Generalized impulse response functions for high dimensional multivariate time series - a tool for macroeconomic analysis

K.G. van Wiggen
VU Student Number: 2675096

supervisor: prof. Q. Wiersma
co-reader: prof. A. Borowska

August 27, 2024

Abstract

In this paper the matrix autoregressive model will be examined on its performance to construct generalized impulse response functions in comparison with various vector autoregressive models in multiple settings where the curse of dimensionality is pronounced. It turned out that the matrix autoregressive model is more robust against the curse of dimensionality. The matrix autoregressive model exhibits particularly strong performance when the autoregressive coefficient to be estimated possesses a Kronecker structure. However, its main drawback is that it generates too weak responses.

Key words: VAR, MAR, gIRF, Ridge regression, Alasso, curse of dimensionality.

Contents

1 Introduction

In the field of econometrics there has been growing interest in modeling high dimensional multivariate time series. Policymakers, investors, or any stakeholder interested in international macroeconomic interactions will be concerned with a macroeconometric model capable of accurately representing multinational relationships. Impulse response functions (IRF) compute the effect of a shock in one or more variables, on other variables over time (Montes-Rojas, 2022). IRFs are suitable to construct a macroeconometric model because it shows the potential range of outcomes for multiple (international) economic indicators followed by a shock in another variable. In most studies, to construct IRFs, some variation of a vector autoregressive model (VAR) is used where the time points are treated as vector. To give some examples among many applications of IRFs in this context, Hafidh (2021) used this approach to investigate the relationship between Islamic banking variables and policy shocks in India, Dey and Tareque (2020) constructed IRFs to illustrate the dynamic effects of external debt shocks and the shocks of the macro-economic variables on economic growth in Bangladesh, and Zhang et al. (2022b) used this procedure to model the impact of oil price uncertainty on China's macro-economy.

A consequential drawback of VAR models is that the inter-relationship between the multivariate time series cannot be modelled. This leads to the loss of the underlying structure of the time series which significantly impairs the interpretability of VAR models (Chen et al., 2021). Imagine some data of multiple economic indicators for multiple countries. It is reasonable that the same indicators of different countries and different indicators of the same country form a strong relationship. By vectorizing the matrix-shaped input, the matrix-shaped structure will be lost. Therefore, it would only be possible to estimate the effect among the indicators, or among the countries. Trying to do both makes it hard to interpret the estimated coefficients. Another downside of vectorizing matrix time series, is the large amount of estimated coefficients in these models. This could lead to the curse of dimensionality, meaning that there are excessively model outputs compared to data observations, which weakens the model performance (Hammer, 1962). Therefore, Chen et al. (2021) consider modelling their matrix-shaped time series with the matrix autoregressive model (MAR). With this model the matrix-shaped structure with its corresponding interpretation stays intact, and the number of estimated coefficients reduces significantly.

Another useful approach to model high dimensional data involves models that incorporate variable selection. Penalized VAR models consider variable selection which is especially useful when the true underlying model has a sparse representation. This means that the model only takes a small number of relevant variables into account, even if many potential variables are available (Zou, 2006). Therefore, a penalized VAR model should hold the oracle properties, meaning that it should identify the right subset model of significant variables, while it has the optimal es-

timation rate. Tibshirani (1996) proposed a method for variable selection, the least absolute shrinkage and selection operator (Lasso), which screens high-dimensional data and sets irrelevant coefficients to zero by using a penalty component. Unfortunately, the Lasso estimator is only consistent under certain conditions and does not have the oracle properties. Zou (2006) introduced an adjustment to the Lasso, the adaptive Lasso (Alasso), which uses weights that differ per variable. This method is consistent and enjoys the oracle properties. A comparable method to the Alasso is the Ridge regression. This regression method proposed by Hoerl and Kennard (1970) is an alternative least squares regression technique that penalizes regression coefficients. The Ridge regression shrinks insignificant variable coefficients towards zero, but not exactly zero.

Although the MAR and Alasso are great approaches to deal with multivariate high dimensional time series, most articles that consider these models are focused on proving predictive power (see Zhang et al. (2022a), Zou (2006) and Chen et al. (2021) for example). In this study the VAR, MAR, Alasso and Ridge regression will be compared to each other regarding their performance of accurately computing IRFs with high dimensional multivariate time series. This study will be unique in the sense that, instead of proving predictive power, these methods will be used to compute IRFs. Furthermore, this study is noteworthy because, whereas most existing literature computes IRFs using various VAR models, this research compares the performance of the VAR with the MAR for this purpose. Furthermore, since the Ridge regression performs almost similar to the Alasso, it will be used to compare with the Alasso. This comparison will tell whether the strict method of the Alasso, by setting invariant variables to zero, contributes to generating gIRFs. However, the Ridge regression estimates as much coefficients as the VAR and will therefore suffer from the curse of dimensionality. Consequently, the Ridge regression will not be a solution to the VAR, but will solely be used to compare with the Alasso.

The results of this research could be used to enable more effective estimation of the impact of a macroeconomic shock on various related international macroeconomic indicators, as models with higher dimensions can be utilized for these estimations. This is relevant for any stakeholder regarding macro-economy because in economic theory, economic indicators are highly related to a lot of other (international) economic indicators (Pesaran et al., 2004). This research might give access to an econometric model that concerns these economic relations properly.

This study will be carried out by doing a simulation study where the VAR, MAR, Ridge and Alasso will be compared on their performance of accurately generating gIRFs in the setting of high dimensional multivariate time series. It will be organized according to the following structure. In Section 2 the models of this study will be explained. Thereafter, Section 3 provides the estimation techniques of these models. Furthermore, the simulation study will be set out in Section 4. The results of this simulation study will be illustrated in Section 5 after which they will be applied in Section 6 with an empirical application. Lastly, Section 7 will conclude this paper.

2 Models

Prior to going into the following sections, mention that in this study symbols will be typed with a bold typeface, and with a capital bold typeface if the symbol represents a vector or matrix, respectively. Otherwise, the symbol just refers to an individual element.

2.1 Impulse response functions for multivariate time series

As mentioned in Section 1, usually IRFs are constructed, using (a variation of) VAR models, to model the response of a variable due to a shock of another variable, over time. To construct an IRF, several assumptions must be made, potentially resulting in multiple complications. Firstly, Sims (1980) introduced the computation of orthogonal impulse responses for VAR models by using the Cholesky decomposition method to orthogonalize the underlying shocks. Unfortunately, this approach depends on the variable ordering in the VAR model, with little direction regarding the appropriate ordering. Therefore, there is a possibility for many potential results of the IRF (Wiesen and Beaumont, 2024). Another problem which had a lot of attention in the literature regards the possible adjustments on the VAR used to calculate the IRF. Researchers disagree on the identifying restrictions, which again leads too many possible results of the IRF (Inoue and Rossi, 2021; Lanne et al., 2017).

Therefore, Koop et al. (1996) and Pesaran and Shin (1998) introduced an order-independent IRF: the generalized impulse response function (gIRF). The gIRF deals with the structural complications by solving the identifiability problem by leaving specific identifying restrictions as ambiguous while constructing consistent IRFs. In other words, the gIRF let the nature of the data generating process depend on the data. Because of this advantage, the gIRF has been widely used in finacial and macro-economic research (see Barigozzi and Conti (2018) and Ewing (2002) for example).

The gIRF can be set out as follows: consider a stable VAR model, such as given in Equation 1

$$\mathbf{Y}_t = \boldsymbol{\alpha} + \boldsymbol{\phi}_1 \mathbf{Y}_{t-1} + ... + \boldsymbol{\phi}_P \mathbf{Y}_{t-p} + \boldsymbol{\epsilon}_t. \tag{1}$$

The difference of deriving the gIRF in contrast with other IRFs, can be noticed from the fact that $\boldsymbol{\epsilon}_t$ can be unrestricted (Wiesen and Beaumont, 2024). Furthermore, consider the vector moving avarage (VMA) of this VAR:

$$\mathbf{Y}_t = \mathbf{c} + \sum_{h=0}^{\infty} \mathbf{A}_h \boldsymbol{\epsilon}_{t-h}. \tag{2}$$

where $\mathbf{c}$ is the constant vector, and $\mathbf{A}_0, \mathbf{A}_1, ...,$ are the VMA coefficient matrices. To estimate the shocks h periods ahead, Equation 3 can be used. The gIRF uses a conditional expectation of the unrestricted shocks.

$$\mathbf{gIRF}(h, \delta_j, \mathbf{\Omega}_{t-1}) = E(\mathbf{Y}_{t+h}|\epsilon_{j,t} = \delta_j, \mathbf{\Omega}_{t-h}) - E(\mathbf{Y}_{t+h}|\mathbf{\Omega}_{t-h}). \tag{3}$$

Here, δ_j is the shock size of variable j and $\mathbf{\Omega}_{t-1}$ is the non-decreasing information set which denotes known history of the other variables. Assuming $\boldsymbol{\epsilon_t}$ has a multivariate normal distribution, Koop et al. (1996) and Pesaran and Shin (1998) showed that

$$E(\epsilon_t|\epsilon_{jt} = \delta_j) = \mathbf{\Sigma}_\epsilon \mathbf{e}_j \sigma_{jj}^{-1} \delta_j, \tag{4}$$

which can be used to rewrite Equation 3 as:

$$\mathbf{gIRF}(h, \delta_j, \mathbf{\Omega}_{t-1}) = \mathbf{A}_h \mathbf{\Sigma}_\epsilon \mathbf{e}_j \sigma_{jj}^{-1} \delta_j. \tag{5}$$

Here $\mathbf{A}_h$ from Equation 2 is used, and $\mathbf{\Sigma}_\epsilon$ is the covariance matrix of $\boldsymbol{\epsilon}$ also from Equation 2. Furthermore, $\mathbf{e}_j$ is the selection vector with a length corresponding to the amount of variables, consisting of all zeros and a one for the jth element. Lastly, σ_{jj} is the variance of shock j. With the gIRF representation of Equation 5 the response of Y due to a shock in $\epsilon_{j,t}$ with size δ_j, can be modeled h periods ahead.

