Penalized Estimation of Panel Vector Autoregressive Models: A Panel LASSO Approach

Annika Schnücker*
Erasmus University Rotterdam

Econometric Institute Report EI-2019-33

Abstract

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

Keywords: Model selection, multi-country model, shrinkage estimation
JEL-codes: C13, C32, C33

*Erasmus School of Economics, Tinbergen Institute, Rotterdam, The Netherlands (schnucker@ese.eur.nl). I thank Helmut Lütkepohl, Efrem Castelnuovo, Richard Paap, Andreas Pick and Tomasz Woźniak 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 University of Melbourne, the Workshop Empirical Macroeconomics at Freie Universität Berlin, the University of Konstanz, the University of Vienna, Erasmus University Rotterdam, the University of Groningen, the University of Mannheim, Stockholm School of Economics and at internal seminars at DIW Berlin. Comments by the participants are gratefully acknowledged.

1 Introduction

Panel vector autoregressive (PVAR) models, such as those surveyed by Canova and Ciccarelli (2013) and Breitung (2015), suffer from heavy parameterization. PVAR models allow to augment unit-specific models with lagged variables of another unit, to model covariances between the error terms of different units, and to specify unit-specific coefficient matrices. To tackle the estimation challenge, Bayesian factor approaches for PVAR models compress information in common, unit- and variable-specific factors (Canova and Ciccarelli 2004, 2009; Ciccarelli et al. 2016; Billio et al. 2016; Koop and Korobilis 2018), Bayesian panel selection priors implement higher shrinkage across panel units (Koop and Korobilis 2016; Korobilis 2016), and restricted ordinary least squares approaches assume no dependence or homogeneity across the panel units (Gnimassoun and Mignon 2016; Attinasi and Metelli 2017; Ciccarelli et al. 2013; Comunale 2017).

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

The specified penalty parameters, regulating the amount of shrinkage and selection, build on autoregressive (AR), vector autoregressive (VAR) and PVAR characteristics. That is, the penalization constraints capture that more recent lags provide more important parts of the dynamics than more distant ones. The penalties allow for different penalization among equations. They can model that lags of variables of the same unit are more important than lags of another unit and that coefficients across units might be homogeneous. A higher penalization on increasing lags is in line with the specification of the Litterman prior for VAR models (Litterman 1986). As demonstrated by Song and Bickel (2011) and Nicholson et al. (2016, 2017), including grouping structures or time series properties in the specification of the LASSO for estimating VAR models can improve forecast accuracy compared to the LASSO penalty for VAR models introduced by Hsu et al. (2008) which is fixed for the whole system. Likewise, contributions on Bayesian selection priors for PVAR

models support that accounting for the inherent panel dimension within the data can enhance forecasting performance (Koop and Korobilis 2016; Korobilis 2016).

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

The proposed LASSO allows to estimate PVAR models without the need to restrict interdependencies across and within cross-sectional units a priori. This is especially useful for macroeconomic applications since theoretical arguments for setting restrictions on interdependencies across countries are often missing or hard to justify. Furthermore, the penalized estimation reduces computational burdens by relying on a coordinate descent algorithm compared to Bayesian alternatives for PVAR models which use Markov Chain Monte Carlo algorithms that are limited for large models. Instead of aggregating information in factors which can have an impact on the dynamics of the model and complicates structural identification as well as interpretation, the selection property of the LASSO for PVAR models enables the estimation of large systems and simplifies the interpretation of the model.

The results of a simulation and an empirical application support the use of the penalization estimation specific to the nature of PVAR models as a frequentist alternative to estimate PVAR models which is competitive to alternative techniques. When forecasting inflation and industrial production growth rates for five countries, the proposed LASSO improves the forecast accuracy in terms of mean squared forecast error loses relative to OLS,

Bayesian estimation and to single country models. Accounting for the panel dimension in the penalty terms increases the forecast performance. The simulation results give lower mean squared forecast errors and mean squared errors of the lasso techniques compared to least squares alternatives.

2 Penalized Estimation for PVAR Models

2.1 PVAR Model

Panel vector autoregressive models include several units, such as countries, and unit-specific variables in one model. PVAR models account for interdependencies and heterogeneities across units by jointly modeling multiple variables of several units. A PVAR model with N units and G variables per unit for $t = 1, ..., T$ periods 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 unit i with $i = 1, ..., N$. The $Y_{t-p} = (y'_{1t-p}, ..., y'_{Nt-p})'$ is a $NG \times 1$ vector and the coefficient matrix A_{ip} is of dimension $G \times NG$ for lags $p = 1, ..., P$. The u_{it} have mean zero and covariance matrix Σ_{ii}. The covariance matrices across units are given by $E(u_{it}u'_{jt}) = \Sigma_{ij}$ for $i \neq j$. A penalized estimation typically requires standardization of the data and, for this reason, equation (1) does not contain an intercept.

In compact form, the PVAR model can be written as

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

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

The unrestricted PVAR model allows for dynamic and static interdependencies as well as for heterogeneities across units. The X_{t-1} includes lagged values of every variable in each equation. The unrestricted B-matrix and the covariance matrix Σ enable unit specific

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

2.2 Penalty Term and Loss Function for PVAR Models

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

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

where tr denotes the trace of the matrix and Ω is the precision matrix, $\Omega = \Sigma^{-1}$. The penalty parameters are collected in Λ which is a vector of dimension $1 \times (NG)^2P$ and $\Phi(B)$ is the penalty function with an output of dimension $(NG)^2P \times 1$.

To allow for specific time series and cross section penalties, the elements of Λ are

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

The penalties can be decomposed into three layers: an autoregressive penalty, a vector autoregressive penalty and a PVAR penalty. The autoregressive or time series penalty, p^α, captures that more recent lags provide more information than more distant ones. The penalty increases with the lag order, $p = 1, ..., P$, for $\alpha > 0$. The VAR penalty, λ_k, varies across equations, providing different shrinkage for the multiple time series included in the system. The PVAR penalties, γ and c, induce shrinkage towards homogeneous coefficients

and penalize lags of variables belonging to another cross-sectional unit. The penalty γ shrinks the first lag of a cross section towards a variable specific average of all units. The cross section penalty, $c > 1$ for variables of a different unit, models that lags of variables of the same unit have a larger impact than lags of variables of another unit. The parameters γ, α and c do not vary across equations.

The model specified so far allows for a variety of penalty functions for different penalized estimators such as LASSO, ridge regression or elastic net. For the LASSO the penalty term for the PVAR model is given by

$$\Lambda|vec(B) - vec(\bar{B})| \tag{4}$$

where $vec(B)$ denotes the vectorization of B and $|vec(B) - vec(\bar{B})|$ the absolute values of each element in $vec(B) - vec(\bar{B})$. The $NG \times NGP$-matrix $\bar{B}$ is needed to allow for shrinkage towards homogeneous coefficients. With this formulation, the specified penalty function resembles ideas of the fused LASSO (Tibshirani et al. 2005) where the difference between successive lags is penalized but adopting it to the panel framework. The penalty function penalizes instead the difference between an autoregressive parameter of the first lags and the group average of this parameter. The group average is a homogeneous parameter across all units. Thus, the matrix $\bar{B}$ is given by

$$\bar{B} = (\bar{B}_1, 0_{NG \times NG(P-1)}) \quad \text{with} \quad \bar{B}_1 = \begin{bmatrix} \bar{b}_{11} & \cdots & \bar{b}_{1G} & & \cdots & \cdots & 0 \\ \vdots & \ddots & \vdots & & & & \vdots \\ \bar{b}_{G1} & \cdots & \bar{b}_{GG} & & & & \vdots \\ & & & \ddots & & & \\ \vdots & & & & \bar{b}_{11} & \cdots & \bar{b}_{1G} \\ \vdots & & & & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & & \bar{b}_{G1} & \cdots & \bar{b}_{GG} \end{bmatrix} \tag{5}$$

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

Compared to the LASSO as in Tibshirani (1996) or Hsu et al. (2008) the loss function of the optimization problem is the weighted sum of squared residuals. The weights are

given by the inverse of the covariance matrix Ω. This is in line with Lee and Liu (2012) who point out that in a VAR model correlations between error terms have an impact on the estimated parameters in a restricted regression.

