of Amisano and Giacomini (2007) test.

** indicates 5% significance level of Amisano and Giacomini (2007) test.

* indicates 10% significance level of Amisano and Giacomini (2007) test.

rolling windows is considerably lower for boom, 0.4, than for recession, 0.92, periods. The relative density forecasting accuracy is stable across multi-step ahead forecasts. That is, the same models have weaker or comparable forecasting performance relative to the benchmark LASSO applied to PVAR models for first horizon forecasts as for higher horizons, as shown in tables A-8 to A-11 in Supplementary Appendix 3.3.

Fig. 3 shows the forecasting performance over time for the LASSO with penalties λ_k, α, c and *CC* as one of the best competitors. The graph plots the time series (first row), one-step ahead squared forecast errors (second row), and logarithmic scores (third row) for different variables and countries over the out-of sample forecasting period. I focus on UK and US as countries where *CC* once produces significantly lower (US) and once higher (UK) point forecasts relative to the LASSO. The plots of the remaining countries are given in Supplementary Appendix 3.3. In general, the performance over time is similar for both methods, with high squared forecast errors and low logarithmic scores for more extreme observations. That is, for example, visible for the spikes in UK IP growth in end of 2012 and US inflation in January 2015. Two points stick out: first, *CC* performs relatively week in forecasting UK IP growth in end of 2012 and, second, *CC* outperforms the LASSO in forecasting US IP growth rates in January 2012. These differences across the methods mainly contribute to the different average performance for UK and US as shown in Table 4.

Fig. 4 gives the sparsity patterns of the autoregressive parameters. The graphs show autoregressive parameters of first and second lag for PVAR models estimated with LASSO with penalties λ_k, α, c and weighted sum of squared residuals. Zero coefficients are coloured in white and non-zero coefficients in grey shades whereby negative values are multiplied by −1 and the darkest shade is given to the highest coefficient value. The grids in the graphs divide into country blocks. Each country block consists of inflation and industrial production growth for both applications and additionally of the real effective