Although the mentioned problems regarding estimating IRFs are solved using gIRFs, the mentioned drawbacks associated with estimating gIRFs using VAR models still apply with multivariate time series. The potential solutions to this problem will be illustrated: the MAR and the Alasso. Furthermore, the Ridge regression will be explained in more detail, as mentioned, to compare with the performance of the Alasso. At first, these models will be set out, after which their estimation will be excessively explained and compared to one another. Thereafter, these models will be used for the estimation of the gIRF.

2.2 The Matrix Autoregressive Model

Consider a three-dimensional time series with length T, where at each time point t, a matrix is observed, say $\mathbf{X}_t$, with dimensions $m \times n$. Using the mentioned example in Section 1, the rows (m) could display the countries and the columns (n) the economic indicators. If a VAR model would be used, the data would be vectorized, denoted by $\text{vec}(\cdot)$, by stacking the columns. This would result in the following VAR(1) model:

$$\text{vec}(\mathbf{X}_t) = \boldsymbol{\phi}\text{vec}(\mathbf{X}_{t-1}) + \text{vec}(\mathbf{E}_t). \tag{6}$$

The roles of m and n are mixed in parameter $\boldsymbol{\phi}$ of Equation 6. Therefore, $\boldsymbol{\phi}$ is difficult to interpret and it does not utilize the matrix structure with the prior knowledge of the correlations among the time series. To overcome this drawback, Chen et al. (2021) proposed the MAR(1):

$$\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1}\mathbf{B}^{'} + \mathbf{E}_t, \tag{7}$$

where $\mathbf{A}$ and $\mathbf{B}$ are $m \times m$ and $n \times n$ autoregressive coefficient matrices, respectively. Furthermore, $\mathbf{A}$ represents the row-wise interactions, while $\mathbf{B}$ reflects the column-wise dependence. Mention that, $\mathbf{A}$ and $\mathbf{B}$ in Equation 7 could be divided and multiplied by the same constant, without changing the model. To solve this identifiability issue, $\mathbf{A}$ should be normalized such that its Frobenius norm is one (Chen et al., 2021). Furthermore, $\mathbf{E}_t$ in Equation 7 is a $m \times n$ matrix at every time point t, that is assumed to be white noise. This means that $\mathbf{E}_t$ and $\mathbf{E}_s$ are uncorrelated for $t \neq s$, but may exhibit concurrent correlation among its own entries.

To model $\mathbf{E}_t$, consider the assumption of Equation 8 with $\boldsymbol{\Sigma_c}$ and $\boldsymbol{\Sigma_r}$ having dimensions $m \times m$ and $n \times n$, respectively.

$$\text{Cov}(\text{vec}(\mathbf{E}_t)) = \boldsymbol{\Sigma_c} \otimes \boldsymbol{\Sigma_r} \tag{8}$$

Here, $\boldsymbol{\Sigma_r}$ represents the covariances across the rows, and $\boldsymbol{\Sigma_c}$ across the columns. To solve an identifiability issue with $\text{vec}(\mathbf{E}_t)$ in Equation 9, the Frobenius norm of $\boldsymbol{\Sigma_r}$ should be one. This assumption will come in handy in Section 3.1.3.

Furthermore, Chen et al. (2021) mention that the MAR in Equation 7 can have the form of a VAR, via the following representation:

$$\text{vec}(\mathbf{X}_t) = (\mathbf{B} \otimes \mathbf{A})\text{vec}(\mathbf{X}_{t-1}) + \text{vec}(\mathbf{E}_t). \tag{9}$$

with $\otimes$ representing the Kronecker product. Regarding the mentioned drawback of the large amount of parameters in VAR models, it can already be seen that $\boldsymbol{\phi}$ in Equation 6 needs m^2n^2 coefficients, while Equation 9 only needs $m^2 + n^2$ coefficients to calculate the autoregressive coefficient.

2.3 Penalized VAR models

Presume a time series with length T, where at every t, vector $\mathbf{y}_t = (y_1, ..., y_d)$ is observed, consisting of $i = 1, ..., d$ dependent variables. Furthermore, consider $\mathbf{X}_t = [x_{11}, ..., x_{dp}]$ the independent variable matrix, for $j = 1, ..., p$ independent predictors. Furthermore, estimation vector $\hat{\boldsymbol{\beta}} = E[y_t|x_t] = \beta_{11}x_{11} + ... + \beta_{dp}x_{dp}$ denotes the estimated coefficient matrix of significant predictors by a certain procedure. Moreover, the true model has a sparse representation and $\mathcal{A} = \{j : \beta_j \neq 0\}$ represents the real subset of significant predictors. Then, according to Zou (2006), $\hat{\boldsymbol{\beta}}$ has the oracle properties if:

1. $\hat{\boldsymbol{\beta}}$ identifies the right subset model, $\{j : \hat{\beta}_j \neq 0\} = \mathcal{A}$
2. $\hat{\boldsymbol{\beta}}$ has the optimal estimation rate, $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{\mathcal{A}}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is the covariance matrix knowing $\mathcal{A}$.

A good procedure for variable selection should have these oracle properties (Fan and Li, 2001; Fan and Peng, 2004).

Therefore, Tibshirani (1996) introduced the Lasso estimator which is shown in Equation 10. The Lasso screens high-dimensional variables to improve the least squares estimation by adding a nonnegative penalty component that is dependent on the sample size T, denoted with λ. The second term in Equation 10 shrinks the coefficients to zero as λ increases, which improves the prediction accuracy and gives the Lasso its unique qualities (Zou, 2006).

$$\hat{\boldsymbol{\beta}}_{Lasso} = \min_{\beta}\{\frac{1}{2}\sum_{i=1}^{d}\left(\mathbf{y}_i - \mathbf{x}_i^T\boldsymbol{\beta}\right)^2 + \lambda\sum_{j=1}^{p}|\boldsymbol{\beta}_j|\} \tag{10}$$

However, Meinshausen and Bühlmann (2006) discovered that because of the sensitivity of the Lasso for correlation and multicollinearity, insignificant variables may be selected. Because of this unnecessary variables in the model, it is harder to fit the model accurately. Therefore, the Lasso estimator is unfortunately inconsistent and does not have the oracle properties. To overcome this, Zou (2006) introduced the Alasso, modeled in Equation 11.

$$\hat{\boldsymbol{\beta}}_{Alasso} = \min_{\beta}\{\frac{1}{2}\sum_{i=1}^{d}\left(\mathbf{y}_i - \mathbf{x}_i^T\boldsymbol{\beta}\right)^2 + \lambda\sum_{j=1}^{p}\mathbf{w}_j|\boldsymbol{\beta}_j|\} \tag{11}$$

With the Alasso, instead of forcing the coefficients to be equally penalized, different weights are assigned to various coefficients. The weight vector, consisting of these weights, can be defined with Equation 12, where $\gamma > 0$. Due to this adjustment to the Lasso estimator, the Alasso enjoys the oracle properties (Zou, 2006).

$$\hat{\mathbf{w}} = 1/|\hat{\boldsymbol{\beta}}|^{\gamma} \tag{12}$$

Another method to estimate $\hat{\boldsymbol{\beta}}$ has been suggested by Hoerl and Kennard (1970), who introduced the Ridge penalty to penalize the least squares estimation. Equation 13 presents the Ridge regression in mathematical form. The Ridge penalty is represented by the second term in Equation 13, where the Ridge regularization parameter λ should be positive.

$$\hat{\boldsymbol{\beta}}_{Ridge} = \min_{\beta}\{\frac{1}{2}\sum_{i=1}^{d}\left(\mathbf{y}_i - \mathbf{x}_i^T\boldsymbol{\beta}\right)^2 + \lambda\sum_{j=1}^{p}\boldsymbol{\beta}_j^2\} \tag{13}$$

Mention that, when $\lambda = 0$ the Lasso, Alasso and Ridge regression all simplify to ordinary least squares estimation. Furthermore, as stated in Section 1, the Lasso and Alasso penalize insignificant variables to zero, whereas the Ridge regression penalizes these variable to a small number. This can explained as follows. The Ridge regression penalizes coefficients with $\boldsymbol{\beta}_j^2$. The derivative of this penalty term with respect to any $\boldsymbol{\beta}_j$ can never be zero because $\boldsymbol{\beta}_j \neq 0$. On the other hand, the derivative of the penalty term of the Lasso and Alasso with respect to $\boldsymbol{\beta}_j$ can be zero because $|\boldsymbol{\beta}_j|$ is not continuous at zero. Because of this jump in $|\boldsymbol{\beta}_j|$, insignificant variables can be set to zero by Lasso and Alasso (Hastie et al., 2009).

3 Estimation

3.1 Estimation of the Matrix Autoregressive Model

Consider the MAR(1) of Equation 7. Coefficient matrices $\mathbf{A}$ and $\mathbf{B}$ can be estimated with three estimation techniques (Chen et al., 2021): the projection method (PROJ), iterated least squares (LSE) and maximum likelihood estimation (MLE). This techniques will consecutively be explained in more detail.