The optimization problem given in (3) is solved for each element, b^{ij}_{klp}, of B.

$$\begin{aligned}\underset{b^{ij}_{klp}}{\operatorname{argmin}} \; & \frac{1}{T}\sum_{t}^{T}\sum_{i,j}^{N}\sum_{k,l}^{G}\omega^{ij}_{kl}\left(Y^{i}_{k,t}-\sum_{p}^{P}b^{ij}_{klp}X^{j}_{lp,t}\right)^{2}+\gamma\sum_{i=j}^{N}\sum_{k,l}^{G}\left|b^{ij}_{kl1}-\bar{b}_{kl}\right| \\ & +\sum_{p=2}^{P}\sum_{i=j}^{N}\sum_{k,l}^{G}\lambda_{k}p^{\alpha}\left|b^{ij}_{klp}\right|+c\sum_{p=1}^{P}\sum_{i\neq j}^{N}\sum_{k,l}^{G}\lambda_{k}p^{\alpha}\left|b^{ij}_{klp}\right|,\end{aligned} \tag{6}$$

where b^{ij}_{klp} is the element in B referring to the p^{th} lag of variable l of unit j in the equation of variable k of unit i for $p = 1, ..., P$, $i, j = 1, ..., N$ and $k, l = 1, ..., G$. The ω^{ij}_{kl} is an element of the inverse of the covariance matrix, $\Sigma^{-1} = \Omega$, corresponding to the l^{th} variable of unit j and the k^{th} variable of unit i. The $Y^{i}_{k,t}$ refers to the k^{th} variable of unit i and $X^{j}_{lp,t}$ is the element in X referring to the p^{th} lag of variable l of unit j in t.

2.3 Penalty Parameters Selection and Estimation of the Covariance and Homogenous Effects

In order to solve (6) we first need to select the optimal penalty parameters, estimate the covariance matrix to obtain an estimate for Ω and estimate the homogeneous effects in $\bar{B}$.

I select the optimal penalty parameters via a rolling cross-validation technique following Song and Bickel (2011) and Nicholson et al. (2016, 2017). The penalty parameters are chosen such that they minimize one-step ahead mean squared forecast errors. This procedure accounts for time dependence. Alternatively, Stock and Watson (2012) and Bergmeir et al. (2018) use leave-m-out cross-validation also for time series data. Like Song and Bickel (2011), I split the sample in three periods: The first period from 1 to $T_1 - 1$ is used for estimating the model, based on the second period from T_1 to $T_2 - 1$ different penalty parameters are evaluated, and the third period from T_2 to the end of the sample is later used for forecast evaluation of the lasso. The model is estimated in a rolling scheme taking the observations from t to $T_1 + t - 1$ for $t = 1, ..., T_2 - T_1$. For each t the out-of-sample forecast accuracy for a specific penalty parameter λ_k is measured by the one-step ahead

mean squared forecast error (MSFE) for variable k, $k = 1, ..., NG$. The search for the optimal λ_k is done over a grid of penalty parameter values whereby at the maximal value all coefficients equal zero while γ, a and c are fixed. This is repeated for different γ, a and c combinations. The optimal combination of all penalty parameters is chosen such that the average MSFEs are minimized. 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 one to $T_2 - 1$.

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

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

$$\log\det(\Omega) - tr(S\Omega) - \rho\,||\Omega|| \tag{7}$$

is maximized with respect to the non-negative definite inverse of the covariance matrix $\Omega = \Sigma^{-1}$. The matrix S is the empirical covariance, $tr(S\Omega)$ is the trace of $S\Omega$ and $||\Omega||$ is the sum of the absolute values of each element of Ω. 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 invertible even in the case when the number of variables is larger than the number of observations due to the regularization using $\rho > 0$. This estimator of the covariance matrix is plugged into equation (6).

An alternative way to estimate the covariance matrix is to use a joint likelihood approach (Lee and Liu 2012; Basu and Michailidis 2015; Davis et al. 2016; Ngueyep and

Serban 2015) or the least squares estimator $\hat{\Sigma} = \frac{1}{T-kk}(Y - \hat{B}X)(Y - \hat{B}X)'$, where kk is the number of degrees of freedom (Tibshirani 1996). In contrast to the GLASSO estimation, this approach can lead to problems for the invertability of the covariance matrix in large systems.

The homogenous coefficients are calculated with an mean-group or aggregated estimator. The mean-group estimator is the average of the coefficients estimated from unit specific VAR models. For the aggregated estimator the data are averaged and a VAR model is estimated with averaged data. Both estimators are computed with ordinary least squares.

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

3 Simulation Studies

3.1 Benchmark Estimation Procedures for PVAR Models

This section evaluates the finite sample performance of the penalized estimators for PVAR models. I consider four variants of the estimator: a LASSO with equation-specific shrinkage by λ_k, lag specific shrinkage by α, higher shrinkage for lags of another unit by c and weighted sum of squared residuals, a LASSO with the same panelty parameters where the covarinace matrix is set to the identity matrix, a LASSO with equation-specific shrinkage by λ_k, lag specific shrinkage by α, higher shrinkage for lags of another unit by c, shrinkage towards homogeneous coefficients by γ, and with weighted sum of squared residuals, and a LASSO with equation-specific shrinkage by λ_k where the covarinace matrix is set to the identity matrix.

The performance is also compared to two alternative estimation procedures for PVAR

models. As a general benchmark model, the PVAR model is estimated with ordinary least squares. However, while it can serve as a benchmark for small models, OLS is unfeasible for larger models for which $T < NGp$. In addition, a block-diagonal system ordering the variables in unit blocks is estimated with OLS, thus, assuming no dynamic interdependencies between units. Such an a-priori assumption can be hard to justify theoretically especially for macroeconomic applications. Furthermore, I compare the performance to separate VAR models for each unit estimated with OLS and with a LASSO with equation-specific shrinkage.

3.2 Performance Criteria

The performance of the estimators is evaluated along the following criteria (similar to e.g. Tibshirani 1996; Ren and Zhang 2010; Kock and Callot 2015).

- **MSFE**. The h-step ahead mean squared forecast error for variable k of unit i for one Monte Carlo replication is calculated as

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

  where $\hat{Y}_{k,t+h}^i$ denotes the iteratively estimated h-step ahead forecast for variable k of unit i for t with $t = T_2, ..., T-1$ and $h = 1, ..., h_{max}$, $h_{max} = 12$.

- **Correct sparsity pattern**. The measure calculates how often the evaluated procedure takes the correct decision whether to include or exclude a variable.

- **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.

- **Number of variables included**. The measure reports the average number of variables included in the model. It evaluates the dimension reduction done by the estimator.