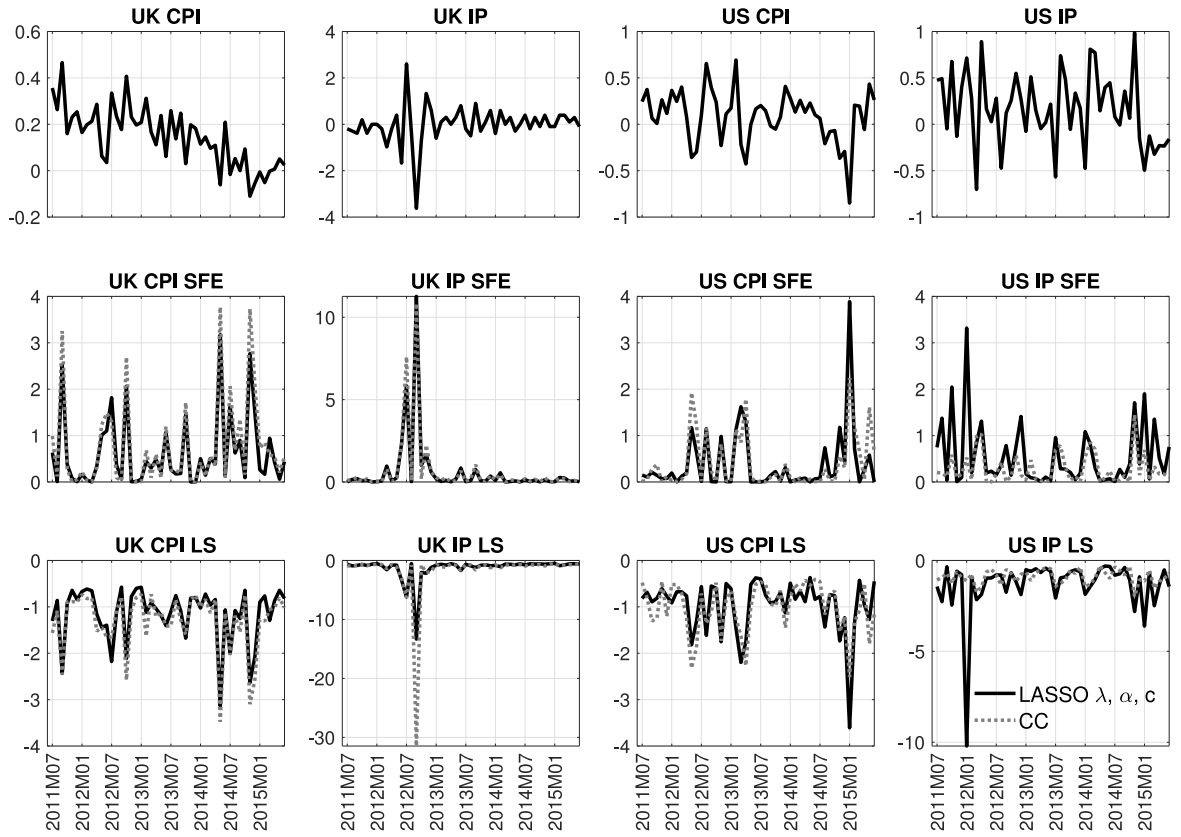

Fig. 3. One-step ahead squared forecast errors and logarithmic scores over time. NOTE: The figure shows time series (first row), one-step ahead squared forecast errors (second row), and logarithmic scores (third row) for CPI and IP of the UK and US for the out-of sample forecasting period. The models are estimated with the LASSO with penalties λ_k, α, c and *CC*.

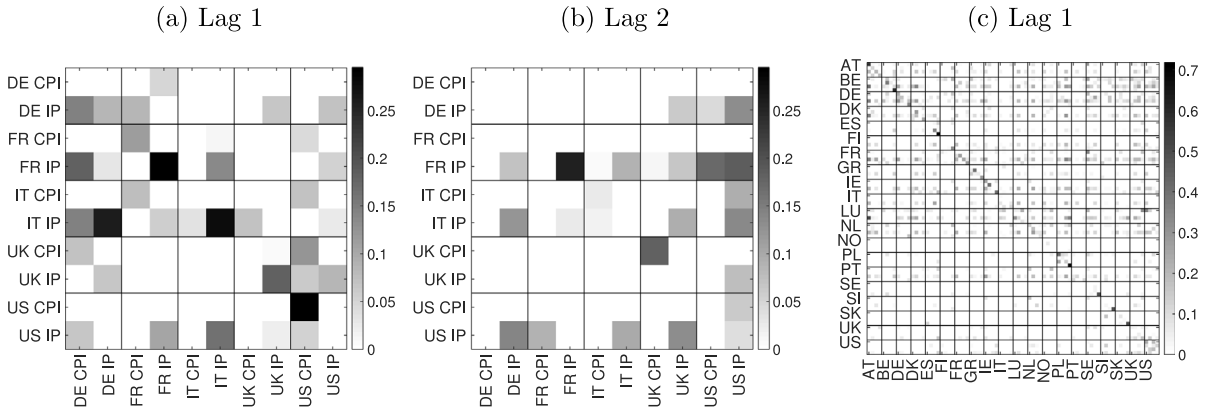

Fig. 4. Sparsity pattern of the autoregressive parameters. NOTE: The figure shows autoregressive parameters of first and second lag for PVAR models estimated with LASSO with penalties λ_k, α, c and weighted sum of squared residuals. Zero coefficients are coloured in white and non-zero coefficients in grey shades whereby negative values are multiplied by -1 and the darkest shade is given to the highest coefficient value. The grids in the graphs divide into country blocks.

exchange rate and the unemployment rate for the larger application.

Own lags have the largest impact, as shown by darker colours for diagonal elements. In addition, US variables affect variables of other countries. For some countries, foreign variables seem to add very little information, e.g., for the UK (in both application) and Spain, Norway, Poland, Sweden, Slovenia, and Slovakia. The variable