3.1.1 The projection method

The projection can be explained as follows. First, the MAR of Equation 7 should be reformulated as the VAR of Equation 9. Furthermore, $\hat{\boldsymbol{\phi}}$ of the VAR in Equation 6 should be estimated. Then, $\hat{\mathbf{A}}_{PROJ}$ and $\hat{\mathbf{B}}_{PROJ}$ can be found by a comparison of $\hat{\boldsymbol{\phi}}$ and $\mathbf{B} \otimes \mathbf{A}$ using the Nearest Kronecker Product of Equation 14. In this equation F indicates the Frobenius norm (Van Loan and Pitsianis, 1993; Van Loan, 2000).

$$(\hat{\mathbf{A}}_{PROJ}, \hat{\mathbf{B}}_{PROJ}) = \arg\min_{\mathbf{A},\mathbf{B}} \|\hat{\boldsymbol{\phi}} - \mathbf{B} \otimes \mathbf{A}\|_F^2 \tag{14}$$

A solution to this minimization problem can be found through multiple calculations (Chen et al., 2021). In these calculations the feature is used that all entries in $\mathbf{B} \otimes \mathbf{A}$ are also in $\text{vec}(\mathbf{A})\text{vec}(\mathbf{B})'$, only the placement of the elements differ. Therefore, the re-arrangement operator of Equation 15 can be constructed.

$$\mathcal{G}(\mathbf{B} \otimes \mathbf{A}) = \text{vec}(\mathbf{A})\text{vec}(\mathbf{B})'. \tag{15}$$

This re-arrangement operator can subsequently be used in the calculations of Equation 16, where $\tilde{\boldsymbol{\phi}} = \mathcal{G}(\hat{\boldsymbol{\phi}})$.

$$
\begin{aligned}
&\min_{\mathbf{A},\mathbf{B}} \|\hat{\boldsymbol{\phi}} - \mathbf{B} \otimes \mathbf{A}\|_F^2 \\
&= \min_{\mathbf{A},\mathbf{B}} \|\mathcal{G}(\hat{\boldsymbol{\phi}}) - \mathcal{G}(\mathbf{B} \otimes \mathbf{A})\|_F^2 \\
&= \min_{\mathbf{A},\mathbf{B}} \|\mathcal{G}(\hat{\boldsymbol{\phi}}) - \text{vec}(\mathbf{A})\text{vec}(\mathbf{B})'\|_F^2 \\
&= \min_{\mathbf{A},\mathbf{B}} \|\tilde{\boldsymbol{\phi}} - \text{vec}(\mathbf{A})\text{vec}(\mathbf{B})'\|_F^2
\end{aligned}
\tag{16}
$$

To make these derivations, the insights are used that $\mathcal{G}$ is a linear operator, and that the Frobenius norm of a matrix is only dependent of the elements in the matrix (Chen et al., 2021). Therefore, if $\mathbf{C}_1$ and $\mathbf{C}_2$ are two constant vectors, $\mathcal{G}(\mathbf{C}_1 + \mathbf{C}_1) = \mathcal{G}(\mathbf{C}_1) + \mathcal{G}(\mathbf{C}_2)$ and $\|\mathcal{G}(\mathbf{C})\|_F = \|\mathbf{C}\|_F$. These calculations of $\tilde{\boldsymbol{\phi}}$ via the re-arrangement operator will be more mathematically explained.

Consider the estimated $\hat{\boldsymbol{\phi}}$ from Equation 6 with dimensions $(m*n)\times(m*n)$. This $\hat{\boldsymbol{\phi}}$ can be transformed in $\tilde{\boldsymbol{\phi}}$ by first dividing matrix $\hat{\boldsymbol{\phi}}$ into blocks with dimensions $m \times m$. Then, within these blocks, create two indicators, for the concerning row and column, and use these to transform the blocks in line with the re-arrangement operator of Equation 15. This will look as follows:

$$
\hat{\boldsymbol{\phi}} \rightarrow \begin{bmatrix} \hat{\phi}_{1,1} & \cdots & \hat{\phi}_{1,m*n} \\ \vdots & \ddots & \vdots \\ \hat{\phi}_{m*n,1} & \cdots & \hat{\phi}_{m*n,m*n} \end{bmatrix} \rightarrow
$$

$$
\left[\begin{array}{ccc|ccc|c|ccc}
\hat{\phi}_{1,1} & \cdots & \hat{\phi}_{1,m} & \hat{\phi}_{1,m+1} & \cdots & \hat{\phi}_{1,m*2} & & \hat{\phi}_{1,m*(n-1)+1} & \cdots & \hat{\phi}_{1,m*n} \\
\vdots & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \\
\hat{\phi}_{m,1} & \cdots & \hat{\phi}_{m,m} & \hat{\phi}_{m,m+1} & \cdots & \hat{\phi}_{m,m*2} & & \hat{\phi}_{m,m*(n-1)+1} & \cdots & \hat{\phi}_{m,m*n} \\
\hline
\hat{\phi}_{m+1,1} & \cdots & \hat{\phi}_{m+1,m} & \hat{\phi}_{m+1,m+1} & \cdots & \hat{\phi}_{m+1,m*2} & & \hat{\phi}_{m+1,m*(n-1)+1} & \cdots & \hat{\phi}_{m+1,m*n} \\
\vdots & & \vdots & \vdots & & \vdots & \cdots & \vdots & & \vdots \\
\hat{\phi}_{m*2,1} & \cdots & \hat{\phi}_{m*2,m} & \hat{\phi}_{m*2,m+1} & \cdots & \hat{\phi}_{m*2,m*2} & & \hat{\phi}_{m*2,m*(n-1)+1} & \cdots & \hat{\phi}_{m*2,m*n} \\
\hline
 & & & & \vdots & & & & & \\
\hline
\hat{\phi}_{m*(n-1)+1,1} & \cdots & \hat{\phi}_{m*(n-1)+1,m} & \hat{\phi}_{m*(n-1)+1,m+1} & \cdots & \hat{\phi}_{m*(n-1)+1,m*2} & & \hat{\phi}_{m\times(n-1)+1,m*(n-1)+1} & \cdots & \hat{\phi}_{m*(n-1)+1,m*n} \\
\vdots & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \\
\hat{\phi}_{m*n,1} & \cdots & \hat{\phi}_{m*n,m} & \hat{\phi}_{m*n,m+1} & \cdots & \hat{\phi}_{m*n,m*2} & & \hat{\phi}_{m*n,m*(n-1)+1} & \cdots & \hat{\phi}_{m*n,m*n}
\end{array}\right]
$$

$$
\rightarrow \tag{17}
$$

$$
\left[\begin{array}{ccccccccccc}
\hat{\phi}_{1,1} & \hat{\phi}_{m+1,1} & & \hat{\phi}_{m*(n-1)+1,1} & \hat{\phi}_{1,m+1} & \hat{\phi}_{m+1,m+1} & & \hat{\phi}_{m*(n-1)+1,m+1} & \hat{\phi}_{1,m*(n-1)+1} & \hat{\phi}_{m+1,m*(n-1)+1} & & \hat{\phi}_{m*(n-1)+1,m*(n-1)+1} \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\
\hat{\phi}_{m,1} & \hat{\phi}_{m*2,1} & & \hat{\phi}_{m*n,1} & \hat{\phi}_{m,m+1} & \hat{\phi}_{m*2,m+1} & & \hat{\phi}_{m*n,m+1} & \hat{\phi}_{m,m*(n-1)+1} & \hat{\phi}_{m*2,m*(n-1)+1} & & \hat{\phi}_{m*n,m*(n-1)+1} \\
\hat{\phi}_{1,2} & \hat{\phi}_{m+1,2} & & \hat{\phi}_{m*(n-1)+1,2} & \hat{\phi}_{1,m+2} & \hat{\phi}_{m+1,m+2} & & \hat{\phi}_{m*(n-1)+1,m+2} & \hat{\phi}_{1,m*(n-1)+2} & \hat{\phi}_{m+1,m*(n-1)+2} & & \hat{\phi}_{m*(n-1)+1,m*(n-1)+2} \\
\vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\
\hat{\phi}_{m,2} & \hat{\phi}_{m*2,2} & & \hat{\phi}_{m*n,2} & \hat{\phi}_{m,m+2} & \hat{\phi}_{m*2,m+2} & & \hat{\phi}_{m*n,m+2} & \hat{\phi}_{m,m*(n-1)+2} & \hat{\phi}_{m*2,m*(n-1)+2} & & \hat{\phi}_{m*n,m*(n-1)+2} \\
 & & & & & & & \vdots & & & & \\
\hat{\phi}_{1,m} & \hat{\phi}_{m+1,m} & & \hat{\phi}_{m*(n-1)+1,m} & \hat{\phi}_{1,m*2} & \hat{\phi}_{m+1,m*2} & & \hat{\phi}_{m*(n-1)+1,m*2} & \hat{\phi}_{1,m*n} & \hat{\phi}_{m+1,m*n} & & \hat{\phi}_{m*(n-1),m*n} \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\
\hat{\phi}_{m,m} & \hat{\phi}_{m*2,m} & & \hat{\phi}_{m*n,m} & \hat{\phi}_{m,m*2} & \hat{\phi}_{m*2,m*2} & & \hat{\phi}_{m*n,m*2} & \hat{\phi}_{m,m*n} & \hat{\phi}_{m*2,m*n} & & \hat{\phi}_{m*n,m*n}
\end{array}\right]
$$

$$\rightarrow \mathcal{G}(\hat{\boldsymbol{\phi}}) = \tilde{\boldsymbol{\phi}}$$

where $\tilde{\boldsymbol{\phi}}$ is of dimension $(m*m)\times(n*n)$. The re-arrangement operator will transform $\hat{\boldsymbol{\phi}}$ in this way for any combination of integers m and n.

