Penalized Estimation of Panel Vector Autoregressive Models: A Lasso Approach

Annika Schnücker

Freie Universität Berlin and DIW Berlin Graduate Center, Mohrenstr. 58, 10117 Berlin, Germany

October 11, 2017

Draft. Please do not cite!

Abstract

This paper proposes a new least absolute shrinkage and selection operator (lasso) for estimating panel vector autoregressive (PVAR) models. By allowing for interdependencies and heterogeneities across cross-sectional units, typically the number of parameters of PVAR models is too large to estimate using ordinary least squares. The penalized regression this paper introduces ensures the feasibility of the estimation by specifying a shrinkage penalty that contains time series and cross section characteristics; thereby, accounting for the inherent panel structure within the data. Furthermore, using the weighted sum of squared residuals as the loss function enables the lasso for PVAR models to take into account correlations between cross-sectional units in the penalized regression. Given large and sparse models, simulation results point towards advantages of using lasso for PVARs over OLS, standard lasso techniques as well as Bayesian estimators in terms of mean squared errors and forecast accuracy. Empirical forecasting applications with up to ten countries and four variables support these findings.

Keywords: Model selection, Lasso, Large vector autoregressions, Penalized regression, Multi-country models, Shrinkage estimation
JEL: C32, C33, C52

Email address: aschnuecker@diw.de (Annika Schnücker)

1. Introduction

Growing international interlinkages in the financial and real sector are a defining feature of the global economy and have risen in importance over recent decades. This involves major economic policy implications as highlighted for example by numerous IMF reports and notes on spillovers. Theoretical papers demonstrate that ignoring international spillovers could lead to biased impulse response functions and to inaccurate forecasts. Georgiadis (2017) stresses that the accuracy of spillover effects increases significantly when they are estimated with multi-country models instead of bilateral vector autoregressive models. Furthermore, leaving out variables capturing international connections could lead to an omitted variable bias impacting structural analyses, as discussed by, for example, Lütkepohl (2014). In addition, Pesaran et al. (2009) point out that not accounting for linkages across countries can lead to less accurate forecasts of macroeconomic variables. Consequently, multi-country models with several variables, such as panel vector autoregressive (PVAR) models, are necessary to capture global spillovers in economic analyses.

The strength of PVARs is to account for interdependencies and heterogeneities across nations by jointly modeling multiple variables of several economies. PVARs enable the modeling of dynamic interdependencies by augmenting country specific models with lagged foreign variables. These models allow for static interdependencies measured by potential non-zero covariances between the error terms of different countries. Moreover, PVARs take cross-country heterogeneities into account by specifying country-specific coefficient matrices. However, estimating these models is challenging because a large number of parameters is usually set against a short time series. Due to the curse of dimensionality estimation of these models is thus often infeasible.

This paper proposes a new least absolute shrinkage and selection operator (lasso) that is suitable for estimating PVAR models. It provides a solution to the curse of dimensionality problem by using the panel structure to ensure the estimation feasibility. The paper establishes the asymptotic oracle properties of the lasso for PVARs. That means, asymptotically the lasso selects the true variables to be included in the model and estimates non-zero parameters as efficiently as if the true underlying model is known. The finite sample properties of the lasso for PVAR models are confirmed in simulation studies.

The proposed lasso estimator takes the panel structure inherent to the data into account and allows for an unrestricted covariance matrix at the same time. This is achieved by including penalty terms incorporating time series and cross-sectional properties. Furthermore, the modification of the penalty term is combined with specifying the loss function of the estimation problem as the weighted sum of squared residuals, thereby, accounting for the correlation between error terms of different cross section units.

In general, the lasso, as proposed by Tibshirani (1996), regulates the dimension of the model by constraining the estimation problem with a linear penalty term. The penalization determines the sum of the absolute values of the regression coefficients, that is, the L_1-norm of the coefficient matrix, to be less than a fixed value. Thus, the penalty term governs the degree of shrinkage. By forcing some coefficients to be zero and shrinking others, the lasso chooses the variables to be included in the model.

The main advantages of the lasso technique applied here are threefold.[1] Firstly, the

[1]Other methods to ensure the feasibility of the estimation are factor approaches, Bayesian shrinkage priors, selection priors, and classical shrinkage methods, such as the ridge regression. Some that are used

specified penalty parameters of the lasso for PVARs account for the inherent panel structure within the data. The penalty terms build on a specific expected structure in panel data models. That is, interdependencies are assumed to only exist between specific variables and cross sections combinations and decrease over lags. The lasso uses this structure to reduce the number of dimensions in the system by setting specific coefficients to zero. In particular, the penalty terms capture that more recent lags provide more information for the estimation than more distant ones and that lags of domestic variables are more important than lags of foreign variables. As demonstrated by Song and Bickel (2011), Nicholson et al. (2016), and Nicholson et al. (2017), including grouping structures or time series properties in the specification of the lasso for estimating VARs can improve forecast accuracy compared to the normal lasso penalty. The authors let the penalty term vary across lags and include grouping structures by using group lasso techniques as proposed by Yuan and Lin (2006). This allows them to capture similar sparsity patterns in the coefficient matrix. Likewise, contributions on Bayesian selection priors for PVARs support that accounting for the inherent panel dimension within the data can enhance forecasting performance.[2]

Secondly, considering an unrestricted covariance matrix by the specification of the loss function includes possible correlations between error terms in the estimation of the parameters. In penalized regressions, coefficients derived using generalized least squares deviate from those derived by ordinary least squares. Using the sum of squared residuals as the loss function disregards possible correlations between variables, thereby restricting the covariance matrix to the identity matrix. Hence, this procedure imposes strict assumptions on the dependence structure between the cross-sectional units. Lee and Liu (2012) show this for the use of lasso for VAR models. Basu and Michailidis (2015), Davis et al. (2015), and Ngueyep and Serban (2015) modify the loss functions in the lasso optimization for VAR models and allow for unrestricted covariances in the penalized estimation.

Thirdly, the lasso for PVARs benefits from the same properties as the lasso proposed by Tibshirani (1996). That is, the lasso reduces the dimension of the estimated model. Thereby, it ensures the feasibility of the estimation if the number of parameters per equation exceeds the number of observations. Furthermore, the lasso simultaneously selects and estimates the model. It allows for a flexible lag structure across equations since the lasso can choose different lag orders for each equation of the model. Moreover, the lasso is able to improve forecast prediction accuracy by reducing the variance of the predicted values.

The lasso for PVARs is of interest for empirical work since it provides a solution to ensure the estimation feasibility for PVAR models. That is relevant since, first, PVAR models are typically large including several countries and variables per country to capture macroeconomic relations. Second, the dimension of PVARs grows fast as adding a country increases the number of equations and columns of the coefficient matrices while adding variables means including them for each country. The lasso for PVARs can be used for estimating reduced form VARs. It can select the subset of variables that should be included in the model and serve as a flexible lag length selection tool. Due to the selection of the relevant variables the PVAR model estimated via lasso is easily interpretable and might be used for further structural analysis or forecasting.

for PVARs are described in section 2.5 in detail.

[2]See, for example, Koop and Korobilis (2015b), Korobilis (2016) and Schnücker (2016).

By introducing the lasso for PVARs, this paper contributes, firstly, to the literature on the use of the lasso techniques for VAR models and, secondly, to the literature on estimation procedures for PVAR models. Hsu et al. (2008) establish the usage of the lasso for VAR models. The authors, along with Kascha and Trenkler (2015), report that the lasso improves forecast performance compared to the use of information criteria for model selection. Ren and Zhang (2010) and Ren et al. (2013) build on Zou (2006), who propose adaptive weights for penalizing coefficients differently, and develop adaptive lasso techniques for VAR models. Their results provide evidence that the adaptive lasso outperforms the lasso in terms of forecasting performance, thus indicating the benefit of coefficient specific penalties. Kock and Callot (2015) establish non-asymptotic oracle inequalities for the lasso and adaptive lasso for high-dimensional VAR models. The authors further show that the lasso provides asymptotically consistent estimates and that the adaptive lasso is asymptotically equivalent to the least squares estimator that only includes true non-zero parameters.[3]

To date, two main extensions of Tibshirani's lasso are proposed in the context of VAR models. As mentioned, one strand of the literature broadens the specification of the penalty term to include special characteristics. The second group modifies the loss function in order to allow for unrestricted covariance matrices. However, the papers are either part of the first or the second group. One exception is Ngueyep and Serban (2015), who propose a penalized log-likelihood scheme applying penalties for higher lags and within group or between group penalties. Thus, the authors take into account the covariance matrix and allow for special characteristics. Yet, they still restrict the covariance matrix in their approach to a block structure by assuming no dependence across groups. This paper fills the gap by combining the weighted sum of squared residuals as the loss function with penalty terms that incorporate data properties.

Furthermore, the paper extends the current literature on the estimation of PVAR models. As yet, the literature mainly uses three kinds of model selection methods. Canova and Ciccarelli (2004) and Canova and Ciccarelli (2009) propose a Bayesian cross-sectional shrinkage approach factorizing the parameters into lower dimensional factors, thereby reducing the number of parameters to estimate. Canova et al. (2012), studying dynamics of the European business cycle, and Ciccarelli et al. (2016), analyzing spillovers in macro-financial linkages across developed economies, apply the cross-sectional shrinkage approach. Billio et al. (2014) extend the approach to a Markov-switching model. Koop and Korobilis (2015a) broaden it to time-varying parameter PVAR models additionally allowing for time-varying covariance matrices. An issue with this procedure is that the structural identification is more complex since the error term includes two components coming from the equation estimating the factorized parameters and from the estimation of the VAR model.

A second Bayesian approach is introduced by Koop and Korobilis (2015b), who suggest a selection prior for PVAR models called stochastic search specification selection.

[3]This paper focuses on the lasso estimated in a frequentist way and does not cover Bayesian lasso approaches. Bayesian lasso variants are, for example, discussed by Park and Casella (2008) and Kyung et al. (2010). Korobilis (2013), Gefang (2014), and Billio et al. (2016) use Bayesian lasso approaches for VAR models. Additionally, papers use the lasso for panel data regressions. Since this paper concentrates on the estimation of panel VAR models, these approaches are not further discussed. Other contributions include Ando and Bai (2016) and Su et al. (2016).

Based on a hierarchical prior, restrictions that specify no dynamic interdependencies, no static interdependencies, and homogeneity across cross-sectional units are searched. Schnücker (2016) develops the approach further by allowing for a more flexible panel structure. These papers provide evidence that accounting for the inherent panel structure in the data is beneficial in terms of improved forecast accuracy.

A third way is to assume no dependence or homogeneity across the panel units.[4] The assumptions must be based on a solid theoretical background. Estimation procedures for these kinds of models are described in Canova and Ciccarelli (2013) and Breitung (2015).

The results of three simulations and an empirical application support the use of the lasso for PVARs. It improves the forecast accuracy measured by mean squared forecast errors relative to estimating the PVAR with OLS, relative to Bayesian panel VAR methods, and relative to single country models. Accounting for the panel dimension in the penalty terms increases the forecast performance as using a lasso approach without such specific penalty terms leads to larger mean squared forecast errors. The gain in forecast accuracy relative to other estimation techniques is, in particular, found for large systems in simulations and an empirical application. For smaller models, the lasso for PVARs performs equally to the models of comparison.

Furthermore, the dimension reduction of the lasso techniques results in smaller mean squared errors for all simulations compared to OLS. The benefit in terms of lower mean squared error is higher for large and sparse models. The mean squared errors of the lasso techniques are in the same range with Bayesian panel VAR methods and single country models.

In the following, the lasso for PVAR models is introduced and its asymptotic properties are discussed. Other estimation strategies for PVARs are reviewed. Next, three simulation studies evaluate the performance of the lasso for PVARs along different criteria. A forecasting exercise is conducted in section four while section five concludes.

2. The lasso for PVARs

2.1. PVAR Model

Panel vector autoregressive models include several countries and country-specific variables in one model. A PVAR with N countries and G variables per country is given by

$$y_{it} = A_{i1}Y_{t-1} + A_{i2}Y_{t-2} + ... + A_{ip}Y_{t-p} + u_{it}, \tag{1}$$

where y_{it} denotes a vector of dimension $[G \times 1]$ for country i with $i = 1, ..., N$.[5] The $Y_{t-P} = (y'_{1t-P}, ..., y'_{Nt-P})'$ are of dimension $[NG \times 1]$ and the coefficient matrices A_{iP} of dimension $[G \times NG]$ for $P = 1, ..., p$. The $u_{it} \sim \mathcal{N}(0, \Sigma_{ii})$ and the covariance matrices across countries are given by Σ_{ij} for $i \neq j$.

In compact form, the PVAR model can be written as

$$Y_t = BX_{t-1} + U_t, \tag{2}$$

[4] Examples setting assumptions include Love and Zicchino (2006), Gnimassoun and Mignon (2016), and Attinasi and Metelli (2017), assuming homogeneity and no dynamic interdependencies, while Ciccarelli et al. (2013) or Comunale (2017) restrict for no dynamic interdependencies. Prez (2015) and Wieladek (2016) use a Bayesian approach and assume no dynamic interdependencies.

[5] Although this specification does not include a constant, it can be extended to include one.

where $Y_t = (y'_{1t}, ..., y'_{Nt})'$ and the coefficient matrix B is of dimension $[NG \times NGp]$. The matrix X_{t-1} includes all lagged variables, $X_{t-1} = (Y_{t-1}, ..., Y_{t-p})'$, and is of dimension $[NGp \times 1]$. The U_t is normally distributed with 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 countries. The X_{t-1} includes lagged values of every variable in each equation. The unrestricted B-matrix and the covariance matrix Σ enable country specific coefficients and correlations between error terms of all possible variable-country combinations. This PVAR model has $(NG)^2 p$ parameters of the B-matrix and $\frac{NG(NG+1)}{2}$ parameters of Σ to estimate. Variables are ordered per country meaning that the first G rows of the system model variables of country one, while the rows $NG - G + 1$ to NG describe the variables of country N. The large number of parameters can lead to the curse of dimensionality problem. The lasso provides a solution to deal with this issue.

2.2. The lasso Estimator

Tibshirani (1996) proposed the lasso for a linear regression model with multiple regressors. The coefficient estimates are obtained by minimizing the sum of squared residuals subject to a linear constraint. The penalization term regulates the sum of the absolute values of the regression coefficients, the L_1-norm of the coefficients, to be less than a fixed value. The lasso forces the coefficients to lie in a specific area that is centered around zero. Thereby, it shrinks some coefficient and constrains other to be equal to zero. The L_1-norm determines the geometric shape of this constraint region. It has two properties that are crucial for the features of the lasso. Coefficients can equal zero due to the possibility of corner solutions and, secondly, the constraint region is convex, which simplifies the optimization procedure.

Introducing a shrinkage penalty in the regression enables coping with situations in which $T < NGp$, can improve prediction accuracy, and produce interpretable models.[6] If $T < NGp$, the number of parameters per equation exceeds the number of observations, ordinary least squares is not feasible since no unique solution exists. If the true model is sparse, meaning that some of the true coefficients are zero, the lasso finds a solution by constraining the estimation. Furthermore, the lasso reduces the variance of the estimated coefficients, thereby improving prediction accuracy. Due to the selection property of the lasso the interpretation of the model is enhanced. By setting some coefficients to zero, a subset of variables that simplifies the identification of core driving variables of the system is selected. The three mentioned properties clarify for which situations the lasso is well suited, namely for large, sparse systems for which the researcher's aim is to provide forecasts and to analyze main driving forces. The bias introduced by the lasso is accepted in order to gain these properties.