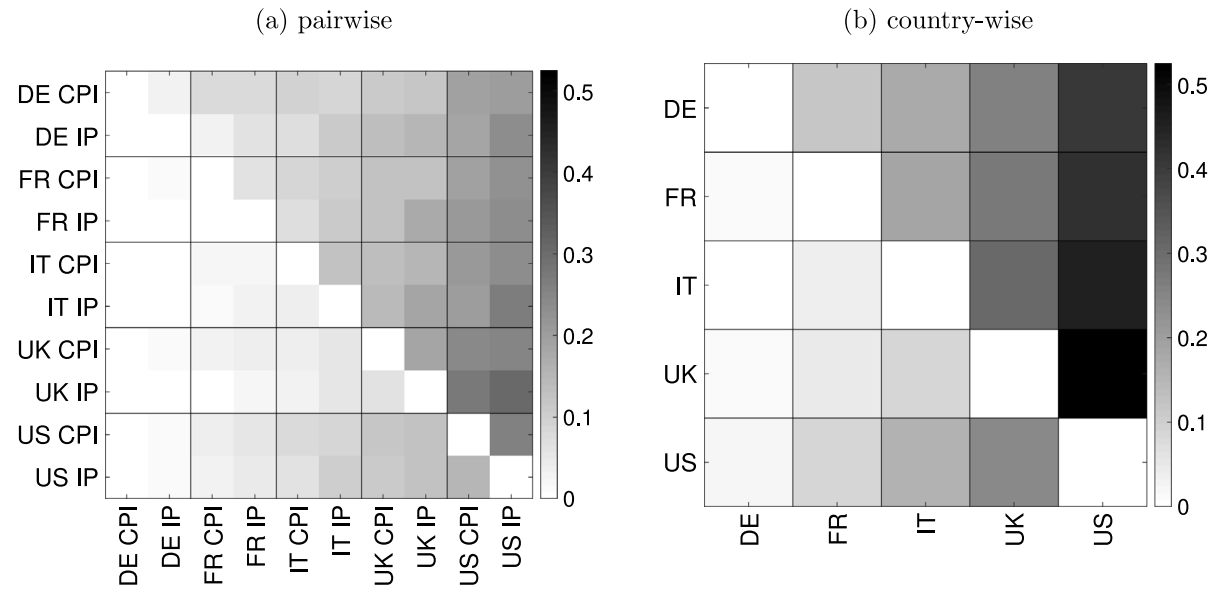

Fig. 5. Pairwise and country-wise connectedness for horizon one - application with two variables of five countries. NOTE: The figure shows pairwise (a) and country-wise (b) conntectedness for the application with two variables of five countries calculated based on a LASSO with penalties λ_k, α, c and weighted sum of squared residuals. Receiving variables of each country (or countries) are in rows and transmitting variables (or countries) in columns. The grids in the graphs divide into country blocks.

most impacted by foreign variables is the REER. This fact is visible by more non-zero (darker) elements for REER equations. The REER is on the third position within each country-block. So in Fig. 4c it can be seen that for Austria, Belgium, Germany, Denmark, France, Ireland, Italy, Luxembourg, The Netherlands, and Portugal the REER is more affected by foreign variables than in other countries. Unemployment, ordered at the last position for each country, is the variable least affected by other variables. This is visible for Belgium, Germany, Spain, Finland, Greece, Ireland, Poland, Sweden, Slovenia, and Slovakia. Here, the domestic variables (diagonal elements in the graph) have the highest impact (darkest colour) while foreign variables seem to have minor explanatory power.

The sparsity pattern is stable over time (Supplementary Appendix 3.3 presents the sparsity patterns of the first lag for each rolling window) and, thus, provides insights into the most important predictors. The combination of penalty parameters might help in guaranteeing a stable sparsity pattern since more conditions are imposed on the parameter vector. While the data support a sparse solution in autoregressive parameters, for static interdependences a dense solution is selected. The optimal penalty parameter (ρ) in the GLASSO estimation of the covariance matrix equals zero - implying no shrinkage. I exclude for the autoregressive parameters a dense solution since the penalty parameter grid starts with values above zero but the optimal selected penalty parameters are usually not the minimal values. Related to the results of Giannone et al. (2021) who find that for a cross-country model (their macro 2 example includes GDP growth rates for 90 countries) a dense model is favored, my findings provide some evidence that the dense relations stem from the static correlations while dynamic interdependencies across countries are sparse.

Figs. 5 and 6 illustrate the pairwise international connectedness for variables and the country-wise connectedness summing over all variables of one country. For the pairwise case (panel a) the graphs show receiving variables of each country in rows and transmitting variables in columns. The grids in the graphs divide into country blocks. For the country-wise case (panel b) rows and columns show receiving and transmitting countries, respectively. The PVAR models are estimated with LASSO with penalties λ_k, α, c and weighted sum of squared residuals and the covariance used to compute $\tilde{C}_{ij}^{lk}$ is estimated with graphical LASSO. In general, there exist strong connections of all countries to the US. Thus, for each country, a large fraction of the one-step ahead forecast error variance can be explained by a shock to US variables. Thus, the US acts as a global transmitter country. Differences across variables are less pronounced. In general, industrial production contributes more to the forecast error variances of other countries than other variables. This holds for both applications.

5. Conclusions

This paper develops LASSO techniques suitable for panel VAR models. The penalty terms incorporate both time series and cross section properties. The regularization constraints autoregressive parameters depending on the lags, parameters depending on the equation, parameters in the equation of another unit and heterogeneous parameters of same variables across units.