After computing $\tilde{\boldsymbol{\phi}}$, its largest singular value, and its first left and right singular vectors, denoted by d_1, $\mathbf{u}_1$ and $\mathbf{v}_1$ can be found with the Singular Value Decomposition (SVD). Finally, these can be used in Equation 18 to find $\text{vec}(\hat{\mathbf{A}})\text{vec}(\hat{\mathbf{B}})$. By converting these vectors of Equation 18 into matrices, with the normalization of $\|\hat{\mathbf{A}}_{PROJ}\|_F$ to ensure stability, $\hat{\mathbf{A}}_{PROJ}$ and $\hat{\mathbf{B}}_{PROJ}$ of Equation 14 can be calculated.

$$\text{vec}(\hat{\mathbf{A}})\text{vec}(\hat{\mathbf{B}})^{'} = d_1\mathbf{u}_1\mathbf{v}_1^{'} \tag{18}$$

Unfortunately, the procedure of the projection method could be inaccurate for moderately large m and n and a finite sample size. Therefore, it will serve as an initial value for the iterated least squares and maximum likelihood estimation (Chen et al., 2021).

3.1.2 Iterated least squares estimation

The iterated least squares estimator matrices can be constructed by solving Equation 19. With this method the assumption is made that all entries in $\mathbf{E}_t$ follow a normal distribution with zero mean and constant variance. However, if the error covariance matrix is arbitrary, this method still results in reasonable estimator matrices (Chen et al., 2021).

$$(\hat{\mathbf{A}}_{LSE}, \hat{\mathbf{B}}_{LSE}) = \min_{\mathbf{A},\mathbf{B}} \sum_t \|\mathbf{X}_t - \mathbf{A}\mathbf{X}_{t-1}\mathbf{B}^{'}\|_F^2 \tag{19}$$

Equation 19 can be solved by first taking the derivatives with respect to $\mathbf{A}$ and $\mathbf{B}$, which are defined in Equation 20

$$\begin{aligned} \sum_t \mathbf{A}\mathbf{X}_{t-1}\mathbf{B}^{'}\mathbf{B}\mathbf{X}_{t-1}^{'} - \sum_t \mathbf{X}_t\mathbf{B}\mathbf{X}_{t-1}^{'} &= \mathbf{0} \\ \sum_t \mathbf{B}\mathbf{X}_{t-1}^{'}\mathbf{A}^{'}\mathbf{A}\mathbf{X}_{t-1} - \sum_t \mathbf{X}_t^{'}\mathbf{A}\mathbf{X}_{t-1} &= \mathbf{0} \end{aligned} \tag{20}$$

Thereafter, $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ should sequentially be iterated by updating either one of them within Equation 19 while maintaining the other one fixed. As mentioned, to start this iterations, $\hat{\mathbf{A}}_{PROJ}$ and $\hat{\mathbf{B}}_{PROJ}$ are used as initial values. Using the derivatives of Equation 20, the iterations of updating $\mathbf{A}$ given $\mathbf{B}$, and the other way around are shown by Equation 21.

$$\begin{aligned}\left(\sum_t \mathbf{X}_t\mathbf{B}\mathbf{X}_{t-1}'\right)\left(\sum_t \mathbf{X}_{t-1}\mathbf{B}'\mathbf{B}\mathbf{X}_{t-1}'\right)^{-1} &\to \mathbf{A}\\ \left(\sum_t \mathbf{X}_t'\mathbf{A}\mathbf{X}_{t-1}\right)\left(\sum_t \mathbf{X}_t'\mathbf{A}'\mathbf{A}\mathbf{X}_{t-1}\right)^{-1} &\to \mathbf{B}\end{aligned} \tag{21}$$

Furthermore, after each iteration $\mathbf{A}$ is normalized such that $\|\mathbf{A}\|_F = 1$ to ensure stability (Chen et al., 2021). When the iterations converge, $\mathbf{A}$ and $\mathbf{B}$ replicate the matrices that solve Equation 19, denoted with $\hat{\mathbf{A}}_{LSE}$ and $\hat{\mathbf{B}}_{LSE}$.

3.1.3 Maximum likelihood estimation

Lastly, the MAR can be estimated with the maximum likelihood estimation. Within this method the assumption that $\mathbf{E}_t$ has the structure of Equation 8, is used to calculate the estimator matrices. The estimates $\hat{\mathbf{A}}_{MLE}$ and $\hat{\mathbf{B}}_{MLE}$ can be found by maximizing the log likelihood of Equation 22.

$$\begin{aligned}&- m(T-1)\log|\boldsymbol{\Sigma_c}| - n(T-1)\log|\boldsymbol{\Sigma_r}|\\ &\qquad - \sum_t \operatorname{tr}(\boldsymbol{\Sigma_r}^{-1}(\mathbf{X}_t - \mathbf{A}\mathbf{X}_{t-1}\mathbf{B}')\boldsymbol{\Sigma_c}^{-1}(\mathbf{X}_t - \mathbf{A}\mathbf{X}_{t-1}\mathbf{B}')')\end{aligned} \tag{22}$$

In line with the estimation of $\hat{\mathbf{A}}_{LSE}$ and $\hat{\mathbf{B}}_{LSE}$, the estimates of $\hat{\mathbf{A}}_{MLE}$ and $\hat{\mathbf{B}}_{MLE}$ can be found by setting the derivatives of this equation equal to zero. These derivatives with respect to $\mathbf{A}$, $\mathbf{B}$, $\boldsymbol{\Sigma_c}$ and $\boldsymbol{\Sigma_r}$ are shown in Equation 23.

$$\begin{aligned}\mathbf{A}\sum_t \mathbf{X}_{t-1}\mathbf{B}'\boldsymbol{\Sigma_c}^{-1}\mathbf{B}\mathbf{X}_{t-1}' - \sum_t \mathbf{X}_t\boldsymbol{\Sigma_c}^{-1}\mathbf{B}\mathbf{X}_{t-1}' &= \mathbf{0}\\ \mathbf{B}\sum_t \mathbf{X}_{t-1}'\mathbf{A}'\boldsymbol{\Sigma_r}^{-1}\mathbf{A}\mathbf{X}_{t-1} - \sum_t \mathbf{X}_t'\boldsymbol{\Sigma_r}^{-1}\mathbf{A}\mathbf{X}_{t-1} &= \mathbf{0}\\ m(T-1)\boldsymbol{\Sigma_c} - \sum_t (\mathbf{X}_t - \mathbf{A}\mathbf{X}_{t-1}\mathbf{B}')'\boldsymbol{\Sigma_r}^{-1}(\mathbf{X}_t - \mathbf{A}\mathbf{X}_{t-1}\mathbf{B}') &= \mathbf{0}\\ n(T-1)\boldsymbol{\Sigma_r} - \sum_t (\mathbf{X}_t - \mathbf{A}\mathbf{X}_{t-1}\mathbf{B}')\boldsymbol{\Sigma_c}^{-1}(\mathbf{X}_t - \mathbf{A}\mathbf{X}_{t-1}\mathbf{B}')' &= \mathbf{0}\end{aligned} \tag{23}$$

The iterations of updating one of this estimators given that the others remain fixed are given in Equation 24. Here, $\mathbf{A}$ and $\boldsymbol{\Sigma_r}$ are normalized after every iteration such that $\|\mathbf{A}\|_F = 1$ and $\|\boldsymbol{\Sigma_r}\|_F = 1$ to ensure stability (Chen et al., 2021). Again, to start this iterations, $\hat{\mathbf{A}}_{PROJ}$ and $\hat{\mathbf{B}}_{PROJ}$ are used as initial values. Furthermore, to find initial values for $\boldsymbol{\Sigma_r}$ and $\boldsymbol{\Sigma_c}$, the covariance matrix of $\operatorname{vec}(\hat{\mathbf{E}}_\mathbf{t})$, denoted with $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{E}}$, resulting from $\hat{\mathbf{A}}_{PROJ}$ and $\hat{\mathbf{B}}_{PROJ}$ in Equation 9 will be calculated. Then, Equation 14 can be used to minimize the difference between $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{E}}$ and $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{c}} \otimes \hat{\boldsymbol{\Sigma}}_{\boldsymbol{r}}$. To do

this, first the rearrangement operator will be applied to calculate $\mathcal{G}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{E}})$, afterwhich $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{c}}$ and $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{r}}$ can be found with Equation 18.

$$\begin{gathered}
\left(\sum_t \mathbf{X}_t \boldsymbol{\Sigma}_{\boldsymbol{c}}^{-1} \mathbf{B} \mathbf{X}_{t-1}^{'}\right)\left(\sum_t \mathbf{X}_{t-1} \mathbf{B}^{'} \boldsymbol{\Sigma}_{\boldsymbol{c}}^{-1} \mathbf{B} \mathbf{X}_{t-1}^{'}\right)^{-1} \rightarrow \mathbf{A} \\
\left(\sum_t \mathbf{X}_t^{'} \boldsymbol{\Sigma}_{\boldsymbol{r}}^{-1} \mathbf{A} \mathbf{X}_{t-1}\right)\left(\sum_t \mathbf{X}_{t-1}^{'} \mathbf{A}^{'} \boldsymbol{\Sigma}_{\boldsymbol{r}}^{-1} \mathbf{A} \mathbf{X}_{t-1}\right)^{-1} \rightarrow \mathbf{B} \\
\mathbf{R}_t = \mathbf{X}_t - \mathbf{A} \mathbf{X}_{t-1} \mathbf{B}^{'} \\
\frac{\sum_t \mathbf{R}_t \boldsymbol{\Sigma}_{\boldsymbol{c}}^{-1} \mathbf{R}_t^{'}}{n(T-1)} \rightarrow \boldsymbol{\Sigma}_{\boldsymbol{r}} \\
\frac{\sum_t \mathbf{R}_t^{'} \boldsymbol{\Sigma}_{\boldsymbol{r}}^{-1} \mathbf{R}_t}{m(T-1)} \rightarrow \boldsymbol{\Sigma}_{\boldsymbol{c}}
\end{gathered} \tag{24}$$

If the iterations converge, $\mathbf{A}$ and $\mathbf{B}$ replicate the matrices that solve Equation 22, denoted by $\hat{\mathbf{A}}_{MLE}$ and $\hat{\mathbf{B}}_{MLE}$, under the assumption that the covariance structure follows Equation 8.

