An alternative procedure allowing an estimation and forecasting VAR models in large data environment is a lasso shrinkage.

The standard OLS -estimation of a VAR model in a low-dimensional space may be written as follows

$$\min_{\Phi_{const},\Phi_j} ||y_t - \Phi_{const} - \Phi_1 y_{t-1} - \Phi_2 y_{t-2} - \ldots - \Phi_p y_{t-p}||_2^2 \tag{1}$$

Lasso shrinkage means taking into account a special penalty function of a certain form:

$$\min_{\Phi_{const},\Phi_j} ||y_t - \Phi_{const} - \Phi_1 y_{t-1} - \Phi_2 y_{t-2} - \ldots - \Phi_p y_{t-p}||_2^2 + \lambda \left(P_y(\Phi) + P_c(\Phi_{const})\right), \tag{2}$$

where λ is a penalty parameter which is selected following a cross-validation procedure, P_y is a penalty function on endogenous coefficients. As for all model coefficients the same penalty parameter is used, all the series under consideration should be on the same scale. Contrary to the standard BVAR which can be estimated with original raw data, the lasso estimation requires the normalization of the series. We use three penalty functions that yuild group shrinkage (group lasso) proposed by (Yuan and Lin, 2006) for VARX models (see also (Nicholson, Matteson, and Bien, 2017)). Three modifications of group lasso penalty functions that are implemented in the study are lag penalty, own/other penalty and sparse lag penalty functions:

Lag penalty function:

$$P_y(\Phi) = m\left(||\Phi_1||_F + \ldots + ||\Phi_p||_F\right) \tag{3}$$

$$P_x(\Phi_{const}) = \sqrt{m}\left(||\Phi_{const,1}||_F + \ldots + ||\Phi_{const,m}||_F\right) \tag{4}$$

This penalty functions splits all the coefficients into $p+1$ groups where p groups consist of lag matrices coefficients $(\Phi_1, \ldots, \Phi_p)$ and one group is a vector of constants forming the matrix Φ_{const}. The lasso shrinkage with lag penalty function results in a special form of sparsity: some lag matrices are active (i.e their coefficient estimates are not zeros) and others are passive (i.e. all coefficient estimates are zeros) but the sparsity within a group is not allowed.

The lag penalty function weights all the coefficients of the lag matrices symmetrically. In fact sometimes it is logical to assume that the coefficients on the main diagonal of a lag matrix are non-zeros even all other coefficients in the same matrix are zeros. It is the case because a variable is more likely affected by its own lags than by the lags of other variables in a set. This is the same idea that is incorporated in a traditional Minnesota prior for BVARs. To take this asymmetry into account the own/other group penalty function can be used.

Own/Other penalty function:

$$P_y(\Phi) = \sqrt{m}\left(||\Phi_1^{on}||_F + \ldots + ||\Phi_p^{on}||_F\right) + \sqrt{m(m-1)}\left(||\Phi_1^{off}||_F + \ldots + ||\Phi_p^{off}||_F\right) \tag{5}$$

$$P_x(\Phi_{const}) = \sqrt{m}\left(||\Phi_{const,1}||_F + \ldots + ||\Phi_{const,m}||_F\right) \tag{6}$$

The estimation with this penalty function allows not only some separated groups to contain non-zero elements but also groups of lag matrices coefficients with all zero elements besides the elements on the main diagonal.

In some cases the restrictions imposed by lag group and own/other lasso penalty functions can be too strict. If just one element in the group is non-zero, it is not sensible to require that all other members to be non-zeros as well. Simon et al.(2013) propose a sparse group penalty function that allows that just some elements in an active group are non-zeros. In this case a penalty function may be defined as: *Sparse Lag penalty function*:

$$P_y(\Phi) = (1-\alpha)m\left(||\Phi_1||_F + \ldots + ||\Phi_p||_F\right) + \alpha||\Phi||_1 \tag{7}$$

$$P_x(\Phi_{const}) = (1-\alpha)\sqrt{m}\left(||\Phi_{const,1}||_F + \ldots + ||\Phi_{const,m}||_F\right) + \alpha||\Phi_{const}||_1, \tag{8}$$

where $0 \leq \alpha \leq 1$ is an additional parameter that controls within-group sparsity. Following (Nicholson, Matteson, and Bien, 2017) we set $\alpha = \frac{1}{k+1}$