2.3. Extended Penalty Term and Loss Function for PVARs

The optimization problem of the lasso for PVAR models modifies the lasso of Tibshirani (1996) in two ways. The weighted sum of squared residuals is used as the loss function instead of the sum of squared residuals. Furthermore, a penalty term capturing

[6] Tibshirani (1996) and Hastie et al. (2015) discuss these three properties in detail.

the time series and cross section properties is introduced. The resulting optimization problem is given by:

$$\begin{aligned}\underset{b_{km}}{\operatorname{argmin}}\ & \frac{1}{T}\sum_{k=1}^{K}\sum_{j=1}^{K}\omega_{kj}\left(Y_k-\sum_{m=1}^{Kp}b_{km}X_m\right)\left(Y_j-\sum_{m=1}^{Kp}b_{jm}X_m\right)' \\ & +\sum_{k=1}^{K}\sum_{m=1}^{Kp}\lambda_{km}\left|b_{km}\right|,\end{aligned}\tag{3}$$

where b_{km} is the element of the B-matrix in the k-th row and m-th column. K is the number of countries times the number of variables, $K = NG$. The Y_j and X_m are of dimension $[1 \times T]$ for $j = 1, ..., K$ and $m = 1, ..., Kp$. The ω_{kj} is an element of the inverse of the covariance matrix, $\Sigma^{-1} = \Omega$. The λ_{km} is the penalty parameter and $|b_{km}|$ denotes the absolute value of b_{km}. As common in the lasso literature, Y is demeaned and standardized. The latter is done in order to have comparable units for all variables when choosing the penalty parameters.

The optimization problem is solved using a coordinate descent algorithm as proposed in Friedman et al. (2007) and Friedman et al. (2010).[7] This iterative algorithm updates from B_n to B_{n+1} by a univariate minimization over a single b_{km}. It iterates over all elements in B till convergence is reached.[8] 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 existence of a global solution. The lasso estimator, which is called *lassoPVAR* in the following, has the form:

$$b_{km}^{lasso} = sign\left(\tilde{b}_{km}\right)\left(\left|\tilde{b}_{km}\right| - \frac{\lambda_{km}T}{2\omega_{kk}X_mX_m'}\right)\tag{4}$$

with

$$\tilde{b}_{km} = \frac{\sum_{j\neq k}^{K}\omega_{jk}(Y_j-\sum_{i=1}^{Kp}b_{ji}X_i)X_m'}{\omega_{kk}X_mX_m'} + \frac{(Y_k-\sum_{i\neq m}^{Kp}b_{ki}X_i)X_m'}{X_mX_m'}.\tag{5}$$

As pointed out by Lee and Liu (2012), in a VAR model correlations between error terms have an impact on the estimated parameters in a restricted regression.[9] It can be easily seen from the above stated lasso estimator b_{km}^{lasso} that the covariance affects the value of b_{km}^{lasso} for elements $\omega_{kk} \neq 1$ and $\omega_{jk} \neq 0$ for $j \neq k$. When Σ equals the identity matrix, the estimator b_{km}^{lasso} reduces to the lasso estimator based on the sum of squared residuals as the loss function.

As yet, the literature on estimating VAR models with the lasso follows two main approaches to estimate the covariance matrix: a two-step approach or a joint likelihood approach. Lee and Liu (2012) describe two plug-in methods, where in a first step either the

[7] The optimization algorithm and the derivation of the lasso estimator are described in detail in Appendix C and Appendix A. For more details regarding the optimization algorithm see Friedman et al. (2007), Friedman et al. (2010) and Hastie et al. (2015).

[8] Convergence is achieved when $max(|B_n - B_{n-1}|) < \epsilon$. 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.

[9] See Lee and Liu (2012) for details. This is similar to the well-known fact that for VAR models, OLS is unequal to GLS in the case of parameter constraints.

coefficient matrix or the covariance is estimated, followed by the estimation of the other. The authors use a graphical lasso (glasso), following, in particular, Friedman et al. (2008). In addition, they present a doubly penalized likelihood approach to jointly estimate the coefficient and covariance matrix in a L1-regularized regression. Basu and Michailidis (2015) propose another option by estimating the covariance matrix using residuals of an initial lasso estimation with sum of squared residuals or a glasso approach. Further, they present a joint penalized maximum likelihood approach. Davis et al. (2015) compare their two-stage approach using tools from the frequency domain with a lasso approach weighted with the inverse covariance matrix. Updating until convergence, the covariance matrix is estimated using the residuals of the lasso estimation. Ngueyep and Serban (2015) propose a penalized log-likelihood scheme applying penalties for higher lags and within group or between group penalties.

In this paper, the covariance matrix Σ is estimated using a two-step approach. The first step estimates the covariance matrix via glasso, while in the second step the estimated $\hat{\Sigma}$ is plugged into the lasso estimation of b_{km}^{lasso}. Friedman et al. (2008) demonstrate that the covariance matrix is estimated by maximizing the Gaussian penalized log-likelihood

$$log\ det(\Omega) - tr(S\,\Omega) - \rho\,||\Omega|| \tag{6}$$

with respect to the nonnegative 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. 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 Appendix B. As pointed out by Banerjee et al. (2008) $\hat{\Sigma}$ is even in the case when the number of variables is larger than the number of observations invertible due to the regularization using $\rho > 0$.

An alternative way to estimate the covariance matrix, as done by, for example, Tibshirani (1996), is to use 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. The degrees of freedom adjusted least square estimator is a consistent estimator for constrained regression problems, although zero restrictions can reduce the number of degrees of freedoms. Another option is to use the number of degrees of freedom for the lasso, which is the number of non-zero parameters.[10] However, in contrast to the glasso estimation, this approach can lead to problems for the invertability of the covariance matrix in large systems. This is why the glasso approach is used here.

The second extension of the lasso for PVARs is the modification of the penalty term. The λ_{km} denotes the penalty parameter. If $\lambda_{km} = 0$, the estimated coefficients equal the OLS solutions. If $\lambda_{km} > 0$, the parameters are shrunk toward zero. To allow for a specific time series and cross section penalty, λ_{km} consists of three parts:

$$\lambda_{km} = \lambda_k\ p^{\alpha}\ c. \tag{7}$$

1. **Basic penalty** - λ_k. This part varies across equations. $\lambda_k > 0$ will force coefficients toward zero.
2. **Time series penalty** - p^{α}. It captures that more recent lags provide more information than more distant ones. The penalty increases with the lag order, p, for $\alpha > 0$.

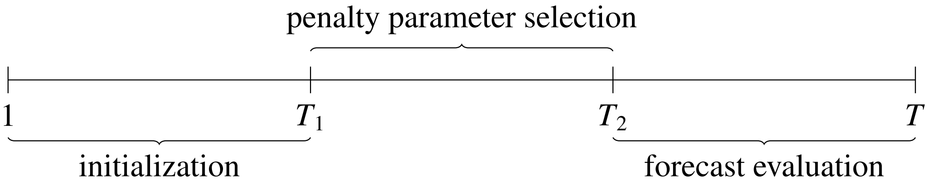
[10]Regarding the degrees of freedom for the lasso see Bühlmann and van de Geer (2011) for details.

The time series penalty part allows the penalty to vary across lagged variables.

3. **Cross section penalty** - $c > 1$, if foreign variable. The penalty models that lags of domestic variables have a larger impact than lags of foreign variables.

The penalty parameters vary across equations (due to λ_k) and across lagged variables (due to p^α and c). The parameters α and c are fixed for the whole model.

Optimal penalty parameters are determined via a rolling cross-validation technique. The penalty parameters are chosen such that they minimize one-step ahead mean squared forecast errors.[11]



Like Song and Bickel (2011), the sample is split 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.[12] The model is estimated in a rolling scheme taking the observations from t to $T_1 + t - 1$ for $t = 1, ..., T_2 - T_1$. For each t the out-of-sample forecast accuracy for a specific penalty parameter λ_{km} is measured by the one-step ahead mean squared forecast error for variable k, $k = 1, ..., NG$:

$$MSFE(\lambda_{km})_k = \frac{1}{T_2 - T_1} \sum_{t=1}^{T_2-T_1} (\hat{Y}_{k,t+1} - Y_{k,t+1})^2$$

where $\hat{Y}_{k,t+1}$ denotes the estimated one-step ahead forecast for variable k. For simplicity only λ_k is determined via cross-validation, while α and c are pre-set to $\alpha = 0.6$ and $c = 1.4$ for the simulation and $\alpha = 0.6$ and $c = 1.8$ for the application. These values are preselected in a small cross-validation exercise. 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.[13] The forecast performance is evaluated for the period T_2 to T by MSFEs based on rolling window forecasts with the fixed penalty parameters determined for the period 1 to $T_2 - 1$.

The application of the lasso for PVAR is not limited to the currently considered panel VAR model in which the cross sections are countries. Other possible cross-sectional

[11]The n-fold cross-validation technique for choosing the optimal penalty parameter is not applied here due to the time dependence in the PVAR model. By choosing the optimal penalty parameter that minimizes one-step ahead mean squared forecast errors, this paper follows Song and Bickel (2011), Nicholson et al. (2016), and Nicholson et al. (2017). However, in contrast, Bergmeir et al. (2015) justify the use of the standard n-fold cross-validation techniques for autoregressive processes.

[12]For estimation, for the simulation the periods are $T_2 = T - 20$ and $T_1 = T_2 - 20$ and for the application $T_2 = T - 20$ and $T_1 = T_2 - 60$. Extending the period for penalty parameter selection comes at the cost of longer computational time.

[13]For the simulations: $\lambda_k^{max} = max(max(X * Y'))$ and λ_k^{grid} are six values between 0.01 and $(1/NGp)\lambda_k^{max}$. For the applications: $\lambda_k^{max} = max(max(X * Y'))$ and λ_k^{grid} are twelve values between 0.01 and $(1/T)\lambda_k^{max}$. See Appendix F.1 for details on the grid values for the application.

dimensions are, for example, industries and regions. More generally, the cross section penalty can be understood as a higher penalty for variables of a cross section unit different than the one of the variable being explained.

2.4. Asymptotic Properties

As a variable selection method, the lasso for PVARs should satisfy the oracle properties.[14] This means, firstly, asymptotically the lasso selects the correct sparsity pattern. With probability tending to one, it sets true zero parameters to zero while setting non-zero parameters unequal zero. Secondly, the non-zero parameters are as efficiently estimated as if the true subset of relevant variables is known. Thus, for the oracle properties to hold, selection consistency and asymptotic normality has to be satisfied as T goes to infinity.[15]

The asymptotic analysis follows the steps established in Song and Bickel (2011) and Lee and Liu (2012). Assume that the data Y_t are generated from an underlying model as in equation (1) where $U_t \sim \mathcal{N}(0, \Sigma)$. The PVAR model is stable. That is, all roots of $det(I_K - A_1 z - A_2 z^2 - ... - A_p z^p)$ are outside the unit circle.

Define the true parameter matrix as B^*. Assume that the covariance matrix is known. The inverse of the covariance matrix is denoted as Ω. If Ω is estimated consistently, it can be easily shown that the results derived in the following hold. The true coefficient in the k-th row and m-th column of B^* is defined as b^*_{km}. The vectorized true coefficient matrix is given by $b^* = vec(B^*)$. Let $J = \{(k, m) : b^*_{km} \neq 0\}$ denote the set of non-zero parameters. The number of non-zero parameters, the cardinality of J, is given by $|J| = s$. The lasso estimator of b^*, as derived from the optimization problem in equation (3) under the $[1 \times NG^2 p]$-vector of penalty parameters, λ, is denoted as $\hat{b}$. The b^*_J is the vector of true non-zero parameters with dimension $[s \times 1]$ and $\hat{b}_J$ is the estimator of b^*_J. Let $Z = I_K \otimes X'$ where X is the $[Kp \times T]$-matrix of right hand side lagged variables.

Define $a_T = \lambda_{km}$ for $k, m \in J$ and $c_T = \lambda_{km}$ for $k, m \notin J$. Assume that the lag length p can increase with growing T. Thus, λ_{km} is time dependent since it depends on p. The a_T is defined as the penalty term λ_{km} for a true non-zero parameter. The c_T gives the penalty term for true zero parameters. The specified penalty terms in *lassoPVAR* allow for different penalization of each variable. The introduction of time series and cross section penalty terms leads to stronger penalization of close to zero coefficients. Thus, the distinction of the penalty term in a_T and c_T is justifiable. Furthermore, the following assumptions are made:

(A1) $\Gamma := plim\, ZZ'/T$ exists and is nonsingular.
(A2) Non-zero parameters exist. The cardinality of J is unequal zero, $|J| = s > 0$.
(A3) Assume that $a_T \sqrt{T} \to 0$.
(A4) Assume that $c_T \sqrt{T} \to \infty$.

Thus, assumptions (A3) and (A4) require different rate of convergences properties for the penalty parameters associated with true zero and true non-zero coefficients.

[14] For the definition of the oracle property, see, for example, Lee and Liu (2012) and Kock and Callot (2015).

[15] An increasing number of cross sections N increases the number of free parameters by adding equations and variables in each existing equation and not the number of observations.

Theorem 1 *Under the assumptions (A1) to (A4) the following results hold:*[16]

(R1) ***Selection consistency**: plim* $\hat{b}_{km} = 0$ *if* $b^*_{km} = 0$.

(R2) ***Asymptotic normality**:* $\sqrt{T}(\hat{b}_J - b^*_J) \xrightarrow{d} \mathcal{N}(0, (\Omega \otimes \Gamma)^{-1}_J)$.

The $(\Omega \otimes \Gamma)_J$ is the covariance matrix obtained by removing the row and column of $\Omega \otimes \Gamma$ corresponding to the elements $(k, m) \notin J$. Results (R1) and (R2) imply that if the penalty parameters satisfy the conditions given in (A3) and (A4), then *lassoPVAR* satisfies asymptotically the oracle properties. Theorem (R1) states the selection consistency. That is, for $T \rightarrow \infty$, a true zero parameter, $b^*_{km} \notin J$, is estimated consistently, meaning that, equaling zero. The second result, (R2), establishes the asymptotic normality for true non-zero parameters, $b^*_{km} \in J$.

2.5. Comparison to Other Estimation Procedures for PVARs

This section describes three further lasso specifications, as well as the alternative existing estimation procedures for PVAR models, in the literature to which the performance of the lasso for PVAR is compared. These are restricted least squares, the selection prior of Koop and Korobilis (2015b), and the cross-sectional shrinkage approach of Canova and Ciccarelli (2009).[17] As a general benchmark model, the PVAR is estimated with ordinary least squared - this model is referred to as *OLS*. However, while it can serve as a benchmark for small models, *OLS* is unfeasible for larger models for which $T < Kp$.

Lasso with basic penalty. The first alternative lasso approach is a lasso with weighted sum of squared residuals as the loss function but without a penalty which explicitly captures panel properties. The time series penalty, α, is set to zero and the cross section penalty, c, equals one. Thus, the penalty parameter λ_{km} reduces to λ_k. In the following, this estimator will be referred to as *lassoVAR*.

Post lasso. Secondly, a post lasso is considered. The post lasso consists of two estimation steps, as explained by Belloni and Chernozhukov (2013). In the first step, a lasso optimization problem is solved based on the proposed specification with weighted sum of squared residuals along with time series and cross section penalties. In the second step, the non-zero elements of the first step are re-estimated with OLS. Thus, the post lasso reduces the bias of the non-zero elements introduced via lasso shrinkage. This estimator is called *post lassoPVAR*.