Chen et al. (2021) showed that the iterated least squares and maximum likelihood coefficient estimates perform equally well to find the true values of $\mathbf{A}$ and $\mathbf{B}$ in most situations.

3.2 Estimation of penalized VAR models

Consider again, $i = 1, ..., d$ dependent variables and $j = 1, ..., p$ independent predictors as in Section 2.3. To find estimation matrix $\hat{\boldsymbol{\beta}}_{Alasso}$ use Equation 10 of the Lasso where $\mathbf{x}_i^*$ of Equation 25 is the input instead of $\mathbf{x}_i$.

$$\mathbf{x}_i^* = \mathbf{x}_i / \hat{\mathbf{w}}_i \tag{25}$$

This accessory estimator matrix is denoted as $\hat{\boldsymbol{\beta}}_{Lasso}^*$. Then, $\hat{\boldsymbol{\beta}}_{Alasso}$ can be found by using Equation 26.

$$\hat{\boldsymbol{\beta}}_{Lasso}^* = \hat{\boldsymbol{\beta}}_{Alasso} * \hat{\mathbf{w}}_i \tag{26}$$

Therefore, to find $\hat{\boldsymbol{\beta}}_{Alasso}$ estimated $\hat{\boldsymbol{\beta}}_{Lasso}^*$ should be rescaled according to Equation 27.

$$\hat{\boldsymbol{\beta}}_{Alasso} = \hat{\boldsymbol{\beta}}_{Lasso}^* / \hat{\mathbf{w}}_i \tag{27}$$

Lastly, cross validation is used to determine optimal values for γ and λ (Zou, 2006).

To use Equation 25 initial values of $\hat{\boldsymbol{\beta}}$ are needed to compute the weights (see Equation 12). In this paper, the estimates of the Ridge regression, are chosen as initial values. This decision is made because using ordinary least squares estimates as initial values could be disadvantageous when collinearity is a concern (Zou, 2006). Furthermore, the estimates of the Lasso complicates the computation of weights when they are used as initial values. The reason behind this is that because $\hat{\boldsymbol{\beta}}_{Lasso}$ contains zero values, and following Equation 12 to calculate the weights, the problem occurs that a division by zero can't be made. The Ridge regression estimates seem suitable as initial values because the Ridge regression scales the insignificant estimates close to zero, whereby this computation problem with calculating the weights doesn't occur. To find $\hat{\boldsymbol{\beta}}_{Ridge}$, Equation 13 should be optimized. Just as with the estimation of the Alasso, cross validation is used to find optimal values for λ.

3.3 Comparison of the model estimates

If the Alasso or Ridge regression would be applied to the same data to calculate the estimates of the MAR, the estimates from the methods could be compared as follows. As mentioned in Section 2.2, the estimates of the MAR, $\mathbf{A}$ and $\mathbf{B}$, are $m \times m$ and $n \times n$ coefficient matrices that represent the row-wise and columns wise dependence, respectively. This estimates could be used to calculate autoregressive coefficient $\mathbf{B} \otimes \mathbf{A}$ in Equation 9, which is of dimension $(m*n) \times (m*n)$. Furthermore, consider i that represents the set of row-column combinations:

$$i = m_1 n_1, m_2 n_1, ..., m_m n_1, m_1 n_2, ..., m_m n_2, ..., m_1 n_n, ..., m_m n_n.$$

These row-column combinations represent the $i = 1, ..., d$ dependent and $j = 1, ..., p$ independent variables used in Sections 2.3 and 3.2. By calculating $\boldsymbol{\beta}_{Alasso}$ or $\boldsymbol{\beta}_{Ridge}$ with all $m * n$ row-column combinations, $\boldsymbol{\beta}_{Alasso}$ or $\boldsymbol{\beta}_{Ridge}$ will have dimensions $(m * n) \times (m * n)$ as well. These matrices can be compared with $\boldsymbol{\phi}$ in Equation 6, where the insignificant values in are set to zero or close to zero.

3.4 Estimation of the Impulse response functions

The gIRF introduced by Koop et al. (1996) and Pesaran and Shin (1998) will have the following application in this paper. The gIRF modeled in Equation 5 has elements $\mathbf{A}_h, \boldsymbol{\Sigma}_\epsilon, \mathbf{e}_j, \sigma_{jj}$, and δ_j. This elements could be computed with the results of the estimation of the MAR, Ridge and Alasso.

If the maximum likelihood estimation of Section 3.1.3 is used to estimate the MAR, $\boldsymbol{\Sigma}_\epsilon$ could be estimated with $\boldsymbol{\Sigma}_\epsilon = \Sigma_c \otimes \Sigma_r$ using Equation 8. If the projection method or iterated least squares method is used to calculate the MAR, or if the Alasso or Ridge regression has been estimated, $\text{vec}(\mathbf{E}_t)$ can be calculated with Equation 9 when the MAR has been estimated, and with Equation 6 when the Alasso or Ridge estimators has been constructed. Thereafter, $\boldsymbol{\Sigma}_\epsilon$ could be calculated as the the covariance matrix of $\text{vec}(\mathbf{E}_t)$.

Subsequently in all cases, σ_{jj} could be estimated as the diagonal of $\mathbf{\Sigma}_\epsilon$. Furthermore, for predetermined h periods ahead, the $t = 1, ..., h$ matrices of $\mathbf{A}$ could be calculated by $\mathbf{A}_h = (\mathbf{B} \otimes \mathbf{A})^h$ if the MAR was estimated, and by $\mathbf{A}_h = \boldsymbol{\beta}^h_{Alasso}$ or $\boldsymbol{\beta}^h_{Ridge}$ if the Alasso or Ridge regression has been used, by Equations 6, 9, 1 and 2. Lastly, with predetermined shock δ_j and an appropriate $\mathbf{e}_j$ as described in Section 2.1, the gIRFs could be computed.

4 Simulation study

In this section, the comparison will be made between the performance of four models, the VAR(1), MAR(1), Ridge, and Alasso, in generating gIRFs under various scenarios. The MAR will be calculated with iterated least squares and maximum likelihood estimation. This decision is made because the estimates of the projection method are less accurate than the estimates of the iterated least squares and maximum likelihood as mentioned in Section 3.1.1. Furthermore, both iterated least squares and maximum likelihood estimation are included to compare with one another in generating gIRFs. As mentioned, the main difference between the iterated least squares and maximum likelihood estimation, is that although both methods assume a Kronecker structure in $\boldsymbol{\phi}$, the maximum likelihood estimation also assumes a Kronecker structure in the covariance matrix of vec($\mathbf{E}_t$) (see Equation 8 and 9). Because vec($\mathbf{E}_t$) is also used to generate gIRFs, it is interesting to see whether this assumption of the maximum likelihood estimation is in its favor or not. As mentioned in Sections 3.1.2 and 3.1.3 the projection method estimates will be used to find the initial values for the iterated least squares and maximum likelihood estimation. Furthermore, as mentioned in Section 1, the performance of the Ridge regression will be taken into account to compare with the Alasso.

These methodologies will be evaluated across three scenarios, specifically within contexts of low and high data length and dimensionality. Mention that the term dimensionality referes to the amount of rows and columns and not to the length of the data. In scenario 1, 2 and 3 the methods will be compared when the data has length, an amount of rows and an amount of columns (T, m, n) being equal to $(200, 4, 3)$, $(200, 8, 10)$ and $(100, 8, 10)$, respectively. As mentioned in Section 1, the VAR should perform worse as the curse of dimensionality intensifies. To see whether this happens two adjustments will be made to the data, increasing the numbers of parameters to be estimated, and reducing the available amount of data. This will be done in the second and third scenario. In the second scenario the data dimensions will increase, whereby more paremeters have to be estimated. In the third scenario both the data dimensions and the length of the data will be adjusted. The second and third scenario will be compared with the first to see if the performance of the VAR decays when the curse of dimensionality intensifies. Furthermore, by making this comparison, the performance of the MAR and Alasso will be compared such that it becomes evident which method performs best in generating gIRFs under

specific conditions of available data and number of parameters to be estimated. In addition, by comparing the second and third scenario, it will become more evident how the methods perform under different stages of the curse of dimensionality.

To compare the methods in these scenarios, a real $\boldsymbol{\phi}$ and a real error term $\text{vec}(\mathbf{E}_t)$ with mean zero, and covariance structure $\boldsymbol{\Sigma_E}$, will be created. The real $\boldsymbol{\phi}$ will ensured to be stable by rescaling it using its eigenvalues and eigenvectors. To do this, the absolute maximum eigenvalue and the eigenvector of $\boldsymbol{\phi}$ will be calculated. If the absolute maximum eigenvalue is larger than one, it is used to rescale itself and the other eigenvalues. Then a new stable $\boldsymbol{\phi}$ will be created by using this rescaled eigenvalues and the eigenvectors (Lütkepohl, 2005; Golub and Van Loan, 2013). Furthermore, $\boldsymbol{\Sigma_E}$ will ensured to be a symmetric positive semidefinite matrix by rescaling $\boldsymbol{\Sigma_E}$ with the Higham method (Higham, 1988). By applying this method, first the matrix will be multiplied with its transpose, and divided by two, to ensure symmetry. Then, the SVD is used to compute the nearest positive semi-definite matrix by reconstructing the matrix from its positive eigenvalues.