- **MSE**. The mean squared error of the parameter estimates for one Monte Carlo

replication is calculated as

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

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

3.3 Simulation Set-Ups

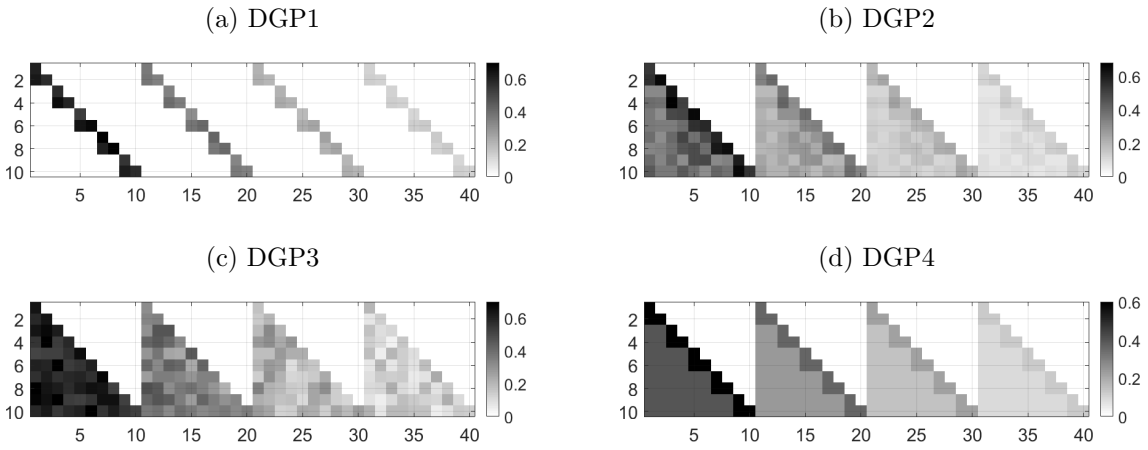
I consider four Monte Carlo simulations, each with sample sizes $T = 100$ and $T = 500$ for number of unites $N = 5$ and number of variables per unit $G = 2$. All DGPs are generated from stationary PVAR models. The idiosyncratic errors, U_t, for all data generating processes (DGP) are assumed to be normally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

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

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

- **DGP3**. Similar to DGP2 the data are generated from a PVAR(4) model where each B_p, for $p = 1, ..., 4$, has a triangular structure. In contrast to DGP3 all non-zero coefficients in B_1 are 0.6 plus a number randomly drawn in the interval $[-0.1, 0.1]$.

Figure 1: Coefficient matrices of the four DGPs



NOTE: Zero coefficients are colored in white and non-zero coefficients in gray shades whereby negative values are multiplied by -1 and the darkest shade is given to the highest coefficient value.

For $p = 2, 3, 4$ the non-zero coefficients equal $(-0.6)^{(p-1)}$ times the corresponding element in B_1. This case presents a sparse model with heterogenous coefficients and smaller coefficients for distant lags. There is no distinction between lags of the same unit and of a different unit.

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

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

Table 1: Mean squared forecast errors relative to OLS

	LASSO techniques					single unit	
	λ_k, c, α	λ_k, c, α $\Sigma = I$	$\lambda_k, c, \alpha, \gamma$ *(AG)*	λ_k $\Sigma = I$	*bl-diag* *OLS*	*OLS*	*LASSO* λ_k
One-step ahead mean squared forecast errors							
DGP1	0.49^{***}	0.48^{***}	0.50^{***}	0.48^{***}	0.51^{***}	0.52^{***}	0.48^{***}
DGP2	0.49^{***}	0.48^{***}	0.51^{***}	0.48^{***}	0.51^{***}	0.51^{***}	0.48^{***}
DGP3	0.47^{***}	0.47^{***}	0.50^{***}	0.47^{***}	0.50^{***}	0.51^{***}	0.47^{***}
DGP4	0.49^{***}	0.48^{***}	0.50^{***}	0.48^{***}	0.51^{***}	0.51^{***}	0.48^{***}
Two-steps ahead mean squared forecast errors							
DGP1	0.58^{**}	0.58^{**}	0.58^{**}	0.58^{**}	0.60^{**}	0.60^{**}	0.58^{**}
DGP2	0.57^{**}	0.57^{**}	0.58^{**}	0.57^{**}	0.59^{**}	0.60^{**}	0.57^{**}
DGP3	0.57^{**}	0.57^{**}	0.57^{**}	0.57^{**}	0.59^{**}	0.60^{**}	0.56^{**}
DGP4	0.57^{**}	0.57^{**}	0.57^{**}	0.57^{**}	0.59^{**}	0.60^{**}	0.57^{**}
Six-steps ahead mean squared forecast errors							
DGP1	0.78	0.78	0.78	0.78	0.78	0.78	0.78
DGP2	0.75	0.75	1.73	0.75	0.76	0.76	0.75
DGP3	0.74	0.74	0.74	0.74	0.74	0.74	0.74
DGP4	0.75	0.76	0.76	0.75	0.76	0.76	0.75

NOTE: MSFEs are relative to a PVAR model estimated by OLS and average over all t, all units and variables and over 1000 Monte Carlo replications. Significance level of Diebold Mariano test is indicated by $***$ (1%), $**$ (5%), and $*$ (10%). Estimators (in columns): LASSO with penalties λ_k, α, c and weighted sum of squared residuals, LASSO with the same penalty parameters where the covariance matrix is set to the identity matrix, LASSO with penalties $\lambda_k, \alpha, c, \gamma$ with aggregated estimator for homogeneous coefficients and with weighted sum of squared residuals, LASSO λ_k where the covariance matrix is set to the identity matrix, block-diagonal system ordering the variables in unit blocks estimated with OLS, separate VAR models for each unit estimated with OLS and with a LASSO with equation-specific shrinkage.

Table 2: Performance evaluation of estimators

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

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

3.4 Simulation Results

Tables 1 and 2 contain the evaluation of the various estimation procedures along the five performance criteria for the four DGPs for $T = 100$. The performance criteria are averages over 1000 Monte Carlo replications. Overall, the simulation studies provide supporting evidence that using LASSO techniques for estimating PVAR models is beneficial in terms of lower mean squared forecast errors and mean squared errors relative to a PVAR model estimated with OLS.

The usage of the selection methods leads to a sizable reduction in mean squared forecast errors compared to OLS for all simulations for the presented one-, two-, and six-steps ahead forecasts, table 1. The largest improvement is found for horizon one. The performance compared among the LASSO alternatives is similar. The gain in forecast performance of the LASSO techniques relative to OLS in a PVAR model is larger than the accuracy gain from a block-diagonal model estimated with OLS or a single unit VAR model estimated with OLS.

The LASSO alternative with λ_k, c and α includes true relevant and discards irrelevant variables in the same range as the other LASSO techniques. The least squares estimator mostly find the correct sparsity pattern in fewer cases but more often detect the fraction of relevant variables included, shown in table 2. The number of detection of the correct sparsity pattern as well as the fraction of relevant variables varies among the four DGPs. The LASSO techniques find more often the correct sparsity pattern for the last three DGPs compared to DGP1.

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

Results for $T = 500$ and $N = 5, G = 5$ are presented in Appendix D.1. As expected, the performance among the LASSO techniques and the least squares estimators become align for $T = 500$ while for $N = 5, G = 5$ the outperformance of the LASSO techniques with respect to the least squares alternatives is pronounced.

4 Forecasting with PVAR Models

Panel VAR models are well suited as forecasting models for macroeconomic time series since they allow for international interdependencies and commonalities. Several studies report good forecasting performance of models that account for international dependences while forecasting national and international key macroeconomic variables (e.g. Dees et al. 2007; Pesaran et al. 2009; Ciccarelli and Mojon 2010; Huber et al. 2016; Bjørnland et al. 2017; Koop and Korobilis 2018).

4.1 Forecasting Application

The forecast performance of the penalized estimators for PVAR models is evaluated for inflation and industrial production forecasts. The PVAR includes monthly log changes in the harmonized index of consumer prices (CPI) and industrial production growth (IP) for five countries: Germany (DE), France (FR), Italy (IT), the United Kingdom (UK), and the United States (US). The time series are seasonally adjusted, de-meaned and standardized. The data are from the OECD and cover the period from 2001:1 to 2016:6. The model includes six lags.

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

The optimal penalty parameters are selected via the cross-validation procedure. The grid of λ_k consists of twelve values between 0.01 and $(1/T)\lambda_k^{max} = \max(\max(XY'))$. The grids for α, c, γ and ρ have a length of five. For each equation the optimal penalty parameters minimize the MSFEs. See Appendix E.2 for details on the penalty selection.

In addition to the alternative models described in 3.1, I consider a LASSO with weighted sum of squared residuals and equation-specific shrinkage by λ_k and a mean-group estimator as an alternative for the estimation of the homogeneous coefficients in the LASSO. I also use an adaptive LASSO. The adaptive LASSO (Zou 2006; Ren and Zhang 2010) penalizes large

non-zero coefficients less than very small coefficients by using the inverse of an unbiased estimator as adaptive weights. The adaptive LASSO applied here uses the weighted sum of squared residuals, λ_k and OLS estimates as weights. An issue with the adaptive LASSO is the choice of the weights as OLS estimates are unfeasible for large systems. Alternatively, ridge regression can be used to estimate the weights. However, this again requires the determination of a penalty parameter.