Adaptive lasso. The third lasso alternative is the adaptive lasso, as proposed by Ren and Zhang (2010) for VAR models following the idea of Zou (2006). While the lasso shrinks all coefficients constantly depending on the penalty parameter, the adaptive lasso penalizes high non-zero coefficient less than very small coefficients. This is achieved by adaptive weights. Zou (2006) proposes weights, which are data-dependent, for the penalty parameter, $\hat{w}_{km} = \frac{1}{|b^{OLS}_{km}|^\gamma}$, where b^{OLS}_{km} is the OLS estimate and γ a constant. OLS estimates close to zero will increase the penalty parameter, leading to increased shrinkage, while high non-zero coefficients will decrease the penalty parameter. The adaptive lasso applied here, referred to as *adaptive lassoVAR*, uses the weighted sum of squared residuals and

[16]The proof of the theorem is provided in Appendix D.

[17]The two Bayesian approaches are only briefly described in this paper. See Koop and Korobilis (2015b), Canova and Ciccarelli (2004), Canova and Ciccarelli (2009), and Canova and Ciccarelli (2013) for details.

sets $\alpha = 0$ and $c = 1$. One issue of the adaptive lasso is the choice of the unbiased estimator for the weights. For very large models, OLS is not feasible if $T < Kp$. An alternative is to use ridge estimates or post lasso estimates as weights.[18]

The *lassoPVAR* allows for different penalty parameters for different coefficients. The specification of the time series and cross section penalties captures close to zero coefficients and penalizes those stronger. Consequently, *lassoPVAR* can be seen as an adaptive lasso.

Restricted least squares. The first estimation approach for PVARs used in the literature, which is discussed here, is a restricted least squares estimation, called *restLS*. Hereby, the restricted LS estimates a block-diagonal system ordering the variables in country blocks. Such a model assumes no dynamic interdependencies between countries. Setting the off-diagonal elements to zero reduces the number of free parameters. However, the assumption of no dynamic interdependencies between various economies is theoretically hard to justify. No restrictions are set on the covariance matrix.

Single-country VAR. This model is closely related to the least squares approach but assumes both a block-diagonal coefficient matrix and a block-diagonal covariance matrix. Hence, the model allows for no interdependencies across countries. Estimating the whole system is equal to an estimation of each single country VAR separately. The model is estimated with OLS. The estimator is called *single VAR*.

Stochastic search specification selection. The second approach is the Bayesian selection prior of Koop and Korobilis (2015b) called stochastic search specification selection (*SSSS*). The authors define weighted normal distributions as prior distributions that center around a restriction with a small or a large variance. Thus, the first part of the distribution shrinks the estimated parameter toward the restriction (small variance) while the second part allows for a more freely estimated parameter (large variance). Depending on a hyperparameter, which is Bernoulli distributed, a parameter is drawn from the first or second part of the distribution. Koop and Korobilis (2015b) specify three different priors based on the possible restrictions: They search for no dynamic interdependencies, no static interdependencies and for homogeneity across coefficient matrices.

The prior centering around the no dynamic interdependency restriction is specified for an off-block-diagonal matrix of B of variables belonging to one country. The dynamic interdependency prior has the following form:

$$\begin{aligned}
B_{ij} &\sim (1-\gamma_{ij}^{DI})\mathcal{N}(0,\tau_1^2 I) + \gamma_{ij}^{DI}\mathcal{N}(0,\tau_2^2 I)\\
\gamma_{ij}^{DI} &\sim Bernoulli(\pi^{DI}), \quad \forall j \neq i
\end{aligned}$$

where B_{ij} is a off-block-diagonal matrix of B and $\tau_1^2 < \tau_2^2$. If $\gamma_{ij}^{DI} = 0$, B_{ij} is shrunk to zero, if $\gamma_{ij}^{DI} = 1$, B_{ij} is more freely estimated. Setting the prior on a block of variables of one country leads to a similar treatment of all variables of one country being either restricted (shrunk to zero) or not. The cross-sectional homogeneity prior is set on the

[18]Compare with, for example, Kock and Callot (2015). However, using the post lasso will increase computational time while using ridge estimation requires further determination of hyperparameters.

diagonal coefficient matrices of the B matrix. The prior has the following form:

$$B_{ii} \sim (1 - \gamma_{ij}^{CSH})\mathcal{N}(B_{jj}, \eta_1^2 I) + \gamma_{ij}^{CSH}\mathcal{N}(B_{jj}, \eta_2^2 I)$$
$$\gamma_{ij}^{CSH} \sim Bernoulli(\pi^{CSH}), \quad \forall j \neq i$$

where B_{ii} and B_{jj} are block-diagonal matrices of B and $\eta_1^2 < \eta_2^2$. If $\gamma_{ij}^{CSH} = 0$, B_{ii} is shrunk to B_{jj}. Koop and Korobilis (2015b) specify a hierarchical normal mixture prior for the off-diagonal elements of the covariance matrix to build in no static interdependencies. Since no restrictions are set on the covariance matrix for the lasso solution and the forecast comparison is done on the reduced form, no restriction search for static interdependencies is done in the following exercises. The covariance is drawn from an inverse Wishart distribution. A Markov Chain Monte Carlo algorithm samples the estimated parameters as the posterior means.

Cross-sectional shrinkage approach. The third estimation procedure is the cross-sectional shrinkage approach proposed by Canova and Ciccarelli (2009). Here, the parameters are factorized into common, country specific, and variable specific time-varying factors. Canova and Ciccarelli (2009) specify the model in a hierarchical structure:

$$\begin{aligned}
b &= \Lambda F + e_t \\
Y_t &= Z_t \Lambda F + \epsilon_t \\
\epsilon_t &= U_t + Z_t e_t \\
e_t &\sim \mathcal{N}(0, \Sigma \otimes \sigma^2 I) \\
\epsilon_t &\sim \mathcal{N}(0, (I + \sigma^2 Z_t' Z_t)\Sigma)
\end{aligned}$$

where Λ is a $[NG^2p \times f]$ matrix of loadings, F is an $[f \times 1]$ vector of factors, and $Z_t = I \otimes X_{t-1}$. Since the factors, F, are of a lower dimension than the vectorized B matrix, b, $f \ll NG^2p$ holds. The specified prior distributions for the covariance matrices are inverse Wishart and $b \sim \mathcal{N}(\Lambda F, \Sigma \otimes \sigma^2 I)$. The number of factors are N common factors for coefficients of each country, G common factors for coefficients of each variable, and one common factor for all coefficients.

An advantage of the approach is that it takes into account time variation. As one limitation, the cross-sectional shrinkage approach groups coefficients due to factorizing, however, it does not consider zero values in a specific way.[19] The procedure does not use possible sparsity for dimension reduction.

3. Simulation Studies

3.1. Simulation Set-Ups

The finite sample performance of the *lassoPAVR* is evaluated based on three Monte Carlo simulations. In the first simulation set-up data is generated from a stationary PVAR(1) model that includes two countries and two variables per country. The number of time series observations is 100. The underlying PVAR model has the parameter

[19] Korobilis (2016) elaborate further on this point.

matrix

$$A_1^{true} = \begin{bmatrix} 0.9 & 0.8 & 0 & 0 \\ 0 & 0.9 & 0 & 0 \\ 0.6 & 0.6 & 0.9 & 0 \\ 0.6 & 0.6 & 0.8 & 0.9 \end{bmatrix}$$

and normally distributed error terms, $U_t \sim \mathcal{N}(0, \Sigma^{true})$ with Σ^{true} : $\sigma_{ii} = 0.2$ and $\sigma_{ij} = 0.1$ for $i \neq j$. The PVAR model represents a scenario where the second country has no impact on variables of the first country while the first country's variables affect the variables of country 2. This set-up could be a model including one big and one small economy, justifying the block of zeros in the upper part of the A_1-matrix. A second property of the model is that domestic variables have a greater impact than foreign variables have.

The number of parameters of this model is moderate. The coefficient matrix has 16 free coefficients out of which 10 are true non-zero coefficients. As a result, the methods aiming for dimension reduction, such as the lasso approaches and the two Bayesian procedures, are not able to provide substantial benefit by reducing the number of parameters to estimate. Rather, the simulation is conducted to analyze whether these methods perform comparable to standard OLS in terms of mean squared error and forecast accuracy.

In the second simulation data is generated from a stationary PVAR(4) with $U_t \sim \mathcal{N}(0, \Sigma^{true})$, Σ^{true} : $\sigma_{ii} = 0.2$, $\sigma_{ij} = 0.1$ for $i \neq j$, and $T = 100$. The model includes three countries and two variables per country. The set-up illustrates a larger and sparse model with parameter matrices

$$A_1^{true} = \begin{bmatrix} 0.6 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0.6 & 0 & 0 & 0 & 0 \\ 0.4 & 0.4 & 0.6 & 0.5 & 0 & 0.4 \\ 0 & 0.4 & 0 & 0.6 & 0 & 0 \\ 0.4 & 0.4 & 0.4 & 0 & 0.6 & 0 \\ 0 & 0.4 & 0 & 0.4 & 0 & 0.6 \end{bmatrix},$$

$$A_2^{true} = \mathbf{0}, \quad A_3^{true} = \mathbf{0},$$

$$A_4^{true} = \begin{bmatrix} 0.35 & 0.3 & 0 & 0 & 0 & 0 \\ 0 & 0.35 & 0 & 0 & 0 & 0 \\ 0.3 & 0.3 & 0.35 & 0.3 & 0 & 0.3 \\ 0 & 0.3 & 0 & 0.35 & 0 & 0 \\ 0.3 & 0.3 & 0.3 & 0 & 0.35 & 0 \\ 0 & 0.3 & 0 & 0.3 & 0 & 0.35 \end{bmatrix}.$$

The model includes dynamic and static interdependencies as well as cross-sectional heterogeneities. It incorporates a time series pattern by lower coefficients for higher lags. Thus, the impact of a variable is smaller for lag four than for lag one. The second and third lag have no impact. This structure could be motivated by a model using quarterly data depicting seasonal patterns. In addition, foreign variables affect domestic variables less compared to the effect of domestic variables.

The second simulation provides a larger and sparser model than the model in simulation one. The coefficient matrices have 144 free parameters, out of which 34 are true non-zero coefficients, hence 23.61% of all coefficients of B are true non-zero coefficients.

Table 1: Summary of simulation set-ups

Simulation	**(1)**	**(2)**	**(3)**
lag length	1	4	4
number of countries	$N = 2$	$N = 3$	$N = 4$
number of variables	$G = 2$	$G = 2$	$G = 4$
number of non-zeros in B	10	34	432
number of elements in B	16	144	1024
fraction of non-zeros in B	62.50%	23.61%	42.19%

However, this model is still rather of medium size. The simulation enables analyzing whether efficiency gains compared to the benchmark OLS can already be found in medium sized models.

The DGP of the third simulation is based on a PVAR(4) with four countries and four variables per country. The U_t are normally distributed with $U_t \sim \mathcal{N}(0, \Sigma^{true})$, Σ^{true} : $\sigma_{ii} = 0.2$, $\sigma_{ij} = 0.1$ for $i \neq j$, and the length of the time series is $T = 100$. The coefficient matrices for lag $p = 1, ..., 4$ are lower triangular matrices where the diagonal elements are given by

$$(-0.8)^{(p-1)}0.8.$$

A column of the off-diagonal elements below the diagonal is given by

$$[0.5 - (p-1) \quad 0.5 - (p-1) \quad 0.5 - (p-1) \quad 0]'$$

repeated for each country. The coefficient matrices model that foreign lags are less important and that with increasing lag length the impact of the variables decreases. This large and sparse model allows for dynamic and static interdependencies as well as for heterogeneous coefficients across economies. In total, B has 1024 coefficients, of which 432 are true non-zero coefficients, thus 42.19% are non-zero coefficients. The constant is set to zero in all three simulations. Table 1 summarizes the simulation set-ups. The underlying models of simulation one and two are chosen to be all relatively small so that they allow the comparison to Bayesian PVAR methods and least squares estimators. For simulation three some estimators are not feasible.

3.2. Performance Criteria

The performance of the lasso for PVAR models is evaluated along the following criteria.[20]

1. **Correct sparsity pattern**: The measure calculates how often the evaluated procedure takes the correct decision whether to include or exclude a variable. It measures how often are true relevant variables included and true irrelevant discarded averaged over all Monte Carlo replications.
2. **Fraction of relevant variables included**: It counts the number of true relevant variables included in the models relative to the number of all true non-zero coefficients averaged over all Monte Carlo replications.

[20]Tibshirani (1996), Ren and Zhang (2010) or Kock and Callot (2015), for example, use similar criteria to assess the performance of the lasso.

Table 2: Overview of estimators

lassoPVAR	lasso for PVAR with weighted sum of squared residuals, time series and cross section penalties, $\lambda_{km} = \lambda_k p^{\alpha} c$
lassoVAR	lasso for PVAR with weighted sum of squared residuals, $\lambda_{km} = \lambda_k$, $\alpha = 0$ and $c = 1$
post lassoPVAR	post lasso for PVAR: first step estimates lasso for PVAR with weighted sum of squared residuals, time series and cross section penalties, $\lambda_{km} = \lambda_k p^{\alpha} c$ second step re-estimates non-zero elements with OLS
adaptive lassoVAR	adaptive lasso for PVAR with weighted sum of squared residuals, weights depend on OLS estimate, $\lambda_{km} = \lambda_k$, $\alpha = 0$ and $c = 1$
SSSS	selection prior of Koop and Korobilis (2015b)
CC	cross-sectional shrinkage approach of Canova and Ciccarelli (2009)
OLS	ordinary least squares estimation of PVAR model
restLS	restricted least squares, block diagonal system assumption of no dynamic interdependencies
single VAR	least squares, block diagonal system for coefficient matrix and covariance assumption of no dynamic and static interdependencies

3. **Number of variables included**: Reports the average number of variables included in the model. This measure evaluates the dimension reduction done by the estimator.
4. **MSE**: The mean squared error of the parameter estimates for one Monte Carlo replication is calculated as

$$MSE = \frac{1}{K^2 p} \sum_{k=1}^{K} \sum_{m=1}^{Kp} (\hat{b}_{km} - b_{km}^{true})^2$$

where $\hat{b}_{km}$ is the estimate of the true parameter b_{km}^{true}. The MSEs are averaged over all Monte Carlo replications.
5. **MSFE**: The h-step ahead mean squared forecast error for one Monte Carlo replication is calculated as

$$MSFE = \frac{1}{T - h - T_2 - 1} \sum_{t=T_2}^{T-h_{max}} \left[\frac{1}{K} \sum_{j=1}^{K} (\hat{Y}_{j,t+h} - Y_{j,t+h})^2 \right]$$

where $\hat{Y}_{j,t+h} = \hat{B}\hat{X}_{t+h_{max}-1}$ denotes the iteratively estimated h-step ahead forecast for

Table 3: Performance evaluation of estimators

	lasso techniques				Bayesian methods		least squares		
	lasso PVAR	*lasso VAR*	*post lasso*	*adaptive lasso*	*SSSS*	*CC*	*restLS*	*single VAR*	*OLS*
Correct sparsity pattern in %									
(1)	55.13	54.69	55.13	53.06	-	-	37.50	37.50	37.50
(2)	40.83	54.54	40.83	51.43	-	-	34.72	34.72	76.39
(3)	47.91	51.96	47.91	51.52	-	-	39.06	39.06	57.81
Fraction of relevant variables included in %									
(1)	31.90	34.40	31.90	38.50	-	-	60.00	60.00	100.00
(2)	38.41	57.26	38.41	51.29	-	-	47.06	47.06	100.00
(3)	44.18	62.85	44.18	59.98	-	-	33.33	33.33	100.00
Number of variables included									
(1)	5.20	5.63	5.20	6.19	16	16	8	8	16
(2)	50.91	83.48	50.91	74.94	144	144	48	48	144
(3)	440.38	643.12	440.38	613.76	-	-	256	256	1024
Mean squared error relative to OLS									
(1)	0.9649	0.9654	0.9735	0.9578	0.9753	0.9631	0.9707	0.9700	-
(2)	0.5806	0.6426	0.6890	0.6335	0.8723	0.5755	0.5987	0.6232	-
(3)	0.2487	0.2819	-	0.2827	-	-	-	0.2394	-

Note: (1): Simulation 1 - DGP of simulation 1 is generated from a two-country, two-variable model with one lag, B has 16 coefficients, 10 true non-zero. (2): Simulation 2 - DGP of simulation 2 is generated from a sparse three-country two-variable model with four lags, B has 144 coefficients, 34 are true non-zero. (3): Simulation 3 - DGP of simulation 3 is generated from a four-country four-variable model with four lags, B has 1024 coefficients, 432 are true non-zero. The correct sparsity pattern measures how often true relevant variables are included and irrelevant ones excluded. The fraction of relevant variables included counts the number of true relevant variables included in the models relative to the number of all true relevant variables. The number of variables included measured the dimension reduction. MSEs are relative to OLS. All measures are averaged over 100 Monte Carlo replications.

t with $t = T_2, ..., T-1$ and $h = 1, ..., h_{max}$, $h_{max} = 12$. The MSFEs are averaged over t, over all variables and over all Monte Carlo replications.