Then, while using the first row of $\text{vec}(\mathbf{E}_t)$ as initial values, $\text{vec}(\mathbf{X}_t)$ and $\text{vec}(\mathbf{X}_{t-1})$ will be calculated using Equation 6. Furthermore, estimators $\boldsymbol{\phi}$ and $\text{vec}(\mathbf{E}_t)$ will be constructed in multiple ways, which leads to the fourteen data structures that will be used to compare the methods in the three scenarios. The real $\boldsymbol{\phi}$ will be generated: randomly, with a Kronecker structure (suitable for the MAR), or with half of the coefficients set to zero (suitable for the Alasso). In addition, $\boldsymbol{\Sigma_E}$ will be generated randomly, or with a Kronecker structure. Subsequently, the estimators $\boldsymbol{\phi}$ and $\boldsymbol{\Sigma_E}$ will be multiplied with a normally distributed error term (before rescaling) with mean zero and standard deviation 0.01, 0.1, 1 or 10. This procedure will be applied except when $\boldsymbol{\phi}$ and $\boldsymbol{\Sigma_E}$ are generated randomly, as adding random noise to an already random matrix is not meaningful. By implementing this approach, the MAR and Alasso will be compared under conditions where they are expected to perform optimally, when $\boldsymbol{\phi}$ is constructed ideally and the error term exhibits a low standard deviation, and under conditions where their performance may be suboptimal, when $\boldsymbol{\phi}$ is constructed suboptimal or the error term exhibits a high standard deviation. In addition, by using these structures of $\boldsymbol{\Sigma_E}$, the comparison can be made between the two estimation methods to calculate the MAR, because the covariance structure is created randomly or in accordance with Equation 8 with some noise. However, $\boldsymbol{\Sigma_E}$ will only be calculated with a Kronecker structure as well if $\boldsymbol{\phi}$ is ideally for the MAR. It makes no sense to generate $\boldsymbol{\Sigma_E}$ with a Kronecker structure when $\boldsymbol{\phi}$ is generated randomly or ideally for the Alasso.

The performance of the methods will be measured by estimating the mean squared error (MSE) between the gIRFs calculated with the real $\boldsymbol{\phi}$, $\text{vec}(\mathbf{E}_t)$, $\text{vec}(\mathbf{X}_t)$ and $\text{vec}(\mathbf{X}_{t-1})$, and the gIRFs calculated with the different methods, using the same $\text{vec}(\mathbf{X}_t)$ and $\text{vec}(\mathbf{X}_{t-1})$. The gIRFs will be estimated ten periods ahead, were one of the variables gets a shock with the size of the standard deviation of this variable.

Table 1: Scenario 1

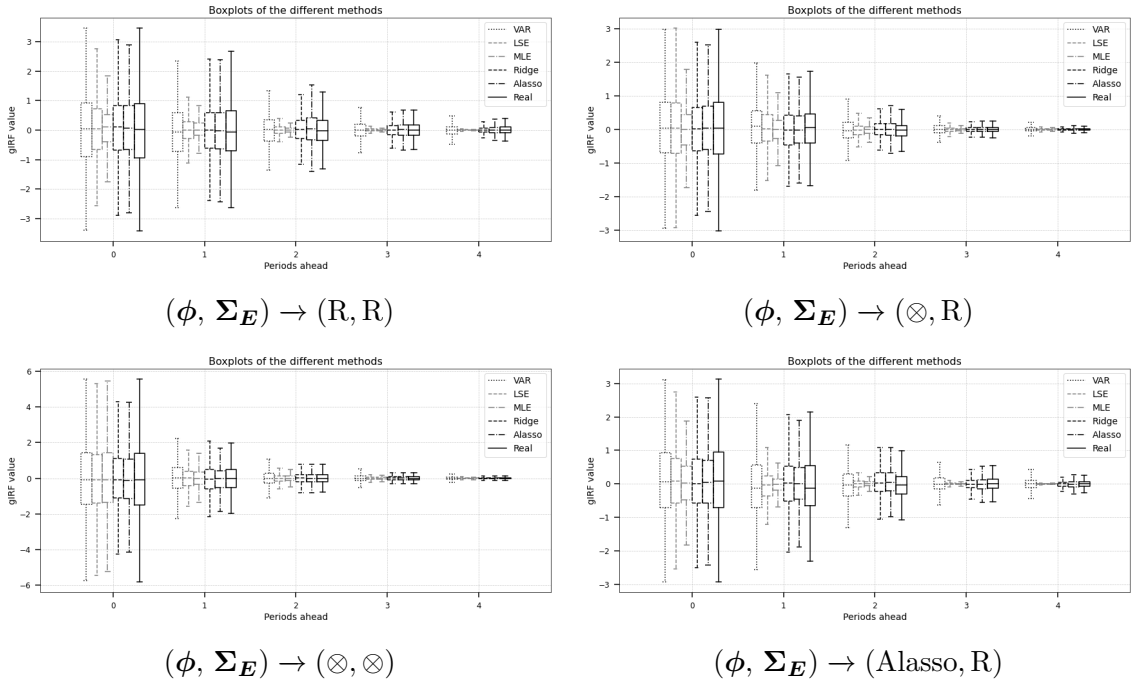
			VAR		LSE		MLE		Ridge		Alasso	
$\boldsymbol{\phi}$	$\boldsymbol{\Sigma_E}$	Error	$\mathbb{E}$	σ	$\mathbb{E}$	σ	$\mathbb{E}$	σ	$\mathbb{E}$	σ	$\mathbb{E}$	σ
R	R	-	0.023	0.025	0.100	0.109	0.201	0.210	0.360	0.410	0.407	0.477
MAR	R	0.01	0.015	0.017	0.003	0.005	0.111	0.144	0.173	0.284	0.193	0.318
	$\otimes$	0.01	0.032	0.125	0.009	0.024	0.013	0.034	0.479	1.188	0.528	1.270
	R	0.1	0.015	0.016	0.008	0.010	0.108	0.139	0.183	0.250	0.201	0.286
	$\otimes$	0.1	0.029	0.051	0.019	0.028	0.037	0.067	0.596	1.842	0.655	2.105
	R	1	0.019	0.019	0.095	0.095	0.210	0.224	0.318	0.369	0.351	0.404
	$\otimes$	1	0.042	0.058	0.291	0.574	0.336	0.550	0.773	1.210	0.821	1.221
	R	10	0.021	0.022	0.107	0.108	0.212	0.226	0.333	0.376	0.364	0.406
	$\otimes$	10	0.053	0.064	0.329	0.576	0.560	0.921	0.931	1.355	1.001	1.403
Alasso	R	0.01	0.020	0.019	0.081	0.091	0.177	0.189	0.277	0.383	0.296	0.417
	R	0.1	0.019	0.019	0.086	0.098	0.207	0.224	0.309	0.363	0.334	0.391
	R	1	0.020	0.020	0.087	0.083	0.201	0.218	0.314	0.390	0.353	0.432
	R	10	0.020	0.020	0.102	0.117	0.213	0.238	0.353	0.392	0.390	0.431

The mean and standard deviation of the MSE between the real and calculated gIRFs by multiple methods, where the real $\boldsymbol{\phi}$ *and* $\boldsymbol{\Sigma_E}$ *have different structures and* $(T, m, n) = (200, 4, 3)$.

This approach will be repeated 500 times in scenario 1 and 250 times in scenario 2 and 3. The reason behind this is that the computation time of the simulations in scenario 2 and 3 is a lot higher.

To compare the performance of the models, the mean and standard deviation of these MSEs will be calculated. Furthermore, a boxplot of the gIRFs constructed across all scenarios, incorporating all structures of $\boldsymbol{\phi}$ and $\boldsymbol{\Sigma_E}$ with an error term of 0.1, for each method, will provide greater insight into the performance of the methods. This approach allows for the presentation of boxplots for the different methods under conditions where $\boldsymbol{\phi}$ and $\boldsymbol{\Sigma_E}$ are either ideally constructed or not. The plots with an error term of 0.1 have been selected because, with a higher error, the data structure would shift towards the completely random structure, making it less apparent how the methods perform under optimal conditions. On the other hand, an error term of 0.1 is chosen instead of 0.01 because these boxplots represent a bit better how these methods would work in a general setting. The boxplots with an error term of 0.01 mainly show how the boxplots work in an almost ideal situation. However, the performance of the methods in the setting where both $\boldsymbol{\phi}$ and $\boldsymbol{\Sigma_E}$ are random indicates best how these methods would work when used in other situations. Furthermore, these boxplots will be generated for the first five periods only, as beyond this point, most gIRFs converge to values close to zero and exhibit minimal differences.

Figure 1: Boxplots of the gIRFs in scenario 1



$(\boldsymbol{\phi}, \boldsymbol{\Sigma_E}) \to (\mathrm{R}, \mathrm{R})$ $(\boldsymbol{\phi}, \boldsymbol{\Sigma_E}) \to (\otimes, \mathrm{R})$

$(\boldsymbol{\phi}, \boldsymbol{\Sigma_E}) \to (\otimes, \otimes)$ $(\boldsymbol{\phi}, \boldsymbol{\Sigma_E}) \to (\mathrm{Alasso}, \mathrm{R})$

Boxplots of the gIRFs produced by the different methods. $(T, m, n) = (200, 4, 3)$, $\boldsymbol{\phi}$ *and* $\boldsymbol{\Sigma_E}$ *constructed with different structures and an error term of 0.1.*

4.1 Performance in low-dimensional settings

In Tables 1, 2 and 3 the resulting means and standard deviations of the MSEs between the methods and the real gIRFs are shown in the different scenarios. Additionally, Figures 1, 2 and 3 represent the boxplots of the generated gIRFs by the different methods. In these tables and figures R and $\otimes$ represent a random and Kronecker structure, respectively. Furthermore, LSE and MLE refer to the estimation method used to calculated the MAR estimates.