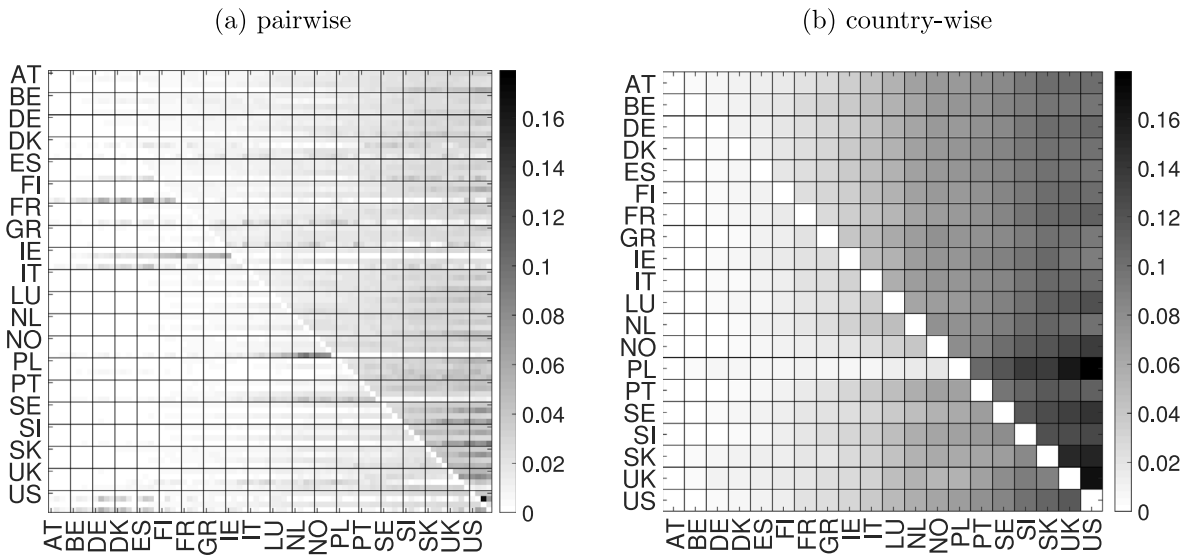

Fig. 6. Pairwise and country-wise connectedness for horizon one - application with four variables of 20 countries. NOTE: The figure shows pairwise (a) and country-wise (b) conntectedness for the application with four variables of 20 countries calculated based on a LASSO with penalties λ_k, α, c and weighted sum of squared residuals. Receiving variables of each country (or countries) are in rows and transmitting variables (or countries) in columns. The grids in the graphs divide into country blocks.

The main results of the paper are as follows. Simulation results show that estimating PVAR models with LASSO techniques achieves lower mean squared forecast errors and larger average logarithmic scores, thus increasing forecasting performance compared to estimating the PVAR models with OLS. A forecasting exercise that includes up to 20 advanced economies and four macroeconomic variables provides evidence that accounting for time series and cross section properties in the penalty term is beneficial for the forecast performance as a LASSO with penalization constraints specific to the nature of PVAR models outperforms a LASSO estimator without specific penalties. Compared to other Bayesian panel VAR methods, Bayesian GVAR models and single county models, the LASSO specific to PVAR models improves forecasts accuracy. The increase in forecasting performance is visible for point and density forecasts with largest gains in expansion periods and point forecast.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

I thank the Editor, two anonymous referees, Helmut Lütkepohl, Efrem Castelnuovo, Richard Paap, Andreas Pick, Wendun Wang and Tomasz Woźniak for insightful comments. Earlier versions of the paper were presented at SMYE 2017, Halle, Barcelona GSE Summer Forum 2017, Barcelona, EEA-ESEM 2017, Lisbon, annual meeting of the Verein für Socialpolitik 2017, Vienna, University of Sydney, University of Melbourne, Workshop Empirical Macroeconomics at Freie Universität Berlin, University of Konstanz, University of Vienna, Erasmus University Rotterdam, University of Groningen, University of Mannheim, Stockholm School of Economics, NESG 2019, Amsterdam, and at internal seminars at DIW Berlin. Comments by the participants are gratefully acknowledged. This work was carried out on the Dutch national e-infrastructure with the support of SURF Cooperative.

Appendix A. The LASSO estimator for PVAR models

In order to allow for shrinkage towards homogeneous coefficient, define $\underline{B} = B - \bar{B}$. The matrix $\bar{B}$ is given by

$$\bar{B} = (\bar{B}_1, 0_{NG\times NG(P-1)}) \text{ with}$$

$$\bar{B}_1 = \begin{bmatrix} \bar{b}_{11} & \cdots & \bar{b}_{1G} & & \cdots & \cdots & 0 \\ \vdots & \ddots & \vdots & & & & \vdots \\ \bar{b}_{G1} & \cdots & \bar{b}_{GG} & & & & \vdots \\ & & & \ddots & & & \\ \vdots & & & & \bar{b}_{11} & \cdots & \bar{b}_{1G} \\ \vdots & & & & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & & \bar{b}_{G1} & \cdots & \bar{b}_{GG} \end{bmatrix} \tag{9}$$

The elements on the block-diagonal, $\bar{b}_{kl}$ for $k, l = 1, \ldots, G$, are variable-specific coefficients which are homogeneous over all cross-sectional units. The coefficient $\bar{b}_{kl}$ refers to the impact of the lth variable on variable k.