Furthermore, two Bayesian alternative estimation procedures for PVAR models are used: the Bayesian selection prior of Koop and Korobilis (2016) called stochastic search specification selection and the cross-sectional shrinkage approach proposed by Canova and Ciccarelli (2004, 2009). The use of the selection prior is limited since it relies on a Markov Chain Monte Carlo algorithm not suitable for large systems. The cross-sectional shrinkage approach groups coefficients due to factorizing, however, it does not use possible sparsity for dimension reduction (see Korobilis 2016). Appendix E.1 gives details on the Bayesian methods.

4.2 Results of the Forecasting Exercises

Table 3 and 4 present variable- and country-specific one-step ahead mean squared forecast errors relative to a PVAR model estimated by OLS. The stars in the tables indicate the significance levels of Diebold Mariano tests. MSFEs for higher forecast horizons, two-steps and six-steps, are given in figure 2.

The use of LASSO techniques for PVAR models improves forecast performance relative to OLS, table 3. The gains are mainly statistically significant. The LASSO alternative with λ_k, c and α produces on average the lowest one-step ahead mean squared forecast errors among all LASSO procedures. On average, the gain in forecast performance of the LASSO with λ_k, c and α is 0.41 relative to OLS and 0.42 for the LASSO with λ_k, c and α without a loss function weighted by the covariance matrix. The slight improvement of the second LASSO version can be explained by the reduced uncertainty due to setting the covariance matrix equal to the identity matrix. Accounting for the time series and cross-sectional characteristics in the penalty terms leads to gains in the forecast accuracy. The first two LASSO techniques outperform on average the LASSOs with λ_k and also the

Table 3: One-step ahead mean squared forecast error of LASSO techniques relative to PVAR model estimated by OLS

	LASSO techniques						
	λ_k, c, α	λ_k, c, α $\Sigma = I$	$\lambda_k, c, \alpha, \gamma$ *(MG)*	$\lambda_k, c, \alpha, \gamma$ *(AG)*	λ_k	λ_k $\Sigma = I$	*adaptive*
Variable specific mean squared forecast errors							
CPI	0.59^{***}	0.59^{***}	0.66^{***}	0.63^{***}	0.60^{***}	0.61^{***}	0.61^{***}
IP	0.59^{***}	0.57^{***}	0.63^{***}	0.92^{*}	0.62^{***}	0.63^{***}	0.61^{***}
Country specific mean squared forecast errors							
DE	0.53^{***}	0.53^{***}	0.67^{***}	0.71^{***}	0.58^{***}	0.57^{***}	0.58^{***}
FR	0.62^{***}	0.62^{**}	0.63^{**}	0.82	0.62^{***}	0.63^{**}	0.62^{**}
IT	0.73^{**}	0.73^{***}	0.80^{**}	1.20	0.74^{**}	0.74^{**}	0.73^{**}
UK	0.65^{***}	0.64^{***}	0.78^{**}	0.68^{**}	0.66^{***}	0.68^{***}	0.70^{***}
US	0.43^{***}	0.40^{***}	0.34^{***}	0.48^{***}	0.46^{***}	0.49^{***}	0.44^{***}
Mean squared forecast errors averaged over countries and variables							
Average	0.59^{***}	0.58^{***}	0.64^{***}	0.78^{***}	0.61^{***}	0.62^{***}	0.61^{***}

NOTE: The forecast period is from 2011:7 to 2016:6. MSFEs are averaged over all t and are relative to a PVAR model estimated by OLS, MSFEs smaller than one indicate better performance relative to OLS. Significance level of Diebold Mariano test is indicated by $***$ (1%), $**$ (5%), and $*$ (10%). Estimators (in columns): LASSO with penalties λ_k, α, c and weighted sum of squared residuals, LASSO with the same penalty parameters where the covariance matrix is set to the identity matrix, LASSO with penalties $\lambda_k, \alpha, c, \gamma$ with mean-group estimator for homogeneous coefficients and with weighted sum of squared residuals, same LASSO with aggregated estimator for homogeneous coefficients, LASSO with weighted sum of squared residuals and λ_k, LASSO λ_k where the covariance matrix is set to the identity matrix, and adaptive LASSO.

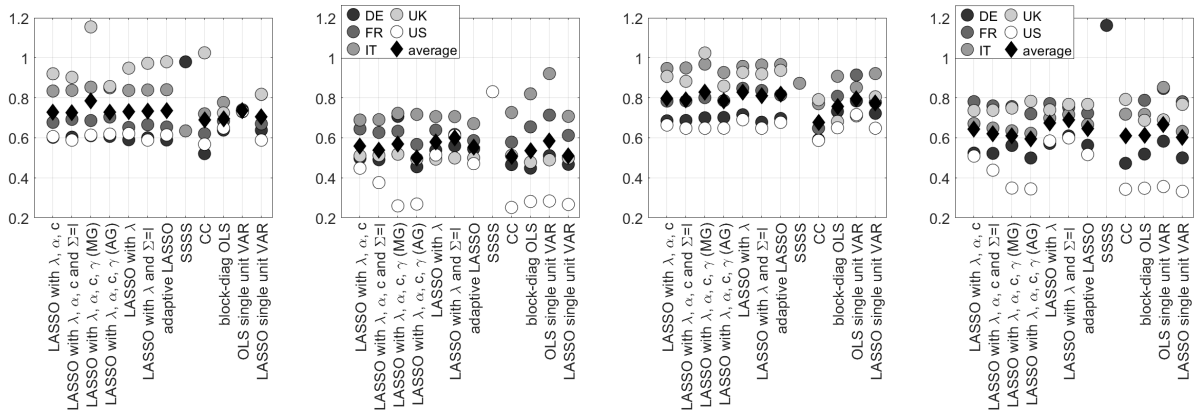
Table 4: One-step ahead mean squared forecast error relative to PVAR model estimated by OLS

		Bayesian methods			single unit	
	LASSO			*bl-diag*		*LASSO*
	λ_k, c, α	*SSSS*	*CC*	*OLS*	*OLS*	λ_k
Variable specific mean squared forecast errors						
CPI	0.59^{***}	1.50	0.60^{***}	0.62^{***}	0.65^{***}	0.61^{***}
IP	0.59^{***}	2.60^{***}	0.65^{***}	0.61^{***}	0.70^{***}	0.61^{***}
Country specific mean squared forecast errors						
DE	0.53^{***}	1.13	0.51^{***}	0.55^{***}	0.63^{***}	0.54^{***}
FR	0.62^{***}	2.43^{***}	0.69^{**}	0.71^{*}	0.80	0.67^{**}
IT	0.73^{**}	2.95^{***}	0.84^{***}	0.85^{**}	0.96^{**}	0.85^{*}
UK	0.65^{***}	2.50^{***}	0.75^{**}	0.61^{***}	0.63^{***}	0.65^{***}
US	0.43^{***}	1.23	0.33^{***}	0.35^{***}	0.37^{***}	0.33^{***}
Mean squared forecast errors averaged over countries and variables						
Average	0.59^{***}	2.05^{***}	0.62^{***}	0.62^{***}	0.67^{***}	0.61^{***}

NOTE: The forecast period is from 2011:7 to 2016:6. MSFEs are averaged over all t and are relative to a PVAR model estimated by OLS, MSFEs smaller than one indicate better performance relative to OLS. Significance level of Diebold Mariano test is indicated by $***$ (1%), $**$ (5%), and $*$ (10%). Estimators (in columns): LASSO with penalties λ_k, α, c and weighted sum of squared residuals, stochastic search specification selection of Koop and Korobilis (2016), cross-sectional shrinkage approach of Canova and Ciccarelli (2004, 2009), block-diagonal system ordering the variables in unit blocks estimated with OLS, separate VAR models for each unit estimated with OLS and with a LASSO with equation-specific shrinkage.