What stands out in Table 1 is that the VAR performs best when $\boldsymbol{\phi}$ and $\boldsymbol{\Sigma_E}$ are generated randomly. Furthermore, in this situation the performance of the LSE, MLE, Ridge and Alasso aligns according to this order. In addition, in the context where $\boldsymbol{\phi}$ is generated suitable for the MAR with low standard deviation of the error, the LSE performs best, regardless of whether $\boldsymbol{\Sigma_E}$ is generated randomly or with a Kronecker structure. When the error increases, once again the VAR outperforms the LSE. Furthermore, the MLE performs significantly better when $\boldsymbol{\Sigma_E}$ is generated with a Kronecker structure compared to a random structure. However, the MLE still performs worse than the LSE, and only performs better than the VAR when the error term is really low. When the error increases, its performance decreases such as with the LSE. Lastly, the Ridge regression performs worse than the other methods except for the Alasso. The Alasso performs the worst when $\boldsymbol{\phi}$ is generated suitable for the MAR in this scenario.

Table 2: Scenario 2

			VAR		LSE		MLE		Ridge		Alasso	
$\boldsymbol{\phi}$	$\boldsymbol{\Sigma_E}$	Error	$\mathbb{E}$	σ	$\mathbb{E}$	σ	$\mathbb{E}$	σ	$\mathbb{E}$	σ	$\mathbb{E}$	σ
R	R	-	0.239	0.157	0.130	0.104	0.233	0.226	0.386	0.349	0.428	0.420
MAR	R	0.01	0.157	0.108	0.004	0.004	0.124	0.141	0.241	0.229	0.172	0.198
	$\otimes$	0.01	0.209	0.246	0.008	0.012	0.049	0.078	0.668	1.157	0.471	1.027
	R	0.1	0.156	0.109	0.008	0.010	0.132	0.169	0.236	0.215	0.176	0.245
	$\otimes$	0.1	0.216	0.228	0.016	0.021	0.051	0.061	0.613	0.691	0.434	0.531
	R	1	0.251	0.169	0.144	0.115	0.291	0.287	0.461	0.448	0.483	0.484
	$\otimes$	1	0.370	0.311	0.391	0.541	0.495	0.769	1.058	1.384	1.094	1.604
	R	10	0.247	0.179	0.137	0.131	0.254	0.255	0.421	0.378	0.440	0.440
	$\otimes$	10	0.410	0.328	0.366	0.366	0.487	0.684	0.991	1.084	1.042	1.293
Alasso	R	0.01	0.255	0.166	0.135	0.118	0.252	0.239	0.405	0.414	0.447	0.496
	R	0.1	0.222	0.150	0.146	0.146	0.264	0.271	0.395	0.371	0.439	0.414
	R	1	0.249	0.156	0.157	0.144	0.251	0.231	0.428	0.410	0.465	0.547
	R	10	0.233	0.175	0.147	0.138	0.263	0.267	0.443	0.461	0.447	0.477

The mean and standard deviation of the MSE between the real and calculated gIRFs by multiple methods, where the real $\boldsymbol{\phi}$ *and* $\boldsymbol{\Sigma_E}$ *have different structures and* $(T, m, n) = (200, 8, 10)$.

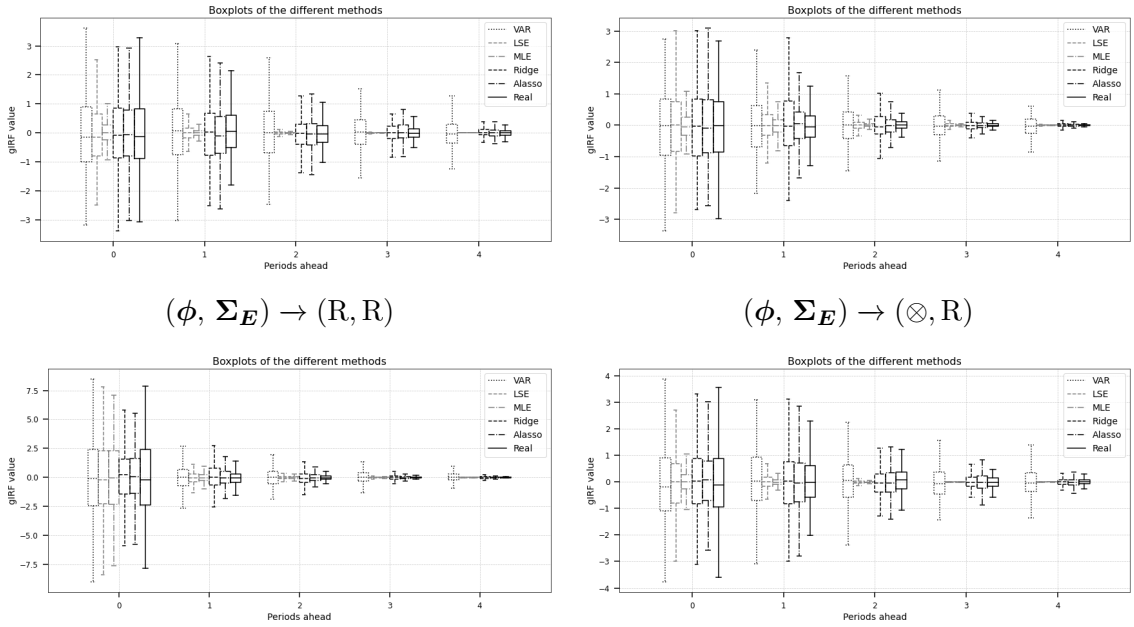
Regarding the situation where $\boldsymbol{\phi}$ is generated ideally for the Alasso, the performance of the methods in line with their performance when $\boldsymbol{\phi}$ and $\boldsymbol{\Sigma_E}$ were generated randomly.

Furthermore, in Figure 1 the boxplots of the different methods are shown for this scenario. In this figure, it stands out that both methods to estimate the MAR tend to generate too weak responses by calculating gIRFs when $\boldsymbol{\phi}$ and $\boldsymbol{\Sigma_E}$ are generated randomly or suitable for the Alasso, especially when the periods ahead is larger than 1. This is because the 95% confidence intervals of the boxplots of these methods are lower the real boxplot in these situations. However, the LSE generates better responses when $\boldsymbol{\phi}$ is generated suitable for the MAR regardless of the structure of $\boldsymbol{\Sigma_E}$. The MLE on the other hand, only performs better responses when both $\boldsymbol{\phi}$ and $\boldsymbol{\Sigma_E}$ have a Kronecker structure.

4.2 Performance in high-dimensional settings

In Table 2 the resulting means and standard deviations in the second scenario are shown. First, what stands out is that in the situation where both $\boldsymbol{\phi}$ and $\boldsymbol{\Sigma_E}$ are generated randomly, the VAR does not outperforms the other models anymore. The LSE performs best in this situation, although it performed a bit worse compared with the first scenario.

Figure 2: Boxplots of the gIRFs in scenario 2



$(\boldsymbol{\phi}, \boldsymbol{\Sigma_E}) \rightarrow (\mathrm{R}, \mathrm{R})$ $(\boldsymbol{\phi}, \boldsymbol{\Sigma_E}) \rightarrow (\otimes, \mathrm{R})$

$(\boldsymbol{\phi}, \boldsymbol{\Sigma_E}) \rightarrow (\otimes, \otimes)$ $(\boldsymbol{\phi}, \boldsymbol{\Sigma_E}) \rightarrow (\mathrm{Alasso}, \mathrm{R})$

Boxplots of the gIRFs produced by the different methods. $(T, m, n) = (200, 8, 10)$, $\boldsymbol{\phi}$ *and* $\boldsymbol{\Sigma_E}$ *constructed with different structures and an error term of 0.1.*

Furthermore, the MLE performed slightly worse than the VAR, while the Ridge and Alasso methods ranked as the second worst and the worst performers, respectively. Regarding the ideal situation for the MAR, it comes to attention that the LSE performs best in all cases except for the case when $\boldsymbol{\Sigma_E}$ had a Kronecker structure and an error term of 1. In addition, MLE outperformed the VAR when the error was 0.01 or 0.1 regardless of the structures of $\boldsymbol{\Sigma_E}$. However, when the error increased the VAR outperformed the MLE once again. Furthermore, it stands out that the Alasso performed a bit better that the Ridge in some situations with this structure of $\boldsymbol{\phi}$. Lastly, Table 2 shows that if $\boldsymbol{\phi}$ is generated ideally for the Alasso the LSE again performs best. Thereafter, ranking second and third, the VAR and MLE, with very similar performance. In addition, the Ridge outperformed the Alasso, which gives the Ridge and Alasso the fourth and fifth place once again.

Moreover, the boxplots regarding the generated gIRFs in this scenario are shown in Figure 2. It is notable that in comparison with the boxplots in Figure 1, the VAR seems to generate too extreme responses to the shock, especially for further periods ahead. Additionally, in line with Figure 1, the MAR generated mostly too weak responses. Only when $\boldsymbol{\phi}$ and $\boldsymbol{\Sigma_E}$ are suitable for the LSE and MLE, they generate responses in line with the real responses, for a few periods ahead.