We can rewrite (3) as

$$\underset{\underline{B}}{\operatorname{argmin}} \ \frac{1}{T} tr\left[(Y - \underline{B}X - \bar{B}X)'\Omega(Y - \underline{B}X - \bar{B}X)\right] + \Phi(\Lambda, \underline{B})$$
$$= \frac{1}{T} tr\left[(\underline{Y} - \underline{B}X)'\Omega(\underline{Y} - \underline{B}X)\right] + \Phi(\Lambda, \underline{B})$$

where $\underline{Y} = Y - \bar{B}X$. Since $\bar{B}$ is known when solving the optimization problem, the solution $\hat{\underline{B}}$ can be decomposed into $\hat{\underline{B}} + \bar{B} = \hat{B}$.

Minimizing the optimization problem in (6) yields the following estimator for b^{ij}_{klp} which is the element in B referring to the pth lag of variable l of unit j in the equation of variable k of unit i for $p = 1, \ldots, P$, $i, j = 1, \ldots, N$ and $k, l = 1, \ldots, G$.

$$\hat{b}^{ij}_{klp} = sign(\tilde{b}^{ij}_{klp}) \left(\left| \tilde{b}^{ij}_{klp} \right| - \frac{\lambda^{ij}_{kp} T}{2\omega_{kl} \sum_l^G \sum_p^P X^j_{lp,t} X^j_{lp,t}} \right) + \bar{b}_{klp} \tag{10}$$

with

$$\tilde{b}^{ij}_{klp} = \sum_t^T \sum_p^P \sum_{i,j}^N \sum_l^G \frac{\sum_{g \neq k}^G \omega^{ij}_{gl} \left(\underline{Y}^i_{g,t} - b^{ij}_{glp} X^j_{lp,t} \right) X^j_{lp,t}}{\omega_{kl} X^j_{lp,t} X^j_{lp,t}} + \frac{\left(\underline{Y}^i_{k,t} - \sum_{g \neq l}^G b^{ij}_{kgp} X^j_{gp,t} \right) X^j_{lp,t}}{X^j_{lp,t} X^j_{lp,t}} \tag{11}$$

and

$$\lambda^{ij}_{kp} = \begin{cases} \gamma & \text{if} \quad i = j, p = 1 \\ \lambda_k p^\alpha & \text{if} \quad i = j, p = 2, \ldots, P \ \text{ and } \ k = 1, \ldots, NG \\ \lambda_k p^\alpha c & \text{if} \quad i \neq j, p = 1, \ldots, P \ \text{ and } \ k = 1, \ldots, NG. \end{cases} \tag{12}$$

The ω^{ij}_{kl} is an element of the inverse of the covariance matrix, $\Sigma^{-1} = \Omega$, corresponding to the lth variable of unit j and the kth variable of unit i. The $\underline{Y}^i_{k,t}$ refers to the kth variable of unit i of $\underline{Y}$ in t, and $X^j_{lp,t}$ is the element in X referring to the pth lag of variable l of unit j in t. The $\bar{b}_{klp}$ is an element of $\bar{B}$.

Appendix B. Optimization algorithm

The optimization problem is solved by a coordinate descent algorithm as proposed in Friedman et al. (2007, 2010). As a starting value B is set equal to a zero matrix. The covariance is estimated in the GLASSO step. The homogeneous coefficients, $\bar{B}$, are computed with a mean-group or averaged estimator. The optimal penalty parameters are determined via a cross-validation technique minimizing MSFEs. The search of the optimal penalty parameters is done over grids of values. The algorithm updates every element b^{ij}_{klp}. The following steps are repeated until convergence is archived. Update b^{ij}_{klp} as follows:

1. Calculate $\tilde{b}^{ij}_{klp}$ according to (11).
2. Set the penalty parameter λ^{ij}_{kp} equal to the optimal chosen.
3. Calculate $\tilde{\lambda}^{ij}_{kp} = \frac{\lambda^{ij}_{kp} T}{2\omega_{kl} \sum_l^G \sum_p^P X^j_{lp,t} X^j_{lp,t}}$
4. Calculate $\hat{b}^{ij}_{klp}$ as

$$\hat{b}^{ij}_{klp} - \bar{b}_{klp} = \begin{cases} \tilde{b}^{ij}_{klp} - \tilde{\lambda}^{ij}_{kp} & \text{for} \ \ \tilde{b}^{ij}_{klp} > 0, \tilde{\lambda}^{ij}_{kp} < |\tilde{b}^{ij}_{klp}| \\ \tilde{b}^{ij}_{klp} + \tilde{\lambda}^{ij}_{kp} & \text{for} \ \ \tilde{b}^{ij}_{klp} < 0, \tilde{\lambda}^{ij}_{kp} < |\tilde{b}^{ij}_{klp}| \\ 0 & \text{for} \ \ \tilde{\lambda}^{ij}_{kp} \geq |\tilde{b}^{ij}_{klp}|. \end{cases}$$