Figure 2: Two-steps and six-steps ahead mean squared forecast errors relative to PVAR model estimated by OLS

(a) $h = 2$: CPI (b) $h = 2$: IP (c) $h = 6$: CPI (d) $h = 6$: IP



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 circles indicate MSFEs for CPI and IP for the five countries. *Average* are the MSFEs additionally averaged over all countries either for CPI or IP.

adaptive LASSO. The performance of the LASSOs allowing also for shrinkage towards homogeneity is weaker.

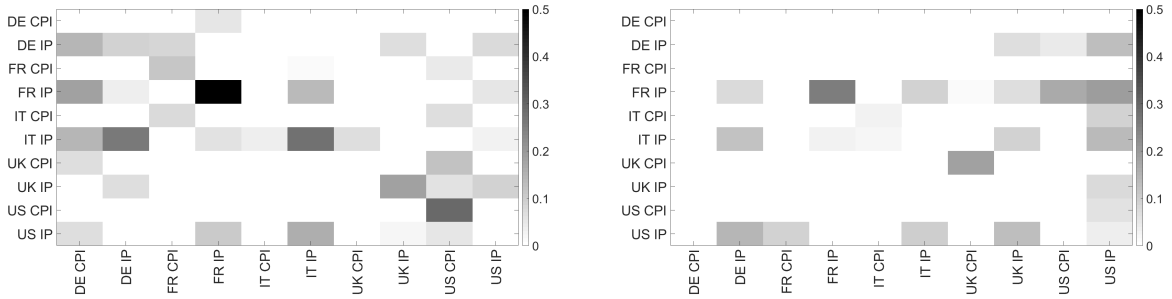
The largest gain in variable-specific forecast performance is found for one-step ahead industrial production growth forecasts for the second LASSO variant. Overall MSFEs are lower for industrial production growth forecasts than for inflation forecasts. The mean squared forecast errors are particularly low for the US for selection methods compared to OLS. Variables of other countries have a low impact on US variables, thus, including these variables does not seem to improve the forecasts for the US.

The LASSO with λ_k, c and α outperforms on average Bayesian methods for estimating PVAR models, block-diagonal VAR models estimated with OLS and single unit VAR models, shown in table 4. While compressing information into factors as done by the approach of Canova and Ciccarelli (2009), *CC*, yields comparable forecasting accuracy to the LASSO, the Bayesian selection prior of Koop and Korobilis (2016), *SSSS*, performs considerable worse.

The results provide some evidence that the use of multi-country models estimated with

Figure 3: Sparsity pattern of the coefficient matrix

(a) Lag 1 (b) Lag 2



NOTE: LASSO with penalties λ_k, α, c and weighted sum of squared residuals estimates for the first and second lags of all variables. Zero coefficients are colored in white and non-zero coefficients in gray shades whereby negative values are multiplied by -1 and the darkest shade is given to the highest coefficient value.

dimension-reduction techniques compared to single-county models is beneficial to improve forecast performance. MSFEs of models accounting for interdependencies across countries are lower than MSFEs of single unit VARs estimated with OLS and in some cases also lower than MSEFs of sinlge unit VARs estimated with LASSO.

The findings for two-steps and six-steps ahead forecasts, figure 2, are in line with the results for one-step ahead forecasts. Further results for additional forecast horizons, given in Appendix E.3, support the good forecast accurancy of LASSO techniques for PVAR models compared to alternative methods.

The sparsity pattern of the coefficient matrix is given in figure 3. Own lags have the largest impact, as shown by the darker colors for diagonal elements. In addition, US variables affect variables of other countries.

5 Conclusions

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

across units.

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

References

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., 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. Annals of Statistics 43, 1535–1567.

Bergmeir, C., Hyndman, R.J., Koo, B., 2018. A Note on the Validity of Cross-validation for Evaluating Autoregressive Time Series Prediction. Computational Statistics & Data Analysis 120, 70 – 83.

Billio, M., Casarin, R., Ravazzolo, F., van Dijk, H.K., 2016. Interconnections Between Eurozone and US Booms and Busts Using a Bayesian Panel Markov-Switching VAR Model. Journal of Applied Econometrics 31, 1352–1370.

Bjørnland, H.C., Ravazzolo, F., Thorsrud, L.A., 2017. Forecasting GDP with Global Components: This Time Is Different. International Journal of Forecasting 33, 153–173.

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

Canova, F., Ciccarelli, M., 2004. Forecasting and Turning Point Predictions in a Bayesian Panel VAR Model. Journal of Econometrics 120, 327–359.

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

Canova, F., Ciccarelli, M., 2013. Panel Vector Autoregressive Models: A Survey. Working Paper Series 1507. European Central Bank.

Ciccarelli, M., Maddaloni, A., Peydró, J.L., 2013. Heterogeneous Transmission Mechanism: Monetary Policy and Financial Fragility in the Eurozone. Economic Policy 28, 459–512.

Ciccarelli, M., Mojon, B., 2010. Global Inflation. Review of Economics and Statistics 92, 524–535.

Ciccarelli, M., Ortega, E., Valderrama, M.T., 2016. Commonalities and Cross-Country Spillovers in Macroeconomic-Financial Linkages. B.E. Journal of Macroeconomics 16, 231–275.

Comunale, M., 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., 2016. Sparse Vector Autoregressive Modeling. Journal of Computational and Graphical Statistics 25, 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, 1–38.

Friedman, J., Hastie, T., Höfling, H., Tibshirani, R., 2007. Pathwise Coordinate Optimization. Annals of Applied Statistics 1, 302–332.

Friedman, J., Hastie, T., Tibshirani, R., 2008. Sparse Inverse Covariance Estimation with the Graphical Lasso. Biostatistics 9, 432–441.

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

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

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

Hsu, N.J., Hung, H.L., Chang, Y.M., 2008. Subset Selection for Vector Autoregressive Processes Using Lasso. Computational Statistics & Data Analysis 52, 3645–3657.

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

Kock, A.B., Callot, L., 2015. Oracle Inequalities for High Dimensional Vector Autoregressions. Journal of Econometrics 186, 325–344.

Koop, G., Korobilis, D., 2016. Model Uncertainty in Panel Vector Autoregressive Models. European Economic Review 81, 115–131.

Koop, G., Korobilis, D., 2018. Forecasting with High-Dimensional Panel VARs. Essex Finance Centre Working Papers 21329. University of Essex, Essex Business School.

Korobilis, D., 2016. Prior Selection for Panel Vector Autoregressions. Computational Statistics and Data Analysis 101, 110–120.

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.

Litterman, R., 1986. Forecasting with Bayesian Vector Autoregressions-Five Years of Experience. Journal of Business & Economic Statistics 4, 25–38.

Marcellino, M., Stock, J.H., Watson, M.W., 2006. A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series. Journal of Econometrics 135, 499–526.

Ngueyep, R., Serban, N., 2015. Large-Vector Autoregression for Multilayer Spatially Correlated Time Series. Technometrics 57, 207–216.

Nicholson, W.B., Matteson, D.S., Bien, J., 2016. High Dimensional Forecasting via Interpretable Vector Autoregression. ArXiv e-prints `arXiv:1412.5250`.

Nicholson, W.B., Matteson, D.S., Bien, J., 2017. VARX-L: Structured Regularization for Large Vector Autoregressions with Exogenous Variables. International Journal of Forecasting 33, 627–651.

Pesaran, M.H., Schuermann, T., Smith, L.V., 2009. Forecasting Economic and Financial Variables with Global VARs. International Journal of Forecasting 25, 642–675.

Ren, Y., Zhang, X., 2010. Subset Selection for Vector Autoregressive Processes via Adaptive Lasso. Statistics & Probability Letters 80, 1705–1712.

Song, S., Bickel, P.J., 2011. Large Vector Auto Regressions. ArXiv e-prints `arXiv:1106.3915`.