Table 2 lists the estimators that are compared in the simulation studies. The OLS estimator serves as a benchmark model. However, for larger models, where $T < Kp$, OLS is not feasible. The lag length of estimated PVAR models is set to the true lag length, which means one in the first simulation and four in the second and third simulation.

3.3. Simulation Results

Table 3 and 4 contain the evaluation of the various estimation procedures along the five performance criteria for simulation one, marked as (1), simulation two, (2), and simulation three, (3). The first four columns present the results for the lasso techniques, the next two columns for the Bayesian methods, and the last three for the least squares

Table 4: Mean squared forecast errors relative to OLS

	lasso techniques				Bayesian methods		least squares	
	lasso PVAR	*lasso VAR*	*post lasso*	*adaptive lasso*	*SSSS*	*CC*	*restLS*	*single VAR*
MSFE for $h = 1$								
(1)	0.9541	0.9551	0.9606	**0.9511**	1.0170	0.9775	0.9531	0.9586
(2)	**0.7318**	0.7731	0.8183	0.7659	1.1948	0.7335	0.7390	0.7613
(3)	**0.1362**	0.1716	-	0.1724	-	-	-	0.1375
MSFE for $h = 2$								
(1)	0.9953	0.9953	0.9958	0.9953	1.0000	1.0000	**0.9948**	**0.9948**
(2)	**0.7707**	0.8170	0.8476	0.8117	1.2549	0.7754	0.7825	0.8056
(3)	**0.1468**	0.1874	-	0.1869	-	-	-	0.1477
MSFE for $h = 6$								
(1)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(2)	**0.9262**	0.9361	0.9410	0.9343	1.3866	0.9451	0.9316	0.9329
(3)	**0.0827**	0.0966	-	0.0967	-	-	-	0.0828
MSFE average over 12 horizons								
(1)	0.9957	0.9958	0.9964	**0.9955**	1.0014	0.9982	0.9956	0.9961
(2)	**0.9083**	0.9248	0.9351	0.9227	1.7136	0.9193	0.9136	0.9216
(3)	**0.0740**	0.0901	-	0.0899	-	-	-	0.0741

Note: (1): Simulation 1 - DGP of simulation 1 is generated from a two-country two-variable model with one lag, B has 16 coefficients, 10 true non-zero. (2): Simulation 2 - DGP of simulation 2 is generated from a sparse three-country two-variable model with four lags, B has 144 coefficients, 34 are true non-zero. (3): Simulation 3 - DGP of simulation 3 is generated from a four-country four-variable model with four lags, B has 1024 coefficients, 432 are true non-zero. MSFEs are relative to OLS and average over all t, all countries and variables and over 100 Monte Carlo replications.

estimators. The performance criteria are averages over 100 Monte Carlo replications.[21]

Overall, the simulation studies provide supporting evidence that the use of the lasso for PVARs is beneficial in terms of lower mean squared errors and mean squared forecast errors relative to *OLS*. The forecast performance is additionally improved relative to the selection prior of Koop and Korobilis (2015b) and the factor approach of Canova and Ciccarelli (2009). Accounting for the panel characteristics in the penalty terms leads to better performance in terms of MSEs and MSFE relative to the *lassoVAR* which does not include time series or cross section properties in the penalty terms.

The *lassoPVAR* includes true relevant and discards irrelevant variables in 55.13% of all simulation draws of the first, in 40.83% of the second, and in 47.91% of the third simulation. The fraction of relevant variables included by *lassoPVAR* is 31.90%, simulation one, 38.41%, simulation two, and 44.18%, simulation three. The other lasso techniques reveal similar numbers while *restLS* and *single VAR* find the correct sparsity pattern in fewer cases but more often detect the fraction of relevant variables included. The number of detection of the correct sparsity pattern as well as the fraction of relevant variables in-

[21] Further results for the simulations are given in Appendix E.1.

cluded are low for all methods. The only exception, in some cases, is *OLS*. However, *OLS* does not reduce the dimension and, hence, is not feasible for larger systems.

The lasso techniques clearly reduce the dimension of the models. The *lassoPVAR* includes 32.50% of all variables in simulation one (number of variables included is on average 5.2), 35.35% in simulation two (50.91 variables included), and 43.01% in simulation three (440.38 variables included). That means that for (1) *lassoPVAR* includes fewer variables than the true number of non-zero coefficients, for (2) it picks too many variables, while for (3) it selects around the true number of non-zero coefficients. Hence, *lassoPVAR* performs best in the largest simulation with true non-zero coefficients around 40%. For model (2), which is the sparsest model, the performance of *lassoPVAR* is weaker. This might be due to the underlying model in simulation two, which sets the whole lags two and three to zero. This structure might be better captured by a model setting a whole block of coefficients to zero. These findings are also partly reflected in the numbers for the correct sparsity pattern and the fraction of relevant variables included.

The lower dimension reduction of *lassoVAR* compared to *lassoPVAR* might be driven by the specification of the penalty terms. The penalty terms of *lassoPVAR* introduce additional penalties on higher lags and foreign variables, which results in more variables excluded. *restLS* and *single VAR* reduce the number of variables by one-half in (1), one-third in (2), and one-fourth in (3). *SSSS* and *CC* are shrinkage approaches. Therefore, *SSSS* includes all variables. Since *CC* uses factors to reduce the number of parameters, the first three performance criteria are not applicable.

Compared to the benchmark *OLS*, all estimators reveal lower mean squared errors in all simulations. As expected, due to the moderate number of parameters in simulation one, the gain - measured in lower MSEs - of using lasso or the Bayesian methods is lower compared to the gain in the larger and sparser set-ups of simulations two and three. The MSEs, relative to OLS for simulation one, are in a range between 0.95 and 0.97 for all estimators. In simulation (2), lassoPVAR leads to a substantial reduction of 0.42 in the MSEs relative to OLS and performs second best compared to all other estimators. Only *CC* has a lower MSE at 0.5755. The *adaptive lassoVAR* and *post lassoPVAR* do not yield improvements compared to *lassoPVAR*. The fact that the second stage *OLS* estimation of *post lassoPVAR* relies on the possible misspecified model of the first step of the lasso estimation could explain the performance of the *post lassoPVAR*. For simulation three some models are infeasible due to invertability issues. The *lassoPVAR* clearly outperforms *OLS* with a MSE of 0.2487. Only *single VAR* has a slightly lower value, 0.2394. The weak performance of *OLS*, particularly in terms of MSE for the larger models, reflects the problem of overfitting.

The usage of the selection methods leads to a sizable reduction in mean squared forecast errors compared to *OLS* for all simulations, as shown in table 4. The presented one-step ahead, two-steps ahead, and six-steps ahead MSFEs are averaged over all t, all countries and variables and over the MC replications. The last three rows show the MSFEs additionally averaged over 12 forecast horizons. Lowest MSFEs per row are marked in bold.

Even in the simulation with a small model, where dimension reduction is not required, MSFEs are lower for all estimators compared to *OLS*, except for *SSSS*, and are in a similar range compared among all estimators. For forecast horizon six, the estimators perform equally good. In the second simulation, the use of *lassoPVAR* improves the forecast

Table 5: Overview of empirical applications

	Model (1)	Model (2)	Model (3)
	N=5, G=2, p=6	N=10, G=2, p=6	N=6, G=4, p=6
countries	DE, FR, IT, UK, US	DE, DK, ES, FR, GR, IE, IT, PT, UK, US	DE, ES, FR, IT, UK, US
variables	CPI, IP	CPI, IP	CPI, IP, REER, UN
# of observations	131	131	131
# of parameters per equation	60	120	144

accuracy for all horizons and produces the lowest MSFEs relative to all other methods. Hence, the results provide evidence that accounting for the inherent panel structure within the data by time series and cross section penalties pays off in terms of improved forecast accuracy. Averaged over 12 horizons, the MSFE is 0.9083, a gain of around 0.09 in forecast performance relative to *OLS*. The largest improvement is found for horizon one with a gain of around 0.27. The *lassoPVAR* also produces the lowest MSFEs in simulation three and substantially improves the forecast accuracy relative to *OLS*, with a MSFE averaged over 12 horizons of 0.0740.

For the covariance estimation of *lassoPVAR* the optimal selected ρ is equal to zero. Estimating the covariance with the two-step least squares procedure leads to similar performance results which can be found in Appendix E.2.

4. Forecasting with Multi-Country Models

4.1. Forecasting Including a Global Dimension

This section assesses the forecasting performance of the PVAR estimated with *lassoPVAR* for an empirical application. Of great interest for applied researches and policy makers are forecasts of macroeconomic variables. The forecasting exercise can shed light on whether forecasts of key macroeconomic variables of interlinked countries have to account for possible spillovers across countries. Since panel VARs can exploit international interdependencies and commonalities, they are well suited as forecasting models including a global dimension.

Several studies stress the benefits of accounting for international dependences while forecasting national and international key macroeconomic variables. Ciccarelli and Mojon (2010) and Bjørnland et al. (2017) use a factor model for inflation and GDP forecasts. The authors report improved forecast performance when accounting for national and global factors. Koop and Korobilis (2015a) indicate that using a panel VAR, estimated by a factor approach, for forecasting key macroeconomic indicators of euro zone countries can lead to improvements in forecasts. Dees et al. (2007) forecast inflation of four euro area countries applying sectoral data. Their results provide evidence that forecasts with sectoral PVARs outperform random walk or autoregressive models in the short run.[22]

4.2. Forecasting Applications

In this paper, forecast performance is evaluated for three different models, described in table 5. The benchmark model, model (1), includes monthly 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 extends the country set to ten countries by adding Denmark (DK), Greece (GR), Ireland (IE), Portugal (PT), and Spain (ES). Finally, the third model additionally uses the changes in unemployment rates (UN) and the real effective exchange rate (REER) for six countries: DE, ES, FR, IT, UK, and US.

The number of parameters per equation is larger than the number of observations for model (3) and very close to it for model (2). Hence, for these two models, *OLS* and estimators dependent on OLS, like *adaptive lasso*, *SSSS*, and *CC*, are not feasible. The data provided by the OECD cover the period from 2001:1 to 2016:6. All models include six lags.[23]

An out-of-sample forecasts exercise is conducted. The forecasts are made for the period from 2011:7 to 2016:6. The estimation is based on the data up to 2011:6 rolling forward so that the same amount of time series observations is used for every forecast. The up to twelve-horizons forecasts are iterated forecasts and are calculated by $\hat{Y}_{t+h} = \hat{B}\hat{X}_{t+h-1}$ for $h = 1, ..., 12$. The estimated coefficient matrix, $\hat{B}$, is computed based on the various compared estimators using the observations $t : T_2 + t - 1$ in t, where T_2 denotes the starting point of the forecasting period, 2011:7. The choice of performing iterated rather than direct forecasts is motivated by the results of Marcellino et al. (2006), according to which iterated forecasts are preferred to direct ones despite theoretical findings demonstrating stronger robustness to model misspecification of the latter. The forecasts are evaluated by mean squared forecast errors. The forecasting performance of *lassoPVAR* is compared to the previously explained variants.

4.3. Results of the Forecasting Exercises

Table 6 presents the averaged mean squared forecast errors relative to *OLS* for one-step, two-steps, six-steps, and twelve-steps ahead forecasts for model (1). Additionally, the last row indicates forecast performance averaged over twelve forecast horizons.[24]

Firstly, the use of *lassoPVAR* improves forecast performance relative to *OLS*. The

[22]Other papers use global VAR (GVAR) models to account for international linkages in forecasts. Pesaran et al. (2009) show that multi-country models obtain more accurate forecasts since GVAR forecasts outperform forecasts based on univariate models. Greenwood-Nimmo et al. (2012), Dovern et al. (2016), Huber et al. (2016), and Garratt et al. (2016) provide further evidence that GVAR models improve forecast performance relative to univariate benchmark models. The first study shows the benefits for higher horizon forecasts and the second for predictive joint densities. The third compares GVAR forecasts under various prior specifications while the latter assesses point and density forecasts for GDP growth as well as for the probability of recessions.

[23]The data are seasonally adjusted. CPI is calculated as the log-differences of consumer price indices. UN is the difference of the unemployment rate from one period to the last period. The time series are stationary, de-meaned and standardized.

[24]Further results on country and variable basis are in Appendix F.2.

Table 6: Mean squared forecast error relative to OLS for model (1)

lasso techniques				Bayesian methods		least squares	
lasso PVAR	*lasso VAR*	*post lasso*	*adaptive lasso*	*SSSS*	*CC*	*restLS*	*single VAR*
MSFE for $h = 1$							
0.5708	0.5839	0.6166	**0.5825**	1.7722	0.6002	0.6114	0.6763
MSFE for $h = 2$							
0.5988	0.5978	0.6524	**0.5976**	1.7055	**0.5749**	0.6050	0.6453
MSFE for $h = 6$							
0.6985	0.7123	0.7372	0.7143	2.6418	**0.6624**	0.7153	0.7532
MSFE for $h = 12$							
0.7873	0.7951	0.8056	0.7953	4.5683	**0.7615**	**0.7775**	0.7790
MSFE average over 12 horizons							
0.6783	0.6869	0.7136	0.6870	2.8079	**0.6528**	0.6884	0.7155

Note: The forecast period is from 2011:7 to 2016:6. MSFEs are averaged over all t and are relative to OLS, MSFEs smaller than one indicate better performance relative to OLS. *Average* are the MSFEs additionally averaged over all horizons.

mean squared forecast error averaged over all countries, variables, t and horizons of *lassoPVAR* has the second lowest value with average MSFE of 0.6783. That means that on average using *lassoPVAR* for forecasting leads to a gain of 0.3217 in forecast accuracy compared to *OLS*. *lassoPVAR* produces stable forecasts over all twelve forecast horizons with MSFE relative to *OLS* in a range of 0.79 and 0.57. The benefit of using *lassoPVAR* relative to *OLS* is greatest for one-step ahead forecasts with a gain in forecast performance of 0.4292. None of the other estimators is statistically significantly better in terms of MSFEs than the *lassoPVAR*.[25]

Secondly, accounting for the time series and cross-sectional characteristics in the penalty terms leads to gains in the forecast accuracy. On average, *lassoPVAR* outperforms *lassoVAR* and all but one of the forecasts horizons. Thirdly, the results provide evidence that the use of multi-country models compared to single-county models is beneficial to improve forecast performance. MSFEs of *lassoPVAR* and *CC*, both models accounting for interdependencies across countries, are lower than for the *single VAR* model. The results of the larger applications, model (2) and model (3), strengthen this finding. Since *OLS* is not feasible, Table 7 compares MSFE of *lassoPVAR* and *single VAR* relative to the mean forecast. On average and for most of the horizons *lassoPVAR* outperforms the mean forecasts and the forecasts based on the single country model.