5. Set the B-matrix of iteration *iter* equal to values obtained in the last iteration, B_{iter-1}, that is $B_{iter} = B_{iter-1}$ for iteration *iter*.

Convergence is achieved when $max(|B_{iter} - B_{iter-1}|) < \epsilon$. The ϵ is chosen such that the LASSO solution converges to the OLS estimate for a penalty parameter set to zero and weighted sum of squared residuals as the loss function.

Appendix C. Penalty parameters

For the empirical applications a grid of twelve values is chosen for λ_k (minimum value of 0.01). The grid for α, c, and γ are: $a_{grid} = [0.2, 0.4, 0.6, 0.8]$, $c_{grid} = [1.2, 1.4, 1.6, 1.8]$, and $\gamma_{grid} = [0.2, 0.4, 0.6, 0.8]$. The grid for ρ is $\rho_{grid} = [0, 0.1, 0.2, 0.3, 0.4]$. Due to the high computational time, for the second application the model with shrinkage towards homogeneous coefficients uses smaller grids (for λ_k six values with minimum of 0.26 which is the smallest value which was chosen for a larger grid, $a_{grid} = [0.4, 0.6]$, $c_{grid} = [1.4, 1.6]$, and $\gamma_{grid} = [0.2, 0.4]$). Table 6 gives the selected penalty parameters for the empricial applications.

Table 6
Choice of penalty parameters.

	Small application			Large application		
	λ_k, c, α	λ_k, c, α $\Sigma = I$	$\lambda_k, c, \alpha, \gamma$	λ_k, c, α	λ_k, c, α $\Sigma = I$	$\lambda_k, c, \alpha, \gamma$
α	0.4	0.4	0.2	0.8	0.2	0.4
c	1.4	1.4	1.8	1.8	1.2	1.6
γ			0.2			0.2

NOTE: Estimators (in columns): LASSO with penalties λ_k, α, c and weighted sum of squared residuals, LASSO with the same penalty parameters where the covariance matrix is set to the identity matrix, LASSO with penalties $\lambda_k, \alpha, c, \gamma$ with aggregated estimator for homogeneous coefficients and with weighted sum of squared residuals.

Appendix D. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijforecast.2022.05.007.

Supplementary appendix: including details on the estimation of the covariance matrix, additional results for the simulation and empirical application, details on the Bayesian benchmark model used in the forecasting application and on the optimal penalty parameters.

Matlab and data (upon request): codes for replicating the simulation study and the forecasting exercise.

References

Alessi, L., & Banbura, M. (2009). *Comparing global macroeconomic forecasts: Technical report*, European Central Bank: Working paper.

Amisano, G., & Giacomini, R. (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics*, *25*, 177–190. http://dx.doi.org/10.1198/073500106000000332.

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

Banerjee, O., El Ghaoui, L., & d'Aspremont, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *Journal of Machine Learning Research*, *9*, 485–516.

Basu, S., & Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics*, *43*, 1535–1567. http://dx.doi.org/10.1214/15-AOS1315.

Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, *120*, 70–83. http://dx.doi.org/10.1016/j.csda.2017.11.003.

Bernanke, B. S., Boivin, J., & Eliasz, P. (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics*, *120*, 387–422.

Billio, M., Casarin, R., Ravazzolo, F., & van Dijk, H. K. (2016). Interconnections between eurozone and US booms and busts using a Bayesian panel Markov-switching VAR model. *Journal of Applied Econometrics*, *31*, 1352–1370. http://dx.doi.org/10.1002/jae.2501.

Billio, M., Casarin, R., & Rossini, L. (2019). Bayesian nonparametric sparse VAR models. *Journal of Econometrics*, *212*, 97–115. http://dx.doi.org/10.1016/j.jeconom.2019.04.

Bjørnland, H. C., Ravazzolo, F., & Thorsrud, L. A. (2017). Forecasting GDP with global components: This time is different. *International Journal of Forecasting*, *33*, 153–173.

Breitung, J. (2015). The analysis of macroeconomic panel data. In *The Oxford handbook of panel data*.