Stock, J.H., Watson, M.W., 2012. Generalized Shrinkage Methods for Forecasting Using Many Predictors. Journal of Business & Economic Statistics 30, 481–493.

Tibshirani, R., 1996. Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Socienty, Series B: Statistical Methodology 58(1), 267–288.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., Knight, K., 2005. Sparsity and smoothness via the fused lasso. Journal of the Royal Statistical Society. Series B (Statistical Methodology) 67.1, 91–108.

Zou, H., 2006. The Adaptive Lasso and Its Oracle Properties. Journal of the American Statistical Association 101(476), 1418–1429.

A The LASSO Estimator for PVAR Models

In order to allow for shrinkage towards homogeneous coefficient, define $\underline{B} = B - \bar{B}$ and rewrite (3) as

$$\begin{aligned}\underset{\underline{B}}{\text{argmin}} \;\; & \frac{1}{T} tr\left[(Y - \underline{B}X - \bar{B}X)'\Omega(Y - \underline{B}X - \bar{B}X)\right] + \Lambda\Phi(\underline{B}) \\ = & \frac{1}{T} tr\left[(\underline{Y} - \underline{B}X)'\Omega(\underline{Y} - \underline{B}X)\right] + \Lambda\Phi(\underline{B})\end{aligned}$$

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

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

$$\hat{b}^{ij}_{klp} = sign(\tilde{b}^{ij}_{klp}) \left(\left| \tilde{b}^{ij}_{klp} \right| - \frac{\lambda^{ij}_{kp} T}{2\omega_{kl} \sum_l^G \sum_p^P X^j_{lp,t} X^j_{lp,t}} \right) + \bar{b}_{klp} \tag{8}$$

with

$$\tilde{b}^{ij}_{klp} = \sum_t^T \sum_p^P \sum_{i,j}^N \sum_l^G \frac{\sum_{g \neq k}^G \omega^{ij}_{gl} \left(\underline{Y}^i_{g,t} - b^{ij}_{glp} X^j_{lp,t} \right) X^j_{lp,t}}{\omega_{kl} X^j_{lp,t} X^j_{lp,t}} + \frac{\left(\underline{Y}^i_{k,t} - \sum_{g \neq l}^G b^{ij}_{kgp} X^j_{gp,t} \right) X^j_{lp,t}}{X^j_{lp,t} X^j_{lp,t}} \tag{9}$$

and

$$\lambda^{ij}_{kp} = \begin{cases} \gamma & \text{if} \quad i = j, p = 1 \\ \lambda_k p^\alpha & \text{if} \quad i = j, p = 2, ..., P \;\; \text{and} \;\; k = 1, ..., NG \\ \lambda_k p^\alpha c & \text{if} \quad i \neq j, p = 1, ..., P \;\; \text{and} \;\; k = 1, ..., NG. \end{cases} \tag{10}$$

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

B Estimation of the Covariance Matrix

The covariance matrix is estimated using a graphical LASSO (GLASSO) approach. 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.

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 homogeneous coefficients, $\bar{B}$, are computed with a mean-group or averaged estimator. The optimal penalty parameters are determined via a cross-validation technique minimizing MSFEs. The search of the optimal penalty parameters is done over grids of values. The algorithm updates every element b_{klp}^{ij}. The following steps are repeated until convergence is archived. Update b_{klp}^{ij} as follows:

1. Calculate $\tilde{b}_{klp}^{ij}$ according to (9).
2. Set the penalty parameter λ_{kp}^{ij} equal to the optimal chosen.
3. Calculate $\tilde{\lambda}_{kp}^{ij} = \frac{\lambda_{kp}^{ij} T}{2\omega_{kl} \sum_l^G \sum_p^P X_{lp,t}^j X_{lp,t}^j}$
4. Calculate $\hat{b}_{klp}^{ij}$ as

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

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

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

D Simulation

D.1 Simulation Results

Table 5: Mean squared forecast errors relative to OLS for $T = 500$

	LASSO techniques					single unit	
		λ_k, c, α	$\lambda_k, c, \alpha, \gamma$	λ_k	*bl-diag*		*LASSO*
	λ_k, c, α	$\Sigma = I$	*(AG)*	$\Sigma = I$	*OLS*	*OLS*	λ_k
One-step ahead mean squared forecast errors							
DGP1	0.92^{***}	0.92^{***}	0.92^{***}	0.92^{***}	0.92^{***}	0.92^{***}	0.92^{***}
DGP2	0.92^{***}	0.92^{***}	0.92^{***}	0.92^{***}	0.92^{***}	0.92^{***}	0.92^{***}
DGP3	0.91^{***}	0.91^{***}	0.91^{***}	0.91^{***}	0.91^{***}	0.91^{***}	0.91^{***}
DGP4	0.92^{***}	0.92^{***}	0.92^{***}	0.92^{***}	0.92^{***}	0.93^{***}	0.91^{***}
Two-steps ahead mean squared forecast errors							
DGP1	0.93^{***}	0.93^{***}	0.93^{***}	0.93^{***}	0.93^{***}	0.93^{***}	0.93^{***}
DGP2	0.93^{***}	0.93^{***}	0.93^{***}	0.93^{***}	0.93^{***}	0.93^{***}	0.93^{***}
DGP3	0.93^{***}	0.93^{***}	0.93^{***}	0.93^{***}	0.93^{***}	0.93^{***}	0.93^{***}
DGP4	0.93^{***}	0.94^{***}	0.94^{***}	0.94^{***}	0.94^{***}	0.94^{***}	0.93^{***}
Six-steps ahead mean squared forecast errors							
DGP1	1^{**}	1^{**}	1^{**}	1^{**}	1^{**}	1^{**}	1^{**}
DGP2	1^{**}	1^{**}	1^{**}	1^{**}	1^{**}	1^{**}	1^{**}
DGP3	1^{**}	1^{**}	1^{**}	1^{**}	1^{**}	1^{**}	1^{**}
DGP4	1^{**}	1^{**}	1^{**}	1^{**}	1^{**}	1^{**}	1^{**}

NOTE: MSFEs are relative to a PVAR model estimated by OLS and average over all t, all units and variables and over 1000 Monte Carlo replications for $T = 500$. Significance level of Diebold Mariano test is indicated by $***$ (1%), $**$ (5%), and $*$ (10%). Estimators (in columns): LASSO with penalties λ_k, α, c and weighted sum of squared residuals, LASSO with the same penalty parameters where the covariance matrix is set to the identity matrix, LASSO with penalties $\lambda_k, \alpha, c, \gamma$ with aggregated estimator for homogeneous coefficients and with weighted sum of squared residuals, LASSO λ_k where the covariance matrix is set to the identity matrix, block-diagonal system ordering the variables in unit blocks estimated with OLS, separate VAR models for each unit estimated with OLS and with a LASSO with equation-specific shrinkage.

Table 6: Mean squared forecast errors relative the single unit VAR for $N = 5$ and $G = 5$

	LASSO λ_k, c, α $\Sigma = I$	*LASSO* λ_k $\Sigma = I$	*bl-diag* *OLS*
One-step ahead mean squared forecast errors			
DGP1	0.77^{***}	0.78^{***}	1.1^{***}
DGP2	0.77^{***}	0.77^{***}	1.1^{***}
Two-steps ahead mean squared forecast errors			
DGP1	0.82^{**}	0.83^{**}	1.13^{**}
DGP2	0.81^{**}	0.81^{**}	1.15^{**}
Six-steps ahead mean squared forecast errors			
DGP1	0.96	0.97	1.25
DGP2	0.94	0.94	1.45

NOTE: MSFEs are relative to a single unit VAR model estimated by OLS and average over all t, all units and variables and over 1000 Monte Carlo replications for $T = 100$. Significance level of Diebold Mariano test is indicated by $***$ (1%), $**$ (5%), and $*$ (10%). Estimators (in columns): LASSO with penalties λ_k, α, c where the covariance matrix is set to the identity matrix, LASSO λ_k where the covariance matrix is set to the identity matrix, block-diagonal system ordering the variables in unit blocks estimated with OLS.