CC shows good performance for small systems, but is infeasible for systems in which the number of parameters per equation exceeds the number of time series observations. That this is a relevant issue for applications, is shown in model (2) and (3) which are of reasonable or even still small size for models addressing potential macroeconomic ques-

[25]Results for the Diebold-Mariano Test assessing the statistical significance of the difference in MSFEs of the models are in Appendix F.2.

Table 7: Mean squared forecast error relative to mean forecast for model (2) and model (3)

N=10, G=2, p=6		N=6, G=4, p=6	
lassoPVAR	*single VAR*	*lassoPVAR*	*single VAR*
MSFE for h = 1			
0.9068	0.9807	**0.9948**	1.0402
MSFE for h = 2			
0.9540	1.0011	1.0476	1.0536
MSFE for h = 6			
0.9588	0.9764	**0.9333**	1.0104
MSFE for h = 12			
0.9519	**0.9253**	**0.9234**	0.9321
MSFE average over 12 horizons			
0.9526	0.9608	**0.9495**	0.9925

Note: The forecast period is from 2011:7 to 2016:6. MSFEs are averaged over all t and are relative to the mean forecast, MSFEs smaller than one indicate better performance relative to the mean forecast. *Average* are the MSFEs additionally averaged over all horizons.

tions in the context of international spillovers. A second issue with the factor approach is the mentioned difficulty for structural identification.

The sparsity pattern of the coefficient matrix for model (1) is given in figure 1.[26] The largest dynamic interdependencies across countries are found for the first and second lags. Own lags have the largest impact, as shown by the darker colors for diagonal elements. In addition, US variables affect variables of other countries, in particular for lags one and two. For benchmark model (1), 600 parameters of the coefficient matrix are estimated. The *lassoPVAR* reduces the dimension by setting 458 coefficients to zero. Thus, 23.67% of the estimated coefficients are non-zero elements. [27]

5. Conclusions

This paper develops a lasso technique for panel VARs, named *lassoPVAR*. It specifies a penalized estimation problem using the weighted sum of squared residuals as the loss function and a penalty incorporating both time series and cross section properties. Thereby, it allows for an unrestricted covariance matrix, meaning that the estimation accounts for possible correlation between variables. The penalty term uses the inherent panel structure within the data. It specifies that more recent or domestic lags provide more information than more distant or foreign lags. As a result, a higher penalty is set for higher lags and foreign variables.

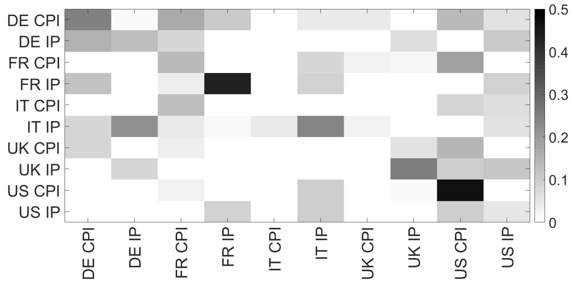
The main results of the paper are as follows. The *lassoPVAR* has the asymptotic oracle property meaning that selection consistency and asymptotic normality are established.

[26]The sparsity pattern for lag 4 and 5 as well as for the covariance matrix are given in Appendix F.2.

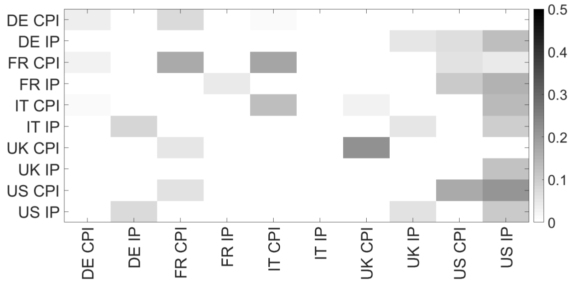
[27]The optimal penalty parameter ρ in the estimation of the covariance is selected to equal zero.

Figure 1: Sparsity pattern of the coefficient matrix for model (1)

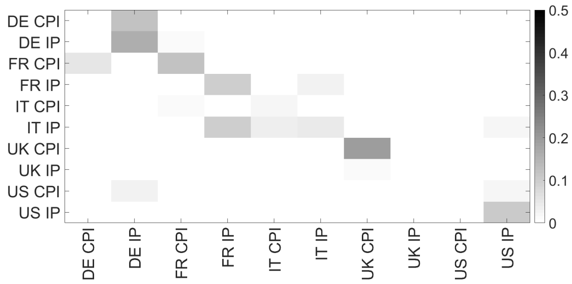
(a) Lag 1



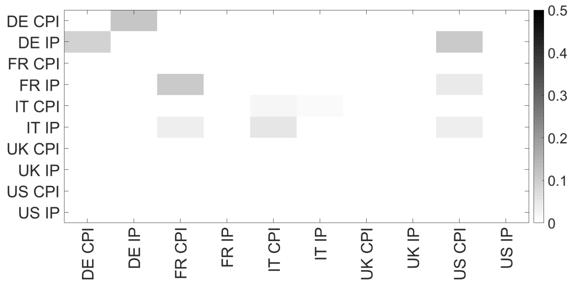
(b) Lag 2



(c) Lag 3



(d) Lag 6



Note: Sparsity pattern of the coefficient matrix B. Negative values are multiplied by -1.

Furthermore, the lasso for PVAR achieves lower mean squared forecast errors, thus increasing forecasting performance compared to estimating the PVAR with OLS. Compared to other Bayesian panel VAR methods and single county models, the *lassoPVAR* improves forecasts, especially for larger models, while mean squared forecast errors are in a similar range for smaller models. These findings are supported by the simulation results and a forecasting exercise that includes up to ten advanced economies and up to four macroeconomic variables. Moreover, accounting for time series and cross section properties in the penalty term is beneficial for the forecast performance as *lassoPVAR* outperforms a lasso estimator without specific penalties. Additionally, the dimension reduction of the lasso techniques leads to reduced mean squared errors compared to OLS in the conducted simulations.

The method proposed in this paper may be of interest for applied researchers, since the lasso for PVAR is able to deal with the curse of dimensionality problem in a multi-country model. *lassoPVAR* ensures the estimation feasibility by using the panel structure in the data and allows at the same time to include interdependencies and heterogeneities across countries in the model. The results presented show that the proposed lasso technique is a useful tool for estimating large panel VAR models.

However, the researcher must be aware that the performance of the lasso is sensitive to the suitability of the analyzed model for the penalized estimation technique. The lasso generally performs well in systems with a large number of parameters and existing sparsity. When few coefficients are large and the others close to zero, the lasso has usually low mean squared errors, while a good performance is not ensured for models deviating from these properties. This point is stressed by Hansen (2016) and is visible in the differences in results for the simulations with DGPs generated from a small and from a larger and sparse model. However, the benefit of the lasso for PVARs is already visible through reduced mean squared errors and improved forecast accuracy in a simulation of moderate size with 165 parameters.

In future research, it may be interesting to further assess different specifications of the penalty term in the context of panel VAR models. One possibility to capture the panel structure is the use of the group lasso, as proposed by Yuan and Lin (2006). The group lasso treats variables in groups, setting whole blocks to zero. This structure might be especially useful for analyses including smaller countries and globally more influential countries. Furthermore, variables in multi-country models might be highly correlated. This issue can be addressed with the elastic-net invented by Zou and Hastie (2005). This procedure is able to select groups of correlated variables while the lasso selects one variable out of a set of correlated variables.

6. Acknowledgements

I thank Helmut Lütkepohl, my supervisor, for insightful comments. Earlier versions of the paper were presented at the SMYE 2017, Halle, the Barcelona GSE Summer Forum Time Series Econometrics and Applications for Macroeconomics and Finance 2017, Barcelona, EEA-ESEM 2017, Lisbon, the annual meeting of the Verein für Socialpolitik 2017, Vienna, the University of Sydney, the Workshop Empirical Macroeconomics at Freie Universität Berlin and at internal seminars at DIW Berlin. Comments by the participants are gratefully acknowledged.

References

Ando, T., Bai, J., 01 2016. Panel Data Models with Grouped Factor Structure Under Unknown Group Membership. Journal of Applied Econometrics 31 (1), 163–191.

Attinasi, M. G., Metelli, L., 2017. Is fiscal consolidation self-defeating? A panel-VAR analysis for the Euro area countries. Journal of International Money and Finance 74, 147 – 164.

Banerjee, O., El Ghaoui, L., d'Aspremont, A., Jun. 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.

Belloni, A., Chernozhukov, V., 05 2013. Least squares after model selection in high-dimensional sparse models. Bernoulli 19 (2), 521–547.

Bergmeir, C., Hyndman, R. J., Koo, B., 2015. A Note on the Validity of Cross-Validation for Evaluating Time Series Prediction. Monash Econometrics and Business Statistics Working Papers 10/15, Monash University, Department of Econometrics and Business Statistics.

Billio, M., Casarin, R., Ravazzolo, F., van Dijk, H. K., Nov. 2014. Interactions between eurozone and US booms and busts: A Bayesian panel Markov-switching VAR model. Working Papers 0026, Centre for Applied Macro- and Petroleum economics (CAMP), BI Norwegian Business School.

Billio, M., Casarin, R., Rossini, L., 2016. Bayesian nonparametric sparse seemingly unrelated regression model (SUR). Working Papers 2016:20, Department of Economics, University of Venice "Ca' Foscari".

Bjørnland, H. C., Ravazzolo, F., Thorsrud, L. A., 2017. Forecasting GDP with global components: This time is different . International Journal of Forecasting 33 (1), 153 – 173.

Breitung, J., 2015. The Analysis of Macroeconomic Panel Data. The Oxford Handbook of Panel Data.

Bühlmann, P., van de Geer, S., 2011. Statistics for High-Dimensional Data. Methods, Theory and Applications. Springer.

Canova, F., Ciccarelli, M., June 2004. Forecasting and turning point predictions in a Bayesian panel VAR model. Journal of Econometrics 120 (2), 327–359.

Canova, F., Ciccarelli, M., 08 2009. Estimating Multicountry VAR Models. International Economic Review 50 (3), 929–959.

Canova, F., Ciccarelli, M., Jan. 2013. Panel vector autoregressive models: A survey. Working Paper Series 1507, European Central Bank.

Canova, F., Ciccarelli, M., Ortega, E., 2012. Do institutional changes affect business cycles? Evidence from Europe. Journal of Economic Dynamics and Control 36 (10), 1520–1533.

Ciccarelli, M., Maddaloni, A., Peydro, J.-L., Mar. 2013. Heterogeneous transmission mechanism: monetary policy and financial fragility in the euro area. Working Paper Series 1527, European Central Bank.

Ciccarelli, M., Mojon, B., August 2010. Global Inflation. The Review of Economics and Statistics 92 (3), 524–535.

Ciccarelli, M., Ortega, E., Valderrama, M. T., January 2016. Commonalities and cross-country spillovers in macroeconomic-financial linkages. The B.E. Journal of Macroeconomics 16 (1), 231–275.

Comunale, M., Feb. 2017. A panel VAR analysis of macro-financial imbalances in the EU. Working Paper Series 2026, European Central Bank.

Davis, R., Zang, P., Zheng, T., 2015. Sparse vector autoregressive modeling. Journal of Computational and Graphical Statistics, 1–53.

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), 1–38.

Dovern, J., Feldkircher, M., Huber, F., 2016. Does joint modelling of the world economy pay off? Evaluating global forecasts from a Bayesian GVAR. Journal of Economic Dynamics and Control 70 (C), 86–100.

Friedman, J., Hastie, T., Höfling, H., Tibshirani, R., 12 2007. Pathwise coordinate optimization. The Annals of Applied Statistics 1 (2), 302–332.

Friedman, J., Hastie, T., Tibshirani, R., 2008. Sparse inverse covariance estimation with the graphical lasso. Biostatistics 9 (3), 432.

Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software 33(1), 1–22.

Garratt, A., Lee, K., Shields, K., 2016. Forecasting global recessions in a GVAR model of actual and expected output. International Journal of Forecasting 32 (2), 374–390.

Gefang, D., 2014. Bayesian doubly adaptive elastic-net Lasso for VAR shrinkage. International Journal of Forecasting 30 (1), 1 – 11.

Georgiadis, G., 2017. To bi, or not to bi? Differences between spillover estimates from bilateral and multilateral multi-country models. Journal of International Economics 107, 1 – 18.

Gnimassoun, B., Mignon, V., 2016. How Do Macroeconomic Imbalances Interact? Evidence from a Panel VAR Analysis. Macroeconomic Dynamics 20 (7), 1717 – 1741.

Greenwood-Nimmo, M., Nguyen, V. H., Shin, Y., 06 2012. Probabilistic forecasting of output growth, inflation and the balance of trade in a GVAR framework. Journal of Applied Econometrics 27 (4), 554–573.

Hansen, B. E., 2016. The Risk of James-Stein and Lasso Shrinkage. Econometric Reviews 35 (8-10), 1456–1470.

Hastie, T., Tibshirani, R., Wainwright, M., 2015. Statistical Learning with Sparsity. The Lasso and Generalizations. Vol. 143 of Monographs on Statistics and Applied Probability.

Hsu, N.-J., Hung, H.-L., Chang, Y.-M., March 2008. Subset selection for vector autoregressive processes using Lasso. Computational Statistics & Data Analysis 52 (7), 3645–3657.

Huber, F., Feldkircher, M., Cuaresma, J. C., 2016. Forecasting with Global Vector Autoregressive Models: A Bayesian Approach. Journal of Applied Econometrics 31 (7), 1371–1391.

Kascha, C., Trenkler, C., 2015. Forecasting VARs, model selection, and shrinkage. Tech. rep., Working Paper ECON 15-07, University of Mannheim.

Kock, A. B., Callot, L., 2015. Oracle inequalities for high dimensional vector autoregressions. Journal of Econometrics 186 (2), 325 – 344, high Dimensional Problems in Econometrics.

Koop, G., Korobilis, D., Nov 2015a. Forecasting With High Dimensional Panel VARs. Working Papers 2015:25, Business School - Economics, University of Glasgow.

Koop, G., Korobilis, D., 2015b. Model uncertainty in panel vector autoregressive models. European Economic Review 81, 115–131.

Korobilis, D., 03 2013. VAR Forecasting Using Bayesian Variable Selection. Journal of Applied Econometrics 28 (2), 204–230.

Korobilis, D., Apr. 2016. Prior selection for panel vector autoregressions. Computational Statistics and Data Analysis 101 (64143), 110–120.

Kyung, M., Gill, J., Ghosh, M., Casella, G., 06 2010. Penalized regression, standard errors, and Bayesian lassos. Bayesian Analysis 5 (2), 369–411.

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.

Love, I., Zicchino, L., May 2006. Financial development and dynamic investment behavior: Evidence from panel VAR. The Quarterly Review of Economics and Finance 46 (2), 190–210.

Lütkepohl, H., 2014. Structural Vector Autoregressive Analysis in a Data Rich Environment: A Survey. Discussion Papers of DIW Berlin 1351, DIW Berlin, German Institute for Economic Research.