Canova, F., & Ciccarelli, M. (2004). Forecasting and turning point predictions in a Bayesian panel VAR model. *Journal of Econometrics*, *120*, 327–359. http://dx.doi.org/10.1016/S0304-4076(03)00216-1.

Canova, F., & Ciccarelli, M. (2009). Estimating multicountry VAR models. *International Economic Review*, *50*, 929–959. http://dx.doi.org/10.1111/j.1468-2354.2009.00554.x.

Canova, F., & Ciccarelli, M. (2013). *Panel vector autoregressive models: a survey: Working paper series 1507*, European Central Bank.

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

Carriero, A., Kapetanios, G., & Marcellino, M. (2009). Forecasting exchange rates with a large Bayesian VAR. *International Journal of Forecasting*, *25*, 400–417.

Ciccarelli, M., Ortega, E., & Valderrama, M. T. (2016). Commonalities and cross-country spillovers in macroeconomic-financial linkages. *B.E. Journal of Macroeconomics*, *16*, 231–275. http://dx.doi.org/10.1515/bejm-2013-0120.

Crespo Cuaresma, J., Feldkircher, M., & Huber, F. (2016). Forecasting with global vector autoregressive models: A Bayesian approach. *Journal of Applied Econometrics*, *31*, 1371–1391. http://dx.doi.org/10.1002/jae.2504.

Davis, R., Zang, P., & Zheng, T. (2016). Sparse vector autoregressive modeling. *Journal of Computational and Graphical Statistics*, *25*, 1–53. http://dx.doi.org/10.1080/10618600.2015.1092978.

Dees, S., di Mauro, F., Smith, L. V., & Pesaran, M. H. (2007). Exploring the international linkages of the euro area: A global VAR analysis. *Journal of Applied Econometrics*, *22*, 1–38.

Demirer, M., Diebold, F. X., Liu, L., & Yilmaz, K. (2018). Estimating global bank network connectedness. *Journal of Applied Econometrics*, *33*, 1–15.

Diebold, F. X., & Yilmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, *182*, 119–134. http://dx.doi.org/10.1016/j.jeconom.2014.04.

Diebold, F., & Yilmaz, K. (2015). Measuring the dynamics of global business cycle connectedness. In S. J. Koopman, & N. Shephard (Eds.), *Unobserved components and time series econometrics*, Oxford University Press, chapter 5.

Doan, T., Litterman, R., & Sims, C. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews*, *3*, 1–100. http://dx.doi.org/10.1080/07474938408800053.

Feldkircher, M., Huber, F., & Pfarrhofer, M. (2020). *Factor augmented vector autoregressions, panel VARs, and global VARs* (pp. 65–93). Cham: Springer International Publishing.

Friedman, J., Hastie, T., & Höfling, R. (2007). Pathwise coordinate optimization. *Annals of Applied Statistics*, *1*, 302–332. http://dx.doi.org/10.1214/07-AOAS131.

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, *9*, 432–441. http://dx.doi.org/10.1093/biostatistics/kxm045.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1–22.

Garcia, M. G., Medeiros, M. C., & Vasconcelos, G. F. (2017). Real-time inflation forecasting with high-dimensional models: The case of Brazil. *International Journal of Forecasting*, *33*, 679–693. http://dx.doi.org/10.1016/j.ijforecast.2017.

Giannone, D., Lenza, M., & Primiceri, G. E. (2021). Economic predictions with big data: The illusion of sparsity. *Econometrica*, *89*(5), 2409–2437. http://dx.doi.org/10.3982/ECTA17842.

Giannone, D., & Reichlin, L. (2009). Comments on forecasting economic and financial variables with global VARs. *International Journal of Forecasting*, *25*, 684–686.

Greenwood-Nimmo, M., & Nguyen, V. H. (2015). *Measuring the connectedness of the global economy: Melbourne institute working paper series wp2015n07*, Melbourne Institute of Applied Economic and Social Research, The University of Melbourne.

Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *volume 143 of monographs on statistics and applied probability, Statistical learning with sparsity. the lasso and generalizations*.

Hsu, N. J., Hung, H. L., & Chang, Y. M. (2008). Subset selection for vector autoregressive processes using lasso. *Computational Statistics & Data Analysis*, *52*, 3645–3657. http://dx.doi.org/10.1016/j.csda.2007.12.004.

Kilian, L., & Lütkepohl, H. (2017). *Structural vector autoregressive analysis* (pp. 535–574). Cambridge University Press., chapter 16. Structural VAR Analysis in a Data-Rich Environment.

Kock, A. B., & Callot, L. (2015). Oracle inequalities for high dimensional vector autoregressions. *Journal of Econometrics*, *186*, 325–344. http://dx.doi.org/10.1016/j.jeconom.2015.02.013.