Table 7: Performance evaluation of estimators for $T = 500$

	LASSO techniques					single unit		
		λ_k, c, α	$\lambda_k, c, \alpha, \gamma$	λ_k	*bl-diag*		*LASSO*	*PVAR*
	λ_k, c, α	$\Sigma = I$	*(AG)*	$\Sigma = I$	*OLS*	*OLS*	λ_k	*OLS*
Correct sparsity pattern								
DGP1	0.20	0.21	0.18	0.20	0.05	0.05	0.13	0.85
DGP2	0.53	0.53	0.51	0.54	0.45	0.45	0.53	0.45
DGP3	0.54	0.54	0.51	0.54	0.45	0.45	0.53	0.45
DGP4	0.53	0.53	0.51	0.54	0.45	0.45	0.53	0.45
Fraction of relevant variables included								
DGP1	0.09	0.10	0.31	0.07	1.00	1.00	0.23	1.00
DGP2	0.08	0.10	0.14	0.08	0.27	0.27	0.06	1.00
DGP3	0.09	0.09	0.15	0.08	0.27	0.27	0.06	1.00
DGP4	0.09	0.10	0.15	0.08	0.27	0.27	0.06	1.00
Number of variables included								
DGP1	31.10	36.21	49.32	28.78	80.00	80.00	18.53	400.00
DGP2	29.73	35.66	48.00	28.52	80.00	80.00	18.33	400.00
DGP3	34.31	35.85	51.73	30.49	80.00	80.00	18.34	400.00
DGP4	31.08	34.99	48.50	28.76	80.00	80.00	18.37	400.00
Mean squared errors								
DGP1	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
DGP2	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
DGP3	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
DGP4	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05

NOTE: All measures are averaged over 1000 Monte Carlo replications for $T = 500$. Estimators (in columns): LASSO with penalties λ_k, α, c and weighted sum of squared residuals, LASSO with the same penalty parameters where the covariance matrix is set to the identity matrix, LASSO with penalties $\lambda_k, \alpha, c, \gamma$ with aggregated estimator for homogeneous coefficients and with weighted sum of squared residuals, LASSO λ_k where the covariance matrix is set to the identity matrix, block-diagonal system ordering the variables in unit blocks estimated with OLS, separate VAR models for each unit estimated with OLS, with a LASSO with equation-specific shrinkage, and a PVAR estimated with OLS.

Table 8: Performance evaluation of estimators for $N = 5$ and $G = 5$

	LASSO λ_k, c, α $\Sigma = I$	*LASSO* λ_k $\Sigma = I$	*bl-diag* *OLS*	*single unit* *OLS*
Correct sparsity pattern				
DGP1	0.14	0.16	0.08	0.08
DGP2	0.52	0.51	0.48	0.48
Fraction of relevant variables included				
DGP1	0.06	0.05	1.00	1.00
DGP2	0.02	0.03	0.23	0.23
Number of variables included				
DGP1	77.71	121.56	500.00	500.00
DGP2	49.62	69.15	500.00	500.00
Mean squared error				
DGP1	0.02	0.02	0.03	0.02
DGP2	0.04	0.04	0.07	0.06

NOTE: All measures are averaged over 1000 Monte Carlo replications for $T = 100$. Estimators (in columns): LASSO with penalties λ_k, α, c where the covariance matrix is set to the identity matrix, LASSO λ_k where the covariance matrix is set to the identity matrix, block-diagonal system ordering the variables in unit blocks estimated with OLS, separate VAR models for each unit estimated with OLS.

E Forecasting Application

E.1 Bayesian Benchmark Models

Stochastic search specification selection. Koop and Korobilis (2016) 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 set to be

Bernoulli distributed, a parameter is drawn from the first or second part of the distribution. Koop and Korobilis (2016) 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 diagonal coefficient matrices of the B matrix. The prior has the following form:

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

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 (2016) 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. Canova and Ciccarelli (2004, 2009) factorize the parameters 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.

E.2 Penalty Parameters

For the empirical application a grid of twelve values is chosen for

$$\lambda_{k,grid} = [0.6, 0.55, 0.49, 0.44, 0.39, 0.33, 0.28, 0.22, 0.17, 0.12, 0.06, 0.01].$$

The grid for α, c, and γ are four-fold: $a_{grid} = [0.2, 0.4, 0.6, 0.8]$, $c_{grid} = [1.2, 1.4, 1.6, 1.8]$, and $\gamma_{grid} = [0.2, 0.4, 0.6, 0.8]$. The grid for ρ is $\rho_{grid} = [0, 0.1, 0.2, 0.3, 0.4]$.

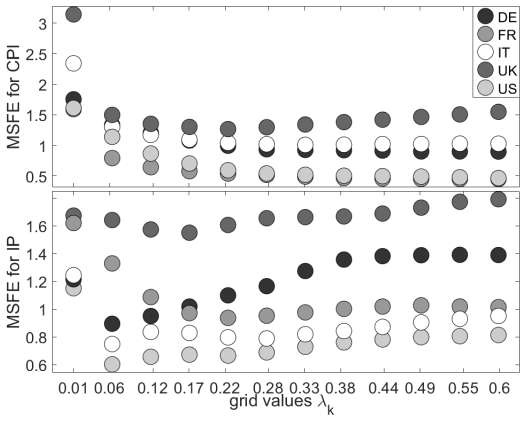
Table 9: Choice of penalty parameters

	LASSO techniques			
	λ_k, c, α	λ_k, c, α $\Sigma = I$	$\lambda_k, c, \alpha, \gamma$ *(MG)*	$\lambda_k, c, \alpha, \gamma$ *(AG)*
α	0.4	0.4	0.6	0.2
c	1.4	1.4	1.4	1.8
γ			0.2	0.2

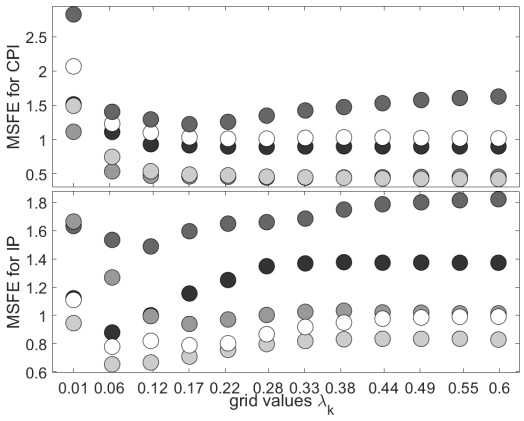
NOTE: Estimators (in columns): LASSO with penalties λ_k, α, c and weighted sum of squared residuals, LASSO with the same penalty parameters where the covariance matrix is set to the identity matrix, LASSO with penalties $\lambda_k, \alpha, c, \gamma$ with mean-group estimator for homogeneous coefficients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

Figure 4: Average MSFE for different values for λ_k per equation and fixed c, α and γ as in table 9

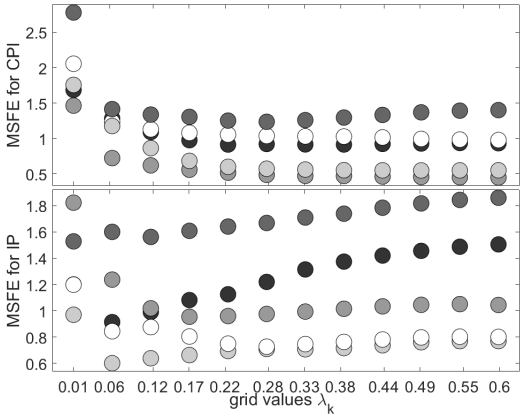
(a) LASSO λ_k, c, α



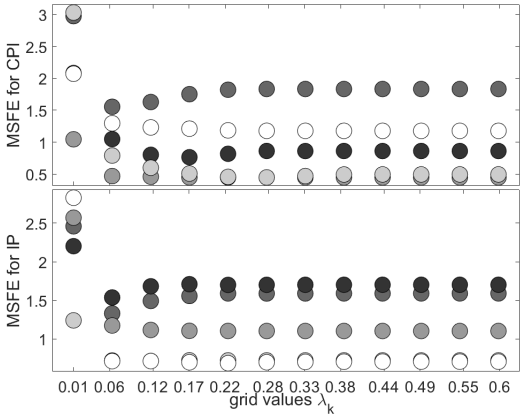
(b) LASSO $\lambda_k, c, \alpha, \Sigma = I$



(c) LASSO $\lambda_k, c, \alpha, \gamma$ (MG)



(d) LASSO $\lambda_k, c, \alpha, \gamma$ (AG)



NOTE: One-step ahead MSFE (averaged over all t in the penalty selection period) are given for each equation in the PVAR model for a grid of λ values. The first graph shows the MSFEs for CPI for the five countries and the second for IP. The λ with the minimal MSFE is chosen for each equation.