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 (1-2), 499–526.

Monache, D., Petrella, I., Venditti, F., 2015. Common faith or parting ways? A time varying parameters factor analysis of euro-area inflation. Birkbeck Working Papers in Economics and Finance 1515, Birkbeck, Department of Economics, Mathematics & Statistics.

Ngueyep, R., Serban, N., 2015. Large-vector autoregression for multilayer spatially correlated time series. Technometrics 57 (2), 207–216.

Nicholson, W. B., Bien, J., Matteson, D. S., Dec. 2016. High dimensional forecasting via interpretable vector autoregression. ArXiv e-prints.

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 (3), 627 – 651.

Park, T., Casella, G., 2008. The bayesian lasso. Journal of the American Statistical Association 103, 681–686.

Pesaran, M. H., Schuermann, T., Smith, L. V., October 2009. Forecasting economic and financial variables with global VARs. International Journal of Forecasting 25 (4), 642–675.

Prez, F., Dec. 2015. Comparing the Transmission of Monetary Policy Shocks in Latin America: A Hierarchical Panel VAR. Working Papers 2015-015, Banco Central de Reserva del Per.

Ren, Y., Xiao, Z., Zhang, X., 2013. Two-step adaptive model selection for vector autoregressive processes. Journal of Multivariate Analysis 116, 349 – 364.

Ren, Y., Zhang, X., 2010. Subset selection for vector autoregressive processes via adaptive lasso. Statistics & Probability Letters 80, 1705 – 1712.

Schnücker, A., 2016. Restrictions search for panel vars. DIW Discussion Papers 1612, Berlin.

Song, S., Bickel, P. J., Jun. 2011. Large Vector Auto Regressions. ArXiv e-prints.

Su, L., Shi, Z., Phillips, P. C. B., 2016. Identifying latent structures in panel data. Econometrica 84 (6), 2215–2264.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B Statistical Methodology 58(1), 267–288.

Wieladek, T., Jan. 2016. The varying coefficient Bayesian panel VAR model. Bank of England working papers 578, Bank of England.

Yuan, M., Lin, Y., 2006. Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68 (1), 49–67.

Zou, H., 2006. The adaptive lasso and its oracle properties. Journal of the American Statistical Association 101(476), 1418–1429.

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B Statistical Methodology 67, 301–320.

Appendix A. The lasso Estimator

The optimization problem of the lasso for PVAR is minimized after b_{km}. The b_{km} is the element of the B-matrix in the k-th row and m-th column. K is the number of countries times the number of variables, $K = NG$. The Y_j and X_m are of dimension $[1 \times T]$ for $j = 1, ..., K$ and $m = 1, ..., Kp$. The ω_{kj} denotes an element of the inverse of the covariance matrix, $\Sigma^{-1} = \Omega$. The λ_{km} is the penalty parameter and $|b_{km}|$ denotes the absolute value of b_{km}. The optimization problem is rewritten as

$$\begin{aligned}
\underset{b_{km}}{\operatorname{argmin}} \;\; \frac{1}{T} \Bigg[& \omega_{kk} \left(Y_k - b_{km} X_k - \sum_{i \neq m}^{Kp} b_{ki} X_i \right) \left(Y_k - b_{km} X_k - \sum_{i \neq m}^{Kp} b_{ki} X_i \right)' \\
& + \sum_{j \neq k}^{K} \omega_{kj} \left(Y_k - b_{km} X_k - \sum_{i \neq m}^{Kp} b_{ki} X_i \right) \left(Y_j - b_{jm} X_k - \sum_{i \neq m}^{Kp} b_{ji} X_i \right)' \\
& + \sum_{j \neq k}^{K} \omega_{jk} \left(Y_j - b_{jm} X_k - \sum_{i \neq m}^{Kp} b_{ji} X_i \right) \left(Y_k - b_{km} X_k - \sum_{i \neq m}^{Kp} b_{ki} X_i \right)' \\
& + \sum_{j \neq k}^{K} \sum_{l \neq k}^{K} \omega_{jl} \left(Y_j - b_{jm} X_k - \sum_{i \neq m}^{Kp} b_{ji} X_i \right) \left(Y_l - b_{lm} X_k - \sum_{i \neq m}^{Kp} b_{li} X_i \right)' \Bigg] \\
& + \lambda_{km} |b_{km}| + \sum_{j \neq k}^{K} \sum_{i \neq m}^{Kp} \lambda_{km} |b_{km}| .
\end{aligned}$$

This simplifies to

$$\begin{aligned}
\frac{1}{T} \Bigg[& \omega_{kk} \left(-2 b_{km} X_m Y_k' + b_{km} X_m X_m' b_{km} + 2 b_{km} X_m \sum_{i \neq m}^{Kp} X_i' b_{ki} + R_1 \right) \\
& + 2 \sum_{j \neq k}^{K} \omega_{jk} \left(-b_{km} X_m Y_j' + b_{km} X_m \sum_{i = m}^{Kp} X_i' b_{ji} + R_2 \right) + R_3 \Bigg] \\
& + \lambda_{km} |b_{km}| + \sum_{j \neq k}^{K} \sum_{i \neq m}^{Kp} \lambda_{km} |b_{km}| ,
\end{aligned}$$

where R_1, R_2 and R_3 collect the terms without b_{km}. Taking the derivative after b_{km}:

$$\begin{aligned}
\frac{1}{T} \Bigg[& \omega_{kk} \left(-2 X_m Y_k' + 2 X_m X_m' b_{km} + 2 X_m \sum_{i \neq m}^{Kp} X_i' b_{ki} \right) \\
& + 2 \sum_{j \neq k}^{K} \omega_{jk} \left(-X_m Y_j' + X_m \sum_{i = m}^{Kp} X_i' b_{ji} \right) \Bigg] + sign(b_{km}) \lambda_{km} \\
& = 0.
\end{aligned}$$

Thus, b_{km}^{lasso} is equal to

$$b_{km}^{lasso} = sign\left(\frac{\sum_{j\neq k}^{K}\omega_{jk}(Y_j - \sum_{i=1}^{Kp} b_{ji}X_i)X_m'}{\omega_{kk}X_mX_m'} + \frac{(Y_k - \sum_{i\neq m}^{Kp} b_{ki}X_i)X_m'}{X_mX_m'}\right)$$
$$\left(\left|\frac{\sum_{j\neq k}^{K}\omega_{jk}(Y_j - \sum_{i=1}^{Kp} b_{ji}X_i)X_m'}{\omega_{kk}X_mX_m'} + \frac{(Y_k - \sum_{i\neq m}^{Kp} b_{ki}X_i)X_m'}{X_mX_m'}\right| - \frac{\lambda_{km}T}{2\omega_{kk}X_mX_m'}\right)$$

Appendix B. Estimation of the Covariance Matrix

Following Friedman et al. (2008) the subgradient of

$$log\ det(\Omega) - tr(S\Omega) - \rho\,||\Omega||$$

with respect to Ω is given by

$$W - S - \rho\Gamma = 0$$

with $W = \hat{\Sigma}$. The elements of Γ give the sign of each element of Ω by being either 1 or -1. For solving the glasso problem the partition

$$\begin{bmatrix} W_{11} & w_{12} \\ w_{12}' & w_{22} \end{bmatrix}\begin{bmatrix} \Omega_{11} & \omega_{12} \\ \omega_{12}' & \omega_{22} \end{bmatrix} = \begin{bmatrix} I & \mathbf{0} \\ \mathbf{0}' & 1 \end{bmatrix}$$

is used. Here, W_{11} is the $(NG-1)\times(NG-1)$ block of W except the j^{th} row and column, w_{12} are the non-diagonal elements of the j^{th} column and row of W and w_{22} is the j^{th} diagonal element of W. The notation is the same for Ω. The partition of the matrix is done rotatively so that each j^{th} row and column is once ordered last. Now, to solve for w_{12} the subgradient is expressed as

$$w_{12} - s_{12} - \rho\gamma_{12} = 0$$
$$W_{11}z - s_{12} + \rho v = 0$$

where γ_{12} is the sign of ω_{12}, $z = -\frac{\omega_{11}}{\omega_{22}} = W_{11}^{-1}w_{12}$, $\gamma_{12} = sign(\omega_{12}) = sign(-\omega_{22}W_{11}^{-1}w_{12})$. Since $\omega_{22} > 0$, $sign(\omega_{12}) = -sign(z)$. The solution of the subgradient $\hat{z}$ gives than the value for w_{12} and $\omega_{12} = -\hat{z}\omega_{22}$. Since the diagonal elements of the covariance matrix are positive, $w_{ii} = s_{ii} + \rho\ \forall\ i$.

The glasso has the following three steps:

1. Set initial value $W = S + \rho I$. For diagonal elements $w_{ii} = s_{ii} + \rho\ \forall\ i$ do not update.
2. For each $j = 1, ..., NG$ update until convergence:
   (a) Partition W and S.
   (b) Solve $W_{11}z - s_{12} + \rho v = 0$.
   (c) $w_{12} = W_{11}\hat{z}$.
3. Compute $\omega_{12} = -\hat{z}\omega_{22}$.

The optimal ρ is chosen over a grid of values by minimizing $BIC_\rho = log(\hat{\Sigma}_\rho) + \frac{log(T_1)}{T_1}df(\rho)$ as done similarly in Kock and Callot (2015). The degrees of freedom, $df(\rho)$, are the number of non-zero elements in $\hat{\Sigma}$. Since the penalty parameter ρ does not vary along the elements of the covariance matrix, the BIC criterion can be used which is faster than the cross-validation technique. The selection of the penalty parameter is done for the period up to T_1.

Appendix C. Optimization Algorithm

The optimization problem is solved by a coordinate descent algorithm as proposed in Friedman et al. (2007) and Friedman et al. (2010). As a starting value B is set equal to a zero matrix. The covariance is estimated in the glasso step. The optimal penalty parameters are determined via a cross-validation technique minimizing MSFEs. The search of the optimal penalty parameters is done over a grid of values. The grid has a length of six for the simulations and twelve for the applications. This rather short length is due to the fact that using a finer grid increases computational time. The algorithm updates every element b_{km} for $k = 1, ..., K$ and $m = 1, ..., Kp$. The following steps are repeated until convergence is archived. Update b_{km} by:

1. Calculate

$$\tilde{b}_{km} = \frac{(Y_k - \sum_{i \neq m}^{Kp} b_{ki} X_i) X'_m}{X_m X'_m} + \frac{\sum_{j \neq k}^{K} \omega_{jk} (Y_j - \sum_{i=1}^{Kp} b_{ji} X_i) X'_m}{\omega_{kk} X_m X'_m}$$

2. Set

$$\lambda_{km} = \begin{cases} \lambda_k p^{\alpha} c & \text{for foreign variables} \\ \lambda_k p^{\alpha} & \text{for domestic variables} \end{cases}$$

   where $\lambda_k > 0$, $\alpha > 0$, $c > 1$, and p is the lag length.
3. Calculate $\tilde{\lambda}_{km} = \frac{\lambda_{km} T}{2 \omega_{kk} X_m X'_m}$
4. Calculate b_{km}^{lasso} by

$$b_{km}^{lasso} = \begin{cases} \tilde{b}_{km} - \tilde{\lambda}_{km} & \text{for } \tilde{b}_{km} > 0, \tilde{\lambda}_{km} < |\tilde{b}_{km}| \\ \tilde{b}_{km} + \tilde{\lambda}_{km} & \text{for } \tilde{b}_{km} < 0, \tilde{\lambda}_{km} < |\tilde{b}_{km}| \\ 0 & \text{for } \tilde{\lambda}_{km} \geq |\tilde{b}_{km}| \end{cases} .$$

5. Set the B-matrix equal to values obtained in the last iteration $B_n = B_{n-1}$.

Convergence is achieved when $max(|B_n - B_{n-1}|) < \epsilon$ where ϵ is a small number. The ϵ is chosen such that the *lassoPVAR* converges to the *OLS* solution for a penalty parameter set to zero and weighted sum of squared residuals as the loss function. For the smaller simulation a conservative value of 0.0000001 is chosen, while for the large simulation (model (3)) $\epsilon = 0.0001$.

Appendix D. Proof of Selection Consistency and Asymptotic Normality

(R1) **Selection consistency**: $plim\ \hat{b}_{km} = 0$ if $b^*_{km} = 0$.

(R2) **Asymptotic normality**: $\sqrt{T}(\hat{b}_J - b^*_J) \xrightarrow{d} \mathcal{N}(0, D^{-1})$.

The vectorized true coefficient matrix is given by $b^* = vec(B^*)$. Let $J = \{(k,m) : b^*_{km} \neq 0\}$ denote the set of non-zero parameters. The lasso estimator of b^* is denoted as $\hat{b}$. The b^*_J is the vector of true non-zero parameters with dimension $[s \times 1]$ and $\hat{b}_J$ is the estimators of b^*_J. Let $Z = I_K \otimes X'$ where X is the $[Kp \times T]$-matrix of right hand side lagged variables. The $y = vec(Y)$ and $u = vec(U)$ are a vector of dimension $[KT \times 1]$. The proof follows Song and Bickel (2011) and Lee and Liu (2012).

Appendix D.1. Proof of asymptotic normality

Let $\beta = \sqrt{T}(b - b^*)$. For the proof it is assumed that Ω is known. If $\hat{\Omega}$ is a consistent estimator of Ω, it can be easily shown that the same steps apply. The lasso optimization problem for the model $y = Zb + u$ is given by:

$$L(\beta) = \left(y - Z\left(b^* + \frac{\beta}{\sqrt{T}}\right)\right)' (\Omega \otimes I_T) \left(y - Z\left(b^* + \frac{\beta}{\sqrt{T}}\right)\right) + T \sum_{k=1}^{K} \sum_{m=1}^{Kp} \lambda_{km} \left| b^*_{km} + \frac{\beta_{km}}{\sqrt{T}} \right|$$

Using $\hat{\beta} = \underset{\beta}{\operatorname{argmin}} L(\beta) = \underset{\beta}{\operatorname{argmin}} (L(\beta) - L(0))$ it follows

$$\begin{aligned}
L(\beta) - L(0) &= \left(y - Z\left(b^* + \frac{\beta}{\sqrt{T}}\right)\right)' (\Omega \otimes I_T) \left(y - Z\left(b^* + \frac{\beta}{\sqrt{T}}\right)\right) \\
&\quad - (y - Zb^*)'(\Omega \otimes I_T)(y - Zb^*) + T \sum_{k=1}^{K} \sum_{m=1}^{Kp} \lambda_{km} \left(\left| b^*_{km} + \frac{\beta_{km}}{\sqrt{T}} \right| - |b^*_{km}| \right) \\
&= (y - Zb^*)'(\Omega \otimes I_T)(y - Zb^*) - \left(Z\frac{\beta}{\sqrt{T}}\right)' (\Omega \otimes I_T)(y - Zb^*) \\
&\quad + \left(Z\frac{\beta}{\sqrt{T}}\right)' (\Omega \otimes I_T) \left(Z\frac{\beta}{\sqrt{T}}\right) - (y - Zb^*)'(\Omega \otimes I_T)\left(-Z\frac{\beta}{\sqrt{T}}\right) \\
&\quad - (y - Zb^*)'(\Omega \otimes I_T)(y - Zb^*) + T \sum_{k=1}^{K} \sum_{m=1}^{Kp} \lambda_{km} \left(\left| b^*_{km} + \frac{\beta_{km}}{\sqrt{T}} \right| - |b^*_{km}| \right) \\
&= \frac{1}{T}\beta' Z'(\Omega \otimes I_T)Z\beta - \frac{2}{\sqrt{T}} (y - Zb^*)' (\Omega \otimes I_T)Z\beta \\
&\quad + T \sum_{k=1}^{K} \sum_{m=1}^{Kp} \lambda_{km} \left(\left| b^*_{km} + \frac{\beta_{km}}{\sqrt{T}} \right| - |b^*_{km}| \right).
\end{aligned}$$