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

Koop, G., & Korobilis, D. (2016). Model uncertainty in panel vector autoregressive models. *European Economic Review*, *81*, 115–131. http://dx.doi.org/10.1016/j.euroecorev.2015.09.006.

Koop, G., & Korobilis, D. (2018). *Forecasting with high-dimensional panel VARs*: *Essex finance centre working papers 21329*, University of Essex, Essex Business School.

Koop, G., & Korobilis, D. (2019). Forecasting with high-dimensional panel VARs. *Oxford Bulletin of Economics and Statistics*, *81*, 937–959.

Korobilis, D. (2016). Prior selection for panel vector autoregressions. *Computational Statistics & Data Analysis*, *101*, 110–120. http://dx.doi.org/10.1016/j.csda.2016.02.011.

Lee, W., & Liu, Y. (2012). Simultaneous multiple response regression and inverse covariance matrix estimation via penalized Gaussian maximum likelihood. *Journal of Multivariate Analysis*, *111*, 241–255. http://dx.doi.org/10.1016/j.jmva.2012.03.013.

Litterman, R. (1986). Forecasting with Bayesian vector autoregressions-five years of experience. *Journal of Business & Economic Statistics*, *4*, 25–38. http://dx.doi.org/10.1080/07350015.1986.10509491.

Louzis, D. (2015). Measuring spillover effects in euro area financial markets: a disaggregate approach. *Empirical Economics*, *49*, 1367–1400.

Marcellino, M., Stock, J. H., & Watson, M. W. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics*, *135*, 499–526. http://dx.doi.org/10.1016/j.jeconom.2005.07.020.

Messner, J. W., & Pinson, P. (2019). Online adaptive lasso estimation in vector autoregressive models for high dimensional wind power forecasting. *International Journal of Forecasting*, *35*, 1485–1498. http://dx.doi.org/10.1016/j.ijforecast.2018.

Ngueyep, R., & Serban, N. (2015). Large-vector autoregression for multilayer spatially correlated time series. *Technometrics*, *57*, 207–216. http://dx.doi.org/10.1080/00401706.2014.902775.

Nicholson, W. B., Matteson, D. S., & Bien, J. (2016). High dimensional forecasting via interpretable vector autoregression. ArXiv e-prints arXiv:1412.5250.

Nicholson, W. B., Matteson, D. S., & Bien, J. (2017). VARX-L: Structured regularization for large vector autoregressions with exogenous variables. *International Journal of Forecasting*, *33*, 627–651. http://dx.doi.org/10.1016/j.ijforecast.2017.01.003.

Nicholson, W. B., Wilms, I., Bien, J., & Matteson, D. S. (2020). High dimensional forecasting via interpretable vector autoregression. *Journal of Machine Learning Research*, *21*, 1–52, URL http://jmlr.org/papers/v21/19-777.html.

Pesaran, M. H., Schuermann, T., & Smith, L. V. (2009). Forecasting economic and financial variables with global VARs. *International Journal of Forecasting*, *25*, 642–675. http://dx.doi.org/10.1016/j.ijforecast.2009.08.007.

Pesaran, M., Schuermann, T., & Weiner, S. (2004). Modeling regional interdependencies using a global error-correcting macroeconometric model. *Journal of Business & Economic Statistics*, *22*, 129–162.

Ren, Y., & Zhang, X. (2010). Subset selection for vector autoregressive processes via adaptive lasso. *Statistics & Probability Letters*, *80*, 1705–1712. http://dx.doi.org/10.1016/j.spl.2010.07.013.

Smeekes, S., & Wijler, E. (2018). Macroeconomic forecasting using penalized regression methods. *International Journal of Forecasting*, *34*, 408–430. http://dx.doi.org/10.1016/j.ijforecast.2018.01.001.

Song, S., & Bickel, P. J. (2011). Large vector auto regressions. ArXiv e-prints http://arxiv.org/abs/1106.3915 [arXiv:1106.3915].

Stock, J. H., & Watson, M. W. (2012). Generalized shrinkage methods for forecasting using many predictors. *Journal of Business & Economic Statistics*, *30*, 481–493. http://dx.doi.org/10.1080/07350015.2012.715956.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, *58*(1), 267–288. http://dx.doi.org/10.1111/j.2517-6161.1996.tb02080.x.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., & Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *67*(1), 91–108. http://dx.doi.org/10.1111/j.1467-9868.2005.00490.x.

Wilms, I., Barbaglia, L., & Croux, C. (2018). Multiclass vector autoregressive models for multistore sales data. *Journal of the Royal Statistical Society. Series C. Applied Statistics*, *67*, 435–452. http://dx.doi.org/10.1111/rssc.12231.