E.3 Application Results

Table 10: Mean squared forecast error of LASSO techniques relative to PVAR model estimated by OLS

		LASSO techniques						
		λ_k, c, α	λ_k, c, α	$\lambda_k, c, \alpha, \gamma$	$\lambda_k, c, \alpha, \gamma$		λ_k	
		λ_k, c, α	$\Sigma = I$	*(MG)*	*(AG)*	λ_k	$\Sigma = I$	*adaptive*
h=2	CPI	0.73^{**}	0.72^{**}	0.78^{**}	0.73^{***}	0.73^{**}	0.73^{**}	0.74^{**}
	IP	0.56^{***}	0.54^{***}	0.57^{***}	0.5^{***}	0.58^{***}	0.6^{***}	0.55^{***}
h=4	CPI	0.74^{*}	0.73^{**}	0.8^{*}	0.73^{**}	0.75^{*}	0.74^{**}	0.75^{*}
	IP	0.59^{***}	0.57^{***}	0.57^{***}	0.52^{***}	0.62^{***}	0.64^{***}	0.59^{***}
h=6	CPI	0.8^{*}	0.79^{*}	0.83	0.78^{*}	0.83	0.81^{*}	0.82
	IP	0.64^{**}	0.62^{**}	0.61^{**}	0.59^{**}	0.67^{**}	0.69^{**}	0.65^{**}
h=12	CPI	0.75	0.74	0.75	0.74	0.77	0.75	0.76
	IP	0.83^{*}	0.83	0.82	0.83	0.84^{*}	0.84^{*}	0.84

NOTE: The forecast period is from 2011:7 to 2016:6. MSFEs are averaged over all t and are relative to a PVAR model estimated by OLS, MSFEs smaller than one indicate better performance relative to OLS. Significance level of Diebold Mariano test is indicated by $***$ (1%), $**$ (5%), and $*$ (10%). Estimators (in columns): LASSO with penalties λ_k, α, c and weighted sum of squared residuals, LASSO with the same penalty parameters where the covarinace matrix is set to the identity matrix, LASSO with penalties $\lambda_k, \alpha, c, \gamma$ with mean-group estimator for homogenous coefficients and with weighted sum of squared residuals, same LASSO with aggregated estimator for homogenous coefficients, LASSO with weighted sum of squared residuals and λ_k, LASSO λ_k where the covarinace matrix is set to the identity matrix, and adaptive LASSO.

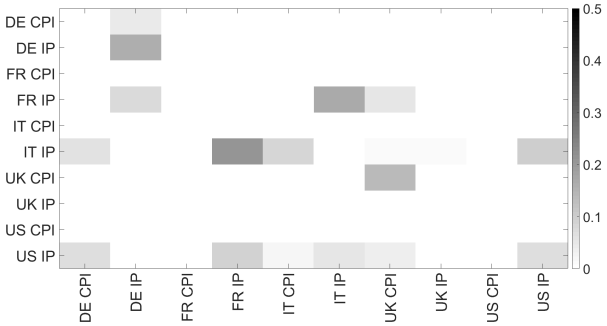
Table 11: Mean squared forecast error relative to PVAR model estimated by OLS

			Bayesian methods			single unit	
		LASSO			*bl-diag*		*LASSO*
		λ_k, c, α	*SSSS*	*CC*	*OLS*	*OLS*	λ_k
h=2	CPI	0.73**	1.8**	0.69***	0.69***	0.74***	0.71***
	IP	0.56***	2.08***	0.51***	0.54***	0.58***	0.51***
h=4	CPI	0.74*	1.97**	0.64**	0.68**	0.73**	0.71**
	IP	0.59***	2.81**	0.53***	0.56***	0.61***	0.53***
h=6	CPI	0.8*	2.53**	0.68*	0.76*	0.8*	0.77*
	IP	0.64**	2.89**	0.61**	0.61**	0.67**	0.6**
h=12	CPI	0.75	3.48	0.67	0.73	0.76	0.74
	IP	0.83*	7.45	0.86*	0.82	0.80	0.82

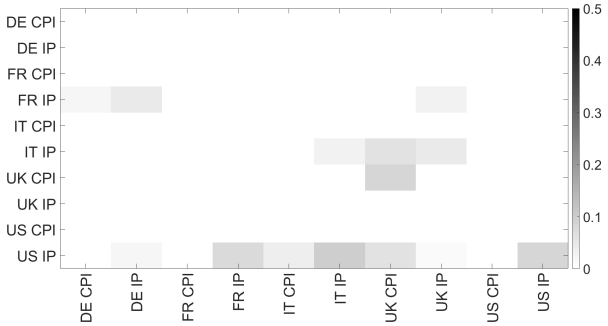
NOTE: The forecast period is from 2011:7 to 2016:6. MSFEs are averaged over all t and are relative to a PVAR model estimated by OLS, MSFEs smaller than one indicate better performance relative to OLS. Significance level of Diebold Mariano test is indicated by $***$ (1%), $**$ (5%), and $*$ (10%). Estimators (in columns): LASSO with penalties λ_k, α, c and weighted sum of squared residuals, stochastic search specification selection of Koop and Korobilis (2016), cross-sectional shrinkage approach of Canova and Ciccarelli (2004, 2009), block-diagonal system ordering the variables in unit blocks estimated with OLS, separate VAR models for each unit estimated with OLS and with a LASSO with equation-specific shrinkage.

Figure 5: Sparsity pattern of the coefficient matrix for model (1): lag 3 to 6.

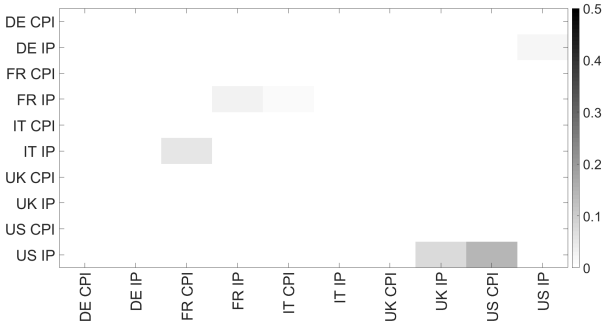
(a) Lag 3



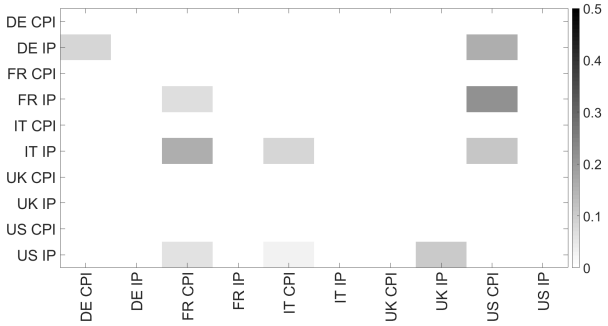
(b) Lag 4



(c) Lag 5



(d) Lag 6



NOTE: LASSO with penalties λ_k, α, c and weighted sum of squared residuals estimates for the third to sixth lags of all variables. Zero coefficients are colored in white and non-zero coefficients in gray shades whereby negative values are multiplied by -1 and the darkest shade is given to the highest coefficient value.