By assumption (A1) for $T \to \infty$

$$\begin{aligned}
\frac{1}{T}\beta' Z'(\Omega \otimes I_T)Z\beta &= \beta' \left(\Omega \otimes \frac{1}{T} Z'Z\right)\beta \\
&\to \beta' (\Omega \otimes \Gamma)\beta
\end{aligned}$$

and since $u \sim \mathcal{N}(0, \Sigma \otimes I)$

$$\begin{aligned}\frac{1}{\sqrt{T}}(y - Zb^*)'(\Omega \otimes I_T)Z &= \frac{1}{\sqrt{T}}u'(\Omega \otimes I_T)Z \\ &\xrightarrow{\text{d}} \mathcal{N}(0, \Omega \otimes \Gamma).\end{aligned}$$

Note that $\Omega = \Sigma^{-1}$ and

$$\begin{aligned}E\left(\frac{1}{\sqrt{T}}Z'(\Omega \otimes I_T)uu'(\Omega \otimes I_T)Z\frac{1}{\sqrt{T}}\right) &= \frac{1}{\sqrt{T}}Z'(\Omega \otimes I_T)E(uu')(\Omega \otimes I_T)Z\frac{1}{\sqrt{T}} \\ &= \frac{1}{T}Z'(\Omega\Sigma \otimes I_T)(\Omega \otimes I_T)Z \\ &= \frac{1}{T}(\Omega \otimes Z'Z) \rightarrow \Omega \otimes \Gamma.\end{aligned}$$

Under assumptions (A2) to (A4) the last term $T\sum_{k=1}^{K}\sum_{m=1}^{Kp}\lambda_{km}(|b^*_{km} + \frac{\beta_{km}}{\sqrt{T}}| - |b^*_{km}|)$ has the following asymptotic behavior for $T \rightarrow \infty$:

$$\begin{cases}\sqrt{T}\lambda_{km}(|b^*_{km} + \frac{\beta_{km}}{\sqrt{T}}| - |b^*_{km}|) \rightarrow 0 & \text{for } b^*_{km} \neq 0 \\ \sqrt{T}\lambda_{km}(|\beta_{km}|) \rightarrow \infty & \text{for } b^*_{km} = 0\end{cases}$$

since for $b^*_{km} = 0$, it holds that $c_T\sqrt{T} \rightarrow \infty$. For $b^*_{km} \neq 0$ since $a_T\sqrt{T} \rightarrow 0$ it follows that $\sqrt{T}\lambda_{km} \rightarrow 0$ and $\sqrt{T}(|b^*_{km} + \frac{\beta_{km}}{\sqrt{T}}| - |b^*_{km}|) \rightarrow \beta_{km}sign(b^*_{km})$. As a result

$$L(\beta) - L(0) \xrightarrow{\text{d}} L(\beta) = \begin{cases}\beta'_J(\Omega \otimes \Gamma)_J\beta_J - 2\beta_J D_J & \text{if } \beta_{km} = 0 \forall (k,m) \notin J \\ \infty & \text{if otherwise}\end{cases}$$

where β_J consists of $\beta_{km} \in J$ and $D_J \xrightarrow{\text{d}} \mathcal{N}(0, (\Omega \otimes \Gamma)_J$. The objective function $L(\beta)$ is minimized by

$$\hat{\beta} = \begin{cases}\hat{\beta}_J & = (\Omega \otimes \Gamma)_J^{-1}D_J \\ \hat{\beta}_{km} & = 0 \ \ \forall (k,m) \notin J\end{cases}$$

Thus, (R2) follows

$$\hat{\beta}_J = \sqrt{T}(\hat{b}_J - b^*_J) \xrightarrow{\text{d}} \mathcal{N}(0, (\Omega \otimes \Gamma)_J)$$

Appendix D.2. Proof of selection consistency

For selection consistency to hold the probability that the estimate of a true zero parameter is unequal zero converges to zero as T goes to infinity, $P(\hat{b}_{km} \neq 0) \rightarrow 0 \ \forall (k,m) \notin J$. Suppose there is a $\hat{b}_{km} \neq 0$ for $(k,m) \notin J$. The Karush-Kuhn-Tucker conditions give the following:

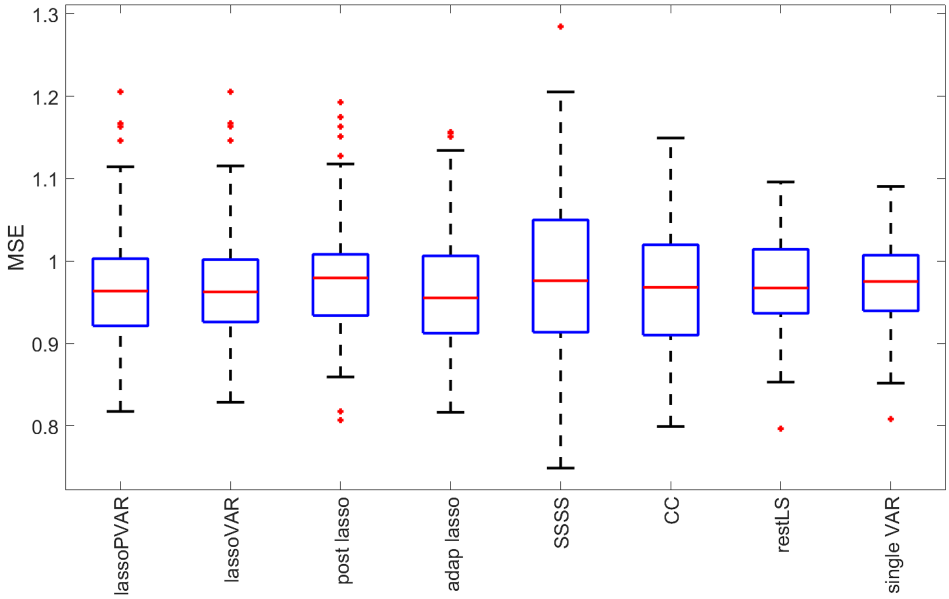
$$0 = \frac{\delta L(\hat{b})}{\hat{b}_{km}} + T\lambda_{km}sign(\hat{b}_{km}).$$

As shown by Song and Bickel (2011) for $T \rightarrow \infty$ the first term is dominated by the second term. Since $c_T\sqrt{T} \rightarrow \infty$, the equation cannot equal zero. Thus, $P(\hat{b}_{km} \neq 0) \rightarrow 0$.

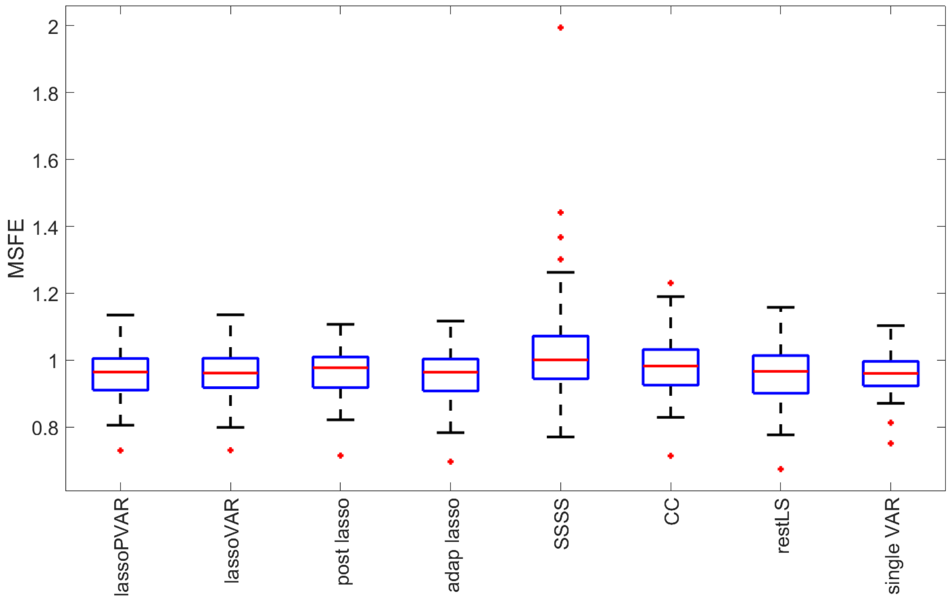
Appendix E. Simulation

Appendix E.1. Simulation Results

Figure E.2: Boxplots of MSEs and MSFEs relative to OLS for simulation 1



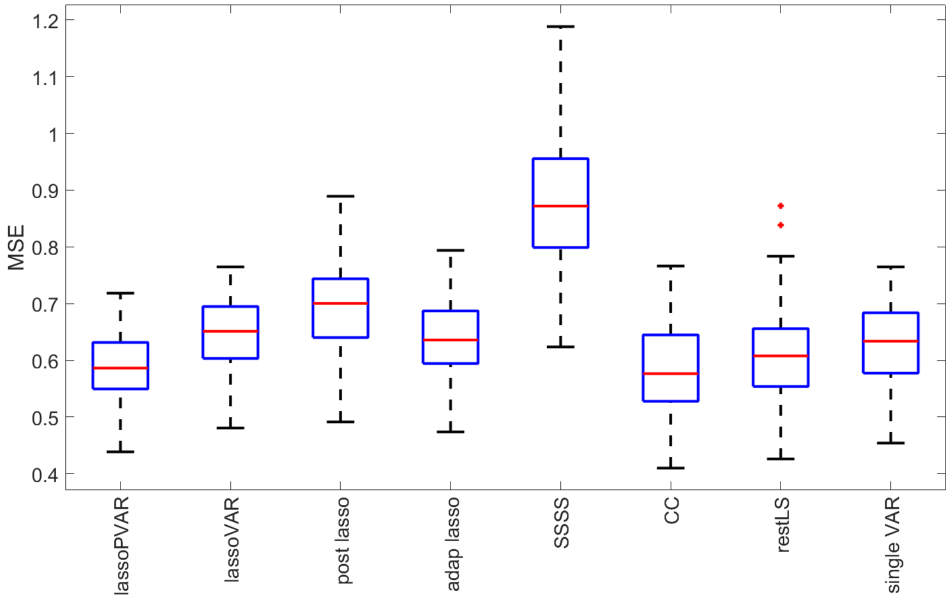
(a) Mean squared errors relative to OLS



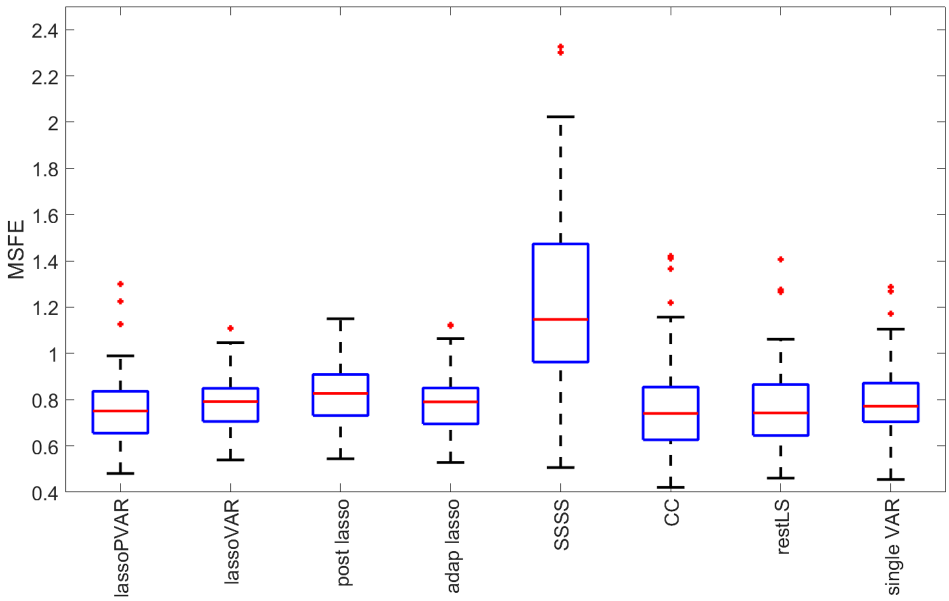
(b) One-step ahead mean squared forecast errors relative to OLS

Note: DGP of simulation 1 is generated from a sparse two-country two-variable model with one lag. The first boxplots show mean squared errors of estimates of parameter matrix B relative to OLS calculated as the average over the deviations of each $\hat{b}_{km}$ from the true value b_{km}^{true} for 100 replications of the simulation. The second figure of boxplots shows one-step ahead mean squared forecast error relative to OLS for 100 simulation replications. MSFE is averaged over t and all variables.

Figure E.3: Boxplots of MSEs and MSFEs relative to OLS for simulation 2



(a) Mean squared errors relative to OLS



(b) One-step ahead mean squared forecast errors relative to OLS

Note: DGP of simulation 2 is generated from a sparse three-country two-variable model with four lags. The first boxplots show mean squared errors of estimates of parameter matrix B relative to OLS calculated as the average over the deviations of each $\hat{b}_{km}$ from the true value b_{km}^{true} for 100 replications of the simulation. The second figure of boxplots shows one-step ahead mean squared forecast error relative to OLS for 100 simulation replications. MSFE is averaged over t and all variables.

Table E.8: Diebold-Mariano Test: test statistic and p-values

	lasso techniques			Bayesian methods		least squares		
	lasso VAR	*post lasso*	*adaptive lasso*	*SSSS*	*CC*	*restLS*	*single VAR*	*OLS*
horizon 1								
(1)	-0.37	-3.33	0.02	-6.53	-3.59	0.09	-0.74	-9.43
	0.36	*0.00*	*0.51*	*0.00*	*0.00*	*0.54*	*0.23*	*0.00*
(2)	-7.95	-9.81	-6.12	-17.90	-0.14	-0.86	-4.75	-15.56
	0.00	*0.00*	*0.00*	*0.00*	*0.45*	*0.20*	*0.00*	*0.00*
(3)	-16.86	-	-16.79	-	-	-	-0.70	-41.44
	0.00	-	*0.00*	-	-	-	*0.24*	*0.00*
horizon 2								
(1)	0.39	-2.69	0.33	-1.86	-1.58	0.78	0.97	-1.46
	0.65	*0.00*	*0.63*	*0.03*	*0.06*	*0.78*	*0.83*	*0.07*
(2)	-2.13	-2.19	-2.14	-2.12	-0.69	-1.41	-1.90	-2.13
	0.02	*0.01*	*0.02*	*0.02*	*0.25*	*0.08*	*0.03*	*0.02*
(3)	-2.11	-	-2.11	-	-	-	-0.29	-2.17
	0.02	-	*0.02*	-	-	-	*0.38*	*0.02*
horizon 6								
(1)	0.01	0.60	1.23	0.91	-0.52	1.99	1.14	1.03
	0.50	*0.73*	*0.89*	*0.82*	*0.30*	*0.98*	*0.87*	*0.85*
(2)	-1.07	-1.07	-1.10	-1.13	-1.11	-1.23	-1.13	-1.11
	0.14	*0.14*	*0.14*	*0.13*	*0.13*	*0.11*	*0.13*	*0.13*
(3)	-1.12	-	-1.13	-	-	-	-0.59	-1.12
	0.13	-	*0.13*	-	-	-	*0.28*	*0.13*

Note: (1): Simulation 1 - DGP of simulation 1 is generated from a two-country two-variable model with one lag, B has 16 coefficients, 10 true non-zero. (2): Simulation 2 - DGP of simulation 2 is generated from a sparse three-country two-variable model with four lags, B has 144 coefficients, 34 are true non-zero. (3): Simulation 3 - DGP of simulation 3 is generated from a four-country four-variable model with four lags, B has 1024 coefficients, 432 are true non-zero. Values of Diebold-Mariano test statistic and p-values which are presented in italic. MSFEs are compared to MSFEs of *lassoPVAR*. MSFEs are averaged over all variables and countries and all MC draws.

Appendix E.2. Simulation Results for the Model with Covariance Estimated with LS

Table E.9: Performance evaluation of estimators: covariance estimated with LS

	lasso PVAR	*lasso VAR*	*post lasso*	*adaptive lasso*
Correct sparsity pattern in %				
(1)	55.31	54.94	55.31	53.37
(2)	37.05	52.57	37.05	48.76
Fraction of relevant variables included in %				
(1)	30.70	33.00	30.70	37.60
(2)	31.35	53.09	31.35	47.68
Number of variables included				
(1)	4.99	5.39	4.99	6.06
(2)	40.67	77.80	40.67	68.64
Mean squared error relative to OLS				
(1)	0.9632	0.9634	0.9692	0.9582
(2)	0.5711	0.6344	0.6512	0.6240

Table E.10: Mean squared forecast errors relative to OLS: covariance estimated with LS

	lasso PVAR	*lasso VAR*	*post lasso*	*adaptive lasso*
MSFE for $h = 1$				
(1)	**0.9550**	0.9555	0.9615	**0.9550**
(2)	**0.7299**	0.7733	0.7991	0.7627
MSFE for $h = 2$				
(1)	0.9953	0.9953	0.9953	0.9953
(2)	**0.7692**	0.8161	0.8276	0.8057
MSFE for $h = 6$				
(1)	1.0000	1.0000	1.0000	1.0000
(2)	**0.9232**	0.9317	0.9313	0.9300
MSFE average over 12 horizons				
(1)	0.9959	0.9959	0.9964	**0.9958**
(2)	**0.9065**	0.9232	0.9262	0.9198

Note: (1): Simulation 1 - DGP of simulation 1 is generated from a two-country two-variable model with one lag, B has 16 coefficients, 10 true non-zero. (2): Simulation 2 - DGP of simulation 2 is generated from a sparse three-country two-variable model with four lags, B has 144 coefficients, 34 are true non-zero. The correct sparsity pattern measures how often true relevant variables are included and irrelevant excluded. The fraction of relevant variables included counts the number of true relevant variables included in the models relative to the number of all true relevant variables. The number of variables included measured the dimension reduction. MSEs are relative to OLS. MSFEs are relative to OLS and average over all t, all countries and variables. All measures are averaged over 100 Monte Carlo replications.

Appendix F. Forecasting Application

Appendix F.1. Penalty Parameters

Table F.11: Grid values for penalty parameters - Application

	Model (1)	Model (2)	Model (3)
λ_k^1	0.376	0.528	0.6193
λ_k^2	0.3427	0.4809	0.5639
λ_k^3	0.3094	0.4338	0.5085
λ_k^4	0.2762	0.3867	0.4531
λ_k^5	0.2429	0.3396	0.3977
λ_k^6	0.2096	0.2925	0.3424
λ_k^7	0.1764	0.2454	0.287
λ_k^8	0.1431	0.1984	0.2316
λ_k^9	0.1098	0.1513	0.1762
λ_k^{10}	0.0765	0.1042	0.1208
λ_k^{11}	0.0433	0.0571	0.0654
λ_k^{12}	0.01	0.01	0.01

Appendix F.2. Application Results

Table F.12: Diebold-Mariano Test: test statistic and p-values. Relative to *lassoPVAR* for model (1)

	lasso techniques			Bayesian methods		least squares		
	lasso VAR	*post lasso*	*adaptive lasso*	*SSSS*	*CC*	*restLS*	*single VAR*	*OLS*
h=1	-0.68	-2.48	-0.82	-7.77	-0.12	-2.48	-0.72	-6.95
	0.25	*0.01*	*0.21*	*0.00*	*0.45*	*0.01*	*0.24*	*0.00*
h=2	0.36	-2.31	0.27	-3.69	1.28	-1.36	-0.03	-3.82
	0.64	*0.01*	*0.61*	*0.00*	*0.90*	*0.09*	*0.49*	*0.00*
h=6	-1.19	-1.75	-1.29	-2.10	1.53	-1.59	-0.70	-1.92
	0.12	*0.04*	*0.10*	*0.02*	*0.94*	*0.06*	*0.24*	*0.03*
h=12	-1.17	-0.94	-1.23	-1.54	1.04	0.67	0.87	-1.61
	0.12	*0.17*	*0.11*	*0.06*	*0.85*	*0.75*	*0.81*	*0.05*

Note: The forecast period is from 2011:7 to 2016:6. Values of Diebold-Mariano test statistic and p-values which are presented in italic. MSFEs are compared to MSFEs of *lassoPVAR*. MSFEs are averaged over all variables and countries.

Table F.13: Diebold-Mariano Test: test statistic and p-values. Relative to *lassoPVAR* for model (2) and model (3)

	N=10, G=2, p=6		N=6, G=4, p=6	
	single VAR	*mean*	*single VAR*	*mean*
h=1	-4.39	-1.03	-0.17	-0.12
	1.00	*0.00*	*0.43*	*0.45*
h=2	-3.00	-0.18	0.62	0.65
	1.00	*0.09*	*0.73*	*0.74*
h=6	-1.88	0.80	-1.90	-1.18
	0.99	*0.10*	*0.03*	*0.12*
h=12	-1.36	1.44	-1.36	-0.55
	0.94	*0.16*	*0.09*	*0.29*

Note: The forecast period is from 2011:7 to 2016:6. Values of Diebold-Mariano test statistic and p-values which are presented in italic. MSFEs are compared to MSFEs of *lassoPVAR*. MSFEs are averaged over all variables and countries.

Table F.14: One-step ahead mean squared forecast error relative to OLS for model (1)

	lasso techniques				Bayesian methods		least squares	
	lasso PVAR	*lasso VAR*	*post lasso*	*adaptive lasso*	*SSSS*	*CC*	*restLS*	*single VAR*
Variable specific mean squared forecast errors								
CPI	0.5755	0.5768	0.6122	0.5777	1.4021	**0.5612**	0.5943	0.6253
IP	**0.5661**	0.5910	0.6211	0.5872	2.1424	0.6392	0.6285	0.7273
Country specific mean squared forecast errors								
DE	0.5531	0.5846	0.6131	0.5908	1.0841	**0.5163**	0.5466	0.6234
FR	**0.6176**	0.6200	0.6809	0.6214	2.1145	0.6771	0.7055	0.7935
IT	**0.7265**	0.7563	0.7258	0.7485	2.3525	0.7998	0.8594	0.9906
UK	**0.5956**	0.6010	0.6119	0.5924	2.3022	0.6804	0.6064	0.6220
US	0.3613	0.3575	0.4515	0.3593	1.0080	**0.3273**	0.3389	0.3520
Mean squared forecast errors averaged over countries and variables								
Average	**0.5708**	0.5839	0.6166	0.5825	1.7722	0.6002	0.6114	0.6763

Note: The forecast period is from 2011:7 to 2016:6. MSFEs are averaged over all t and are relative to OLS, MSFEs smaller than one indicate better performance relative to OLS.

Table F.15: Two-steps ahead mean squared forecast error relative to OLS for model (1)

	lasso techniques				Bayesian methods		least squares	
	lasso PVAR	*lasso VAR*	*post lasso*	*adaptive lasso*	*SSSS*	*CC*	*restLS*	*single VAR*
Variable specific mean squared forecast errors								
CPI	0.6831	0.6936	0.7029	0.6924	1.7323	**0.6530**	0.6676	0.6987
IP	0.5145	0.5020	0.6019	0.5028	1.6788	**0.4968**	0.5423	0.5919
Country specific mean squared forecast errors								
DE	0.5570	0.5499	0.6342	0.5574	1.1186	**0.4938**	0.5380	0.6127
FR	0.6479	0.6352	0.7464	0.6391	1.9491	**0.6046**	0.6771	0.7256
IT	0.7256	0.7162	0.7846	0.7137	1.8606	**0.6898**	0.7822	0.8296
UK	0.6051	0.6276	0.6050	0.6186	2.5384	0.6755	**0.5821**	0.5822
US	0.4584	0.4603	0.4917	0.4592	1.0611	**0.4107**	0.4454	0.4764
Mean squared forecast errors averaged over countries and variables								
Average	0.5988	0.5978	0.6524	0.5976	1.7055	**0.5749**	0.6050	0.6453

Note: The forecast period is from 2011:7 to 2016:6. MSFEs are averaged over all *t* and are relative to OLS, MSFEs smaller than one indicate better performance relative to OLS.

Table F.16: Six-steps ahead mean squared forecast error relative to OLS for model (1)

	lasso techniques				Bayesian methods		least squares	
	lasso PVAR	*lasso VAR*	*post lasso*	*adaptive lasso*	*SSSS*	*CC*	*restLS*	*single VAR*
Variable specific mean squared forecast errors								
CPI	0.7811	0.8054	0.7730	0.8070	2.9802	**0.6926**	0.7749	0.7961
IP	**0.6159**	0.6192	0.7015	0.6216	2.3033	0.6323	0.6556	0.7103
Country specific mean squared forecast errors								
DE	0.6315	0.6349	0.7076	0.6410	1.6743	**0.5625**	0.6466	0.7042
FR	0.7882	0.7950	0.8618	0.7986	3.8808	**0.7126**	0.8389	0.9034
IT	0.7810	0.7865	0.8369	0.7909	1.9969	**0.7648**	0.8561	0.9075
UK	0.7575	0.8015	**0.6934**	0.7964	3.6124	0.7688	0.7036	0.6908
US	0.5342	0.5435	0.5865	0.5444	2.0444	**0.5035**	0.5311	0.5602
Mean squared forecast errors averaged over countries and variables								
Average	0.6985	0.7123	0.7372	0.7143	2.6418	**0.6624**	0.7153	0.7532

Note: The forecast period is from 2011:7 to 2016:6. MSFEs are averaged over all *t* and are relative to OLS, MSFEs smaller than one indicate better performance relative to OLS.

Table F.17: Twelve-steps ahead mean squared forecast error relative to OLS for model (1)

	lasso techniques				Bayesian methods		least squares	
	lasso PVAR	*lasso VAR*	*post lasso*	*adaptive lasso*	*SSSS*	*CC*	*restLS*	*single VAR*
Variable specific mean squared forecast errors								
CPI	0.7792	0.7933	0.7891	0.7934	4.0400	**0.7063**	0.7686	0.7884
IP	0.7954	0.7970	0.8220	0.7972	5.0966	0.8166	0.7864	**0.7697**
Country specific mean squared forecast errors								
DE	0.8006	0.8077	0.8166	0.8084	4.2114	**0.7753**	0.7918	0.7904
FR	0.8124	0.8124	0.8628	0.8123	6.1899	**0.7802**	0.8138	0.8401
IT	0.7929	0.7994	0.8211	0.8001	5.0366	**0.7739**	0.8021	0.7880
UK	0.9751	0.9955	0.9204	0.9950	3.3784	0.9534	0.9274	**0.9201**
US	0.5556	0.5604	0.6068	0.5606	4.0252	**0.5245**	0.5524	0.5565
Mean squared forecast errors averaged over countries and variables								
Average	0.7873	0.7951	0.8056	0.7953	4.5683	**0.7615**	0.7775	0.7790

Note: The forecast period is from 2011:7 to 2016:6. MSFEs are averaged over all *t* and are relative to OLS, MSFEs smaller than one indicate better performance relative to OLS.

Table F.18: Average mean squared forecast error relative to OLS over all forecast horizons for model (1)

	lasso techniques				Bayesian methods		least squares	
	lasso PVAR	*lasso VAR*	*post lasso*	*adaptive lasso*	*SSSS*	*CC*	*restLS*	*single VAR*
Variable specific mean squared forecast errors								
CPI	0.7320	0.7478	0.7428	0.7478	2.6527	**0.6684**	0.7278	0.7517
IP	**0.6247**	0.6261	0.6845	0.6263	2.9632	0.6372	0.6490	0.6793
Country specific mean squared forecast errors								
DE	0.6290	0.6318	0.6828	0.6347	2.3950	**0.5851**	0.6298	0.6706
FR	0.7399	0.7413	0.7990	0.7417	3.7452	**0.7065**	0.7692	0.8065
IT	0.7339	0.7382	0.7742	0.7394	2.7701	**0.7059**	0.7935	0.8345
UK	0.7922	0.8214	0.7547	0.8167	3.1534	**0.7997**	0.7581	0.7544
US	0.4968	0.5020	0.5576	0.5026	1.9760	**0.4667**	0.4914	0.5116
Mean squared forecast errors averaged over countries and variables								
Average	0.6783	0.6869	0.7136	0.6870	2.8079	**0.6528**	0.6884	0.7155

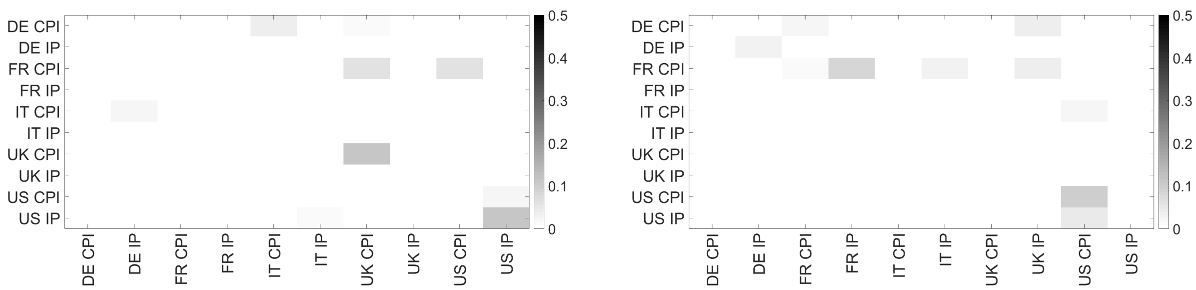
Note: The forecast period is from 2011:7 to 2016:6. MSFEs are averaged over all *t* and are relative to OLS, MSFEs smaller than one indicate better performance relative to OLS.

The tables F.14, F.15, F.16 and F.17 show the forecast evaluation split up into country and variable averages for one-step ahead, two-steps ahead, six-steps ahead and twelve-steps ahead forecasts. Table F.18 presents the average over all twelve forecast horizons. The differences in forecast performance along the two variables are exploited by averaging over all countries for each variable. The differences across countries are evaluated based on the MSFE averaged over the two variables.

lassoPVAR outperforms *OLS* for all variables for all horizons. The same holds for all countries. For one-step ahead forecasts *lassoPVAR* dominates the other estimators for IP and FR, IT and the UK. For higher forecasts horizons *CC* performs best in general. On average as for the six-steps ahead forecasts, *lassoPVAR* has the lowest MSFE for IP forecasts. For all horizons, forecast accuracy of the *lassoPVAR* is improved compared to *lassoVAR* for all countries and variables.

Figure F.4: Sparsity pattern of the coefficient matrix for model (1): lag 4 and 5

(a) Lag 4 (b) Lag 5



Note: Sparsity pattern of the coefficient matrix *B*. Negative values are multiplied by -1.

Figure F.5: Sparsity pattern of the covariance matrix for model (1)