Table 3: Scenario 3

			VAR		LSE		MLE		Ridge		Alasso	
$\boldsymbol{\phi}$	$\boldsymbol{\Sigma_E}$	Error	$\mathbb{E}$	σ	$\mathbb{E}$	σ	$\mathbb{E}$	σ	$\mathbb{E}$	σ	$\mathbb{E}$	σ
R	R	-	5.104	3.681	0.180	0.160	0.335	0.310	0.500	0.463	0.590	0.568
MAR	R	0.01	4.020	3.228	0.013	0.012	0.189	0.212	0.297	0.276	0.243	0.259
	$\otimes$	0.01	4.990	6.192	0.033	0.041	0.096	0.126	0.668	0.925	0.629	1.007
	R	0.1	4.053	2.977	0.016	0.017	0.209	0.265	0.299	0.295	0.280	0.345
	$\otimes$	0.1	5.651	6.001	0.065	0.113	0.115	0.165	0.656	0.748	0.665	0.926
	R	1	4.633	3.269	0.179	0.164	0.306	0.282	0.549	0.516	0.586	0.712
	$\otimes$	1	6.160	6.373	0.563	0.907	0.568	0.959	1.226	1.600	1.276	1.606
	R	10	5.044	3.582	0.203	0.177	0.353	0.355	0.581	0.623	0.709	0.781
	$\otimes$	10	7.230	5.499	0.485	0.503	0.532	0.603	1.112	1.127	1.255	1.288
Alasso	R	0.01	5.144	3.713	0.191	0.155	0.331	0.352	0.487	0.419	0.576	0.517
	R	0.1	5.348	4.467	0.193	0.185	0.311	0.314	0.522	0.454	0.605	0.527
	R	1	5.711	4.947	0.186	0.155	0.302	0.291	0.538	0.518	0.667	0.734
	R	10	5.327	4.113	0.182	0.166	0.344	0.391	0.515	0.474	0.615	0.654

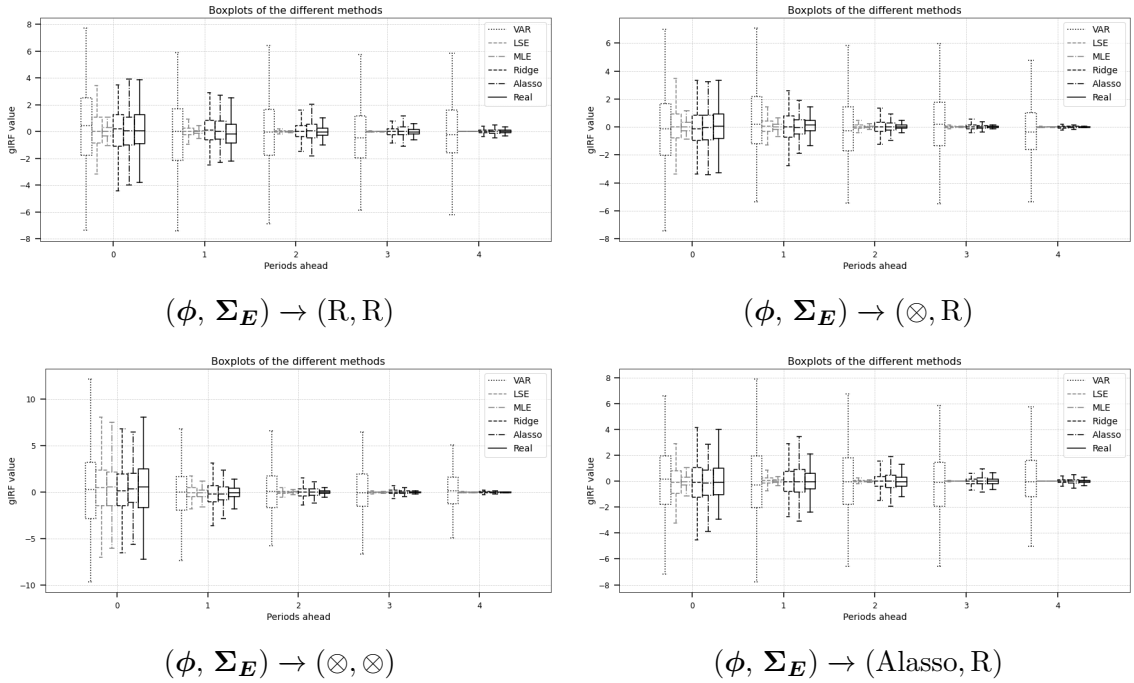
The mean and standard deviation of the MSE between the real and calculated gIRFs by multiple methods, where the real $\boldsymbol{\phi}$ *and* $\boldsymbol{\Sigma_E}$ *have different structures and* $(T, m, n) = (100, 8, 10)$.

4.3 Performance in high-dimensional settings with limited data

Table 3 shows the MSE for the third scenario. It directly comes to attention that the VAR performs by far the worst in this scenario for all data structures of $\boldsymbol{\phi}$ and $\boldsymbol{\Sigma_E}$. Furthermore, the performance of the other methods ranks mostly in the usual order, regardless of the data structure. The LSE demonstrates the best performance, followed by the MLE, then the Ridge, with the Alasso showing the worst performance. However, there are some situations where the Alasso performed slightly better than the Ridge. In addition, such as in the other scenarios, the MLE performs better when the $\boldsymbol{\Sigma_E}$ has a Kronecker structure.

Additionally, the boxplots in Figure 3 show that the VAR performed far too extreme gIRFs in this scenario. Furthermore, it seems that the gIRFs of the VAR do not converge to zero even for a lot of periods ahead. Regarding the gIRFs constructed by the LSE and MLE, such as in scenarios 1 and 2, they mostly generated too weak responses. Moreover, this figure shows that the Ridge and Alasso tend to estimate a bit too extreme responses to the shock, especially when the periods ahead increase. Although, these methods appear to estimate the shock with appropriate magnitude, it does not seem to converge to zero quickly enough. However, this is just a small difference in comparison with the other scenarios.

Figure 3: Boxplots of the gIRFs in scenario 3



$(\boldsymbol{\phi}, \boldsymbol{\Sigma_E}) \rightarrow (\mathrm{R}, \mathrm{R})$ $(\boldsymbol{\phi}, \boldsymbol{\Sigma_E}) \rightarrow (\otimes, \mathrm{R})$

$(\boldsymbol{\phi}, \boldsymbol{\Sigma_E}) \rightarrow (\otimes, \otimes)$ $(\boldsymbol{\phi}, \boldsymbol{\Sigma_E}) \rightarrow (\mathrm{Alasso}, \mathrm{R})$

Boxplots of the gIRFs produced by the different methods. $(T, m, n) = (100, 8, 10)$, $\boldsymbol{\phi}$ *and* $\boldsymbol{\Sigma_E}$ *constructed with different structures and an error term of 0.1.*

5 Results

In Section 4 the performance of the VAR(1), MAR(1), Ridge and Alasso in generating gIRFs have been tested with different data structures and dimensions by doing a simulation study.

This simulation study showed, first of all, that the VAR suffers from the curse of dimensionality. In scenario 1, the VAR performed best of all methods, with most data structures. However, the performance of the VAR decreased already in scenario 2 where only the number of parameters increased. Moreover, in scenario 3, where also the available amount of data decreased, the VAR performed by far the worst of all methods. Moreover, because of the curse of dimensionality, the VAR constructed unstable coefficient matrices in scenario 3. These matrices are used to construct the gIRFs as mentioned in Section 3.4. Because of the instability of its coefficient matrices, the boxplots of the VAR in Figures 3 did not converge to zero anymore, even for a lot of periods ahead.

Furthermore, the LSE seemed to be the best method to estimate gIRFs in most situations. The LSE turned out to be a robust estimation method to generate gIRFs, its performance only decreased a little bit when the curse of dimensionality intensified. Additionally, the LSE outperformed the MLE with all data structures in all scenarios, even when the covariance structure of the error term was generated with a Kronecker structure in accordance with Equation 8. Therefore, the conclusion

can be made that to construct gIRFs, the LSE is a better method to estimate the MAR compared with MLE. In most situations, the LSE performend better than the MLE, because Figures 1, 2 and 3 showed that the performance of the LSE is only dependent on a Kronecker structure in the coefficient matrix of the to be estimated model. On the other hand, these figures showed that the performance of the MLE is dependent on a Kronecker structure in both the coefficient matrix, and the covariance matrix of the error term of the to be estimated model. Furthermore, the LSE performs better than the MLE, even in the ideal situation for the MLE, because although both methods generate too weak responses in most situations, the LSE generates responses with a magnitude closer to the real magnitude of the response. However, apart from reasoning why the LSE turned out to perform better, the weak responses by these methods on itself point out the weakness of these methods to construct gIRFs. As shown in Figures 1, 2 and 3, both estimation methods generate too weak responses, especially for periods further ahead, meaning that these estimation methods have generated estimation coefficients that converge towards zero too quickly. The reason behind this might be related to the rescaling of the coefficient matrices to estimate the coefficient matrices of the LSE and MLE. This sounds straightforward because the rescaling techniques of these methods ensure these coefficient matrices to be stable, and the resulting gIRFs turned out to be 'too stable'. Furthermore, as mentioned earlier, this problem is more notable for the MLE, while the estimates of the MLE are estimated by rescaling both $\mathbf{A}_{MLE}$ and $\mathbf{\Sigma}_{\boldsymbol{r}}$. However, this is just a suggestion, the real cause for this problem lies not within the scope of this research.

Although the MLE has its mentioned weaknesses, whereby the MLE is not the best estimation technique to estimate the MAR and generate gIRFs, it still performed better than the Ridge and Alasso in all scenarios and situations. This is a remarkable result by looking at the boxplots in Figures 1, 2 and 3, because the Ridge and Alasso generated boxplots that were more similar to the real boxplots. This can have multiple reasons. The most obvious reason would be that the Ridge and Alasso might generate the correct response magnitude to the shock, but that the direction of this shock is wrong. A second reason could be that these methods generate a response with a big magnitude, if the real response is small, and the other way around. Subsequently, constructing a boxplot of all simulations would result in the magnitudes of the responses of these methods boxplots aligning with the magnitude indicated by the real boxplot. However, in this instance, the magnitudes of the gIRFs produced by the methods deviate from those of the real gIRFs in each individual simulation.

Lastly, the Ridge regression performed better in most cases than the Alasso. This indicates that to generate gIRFs, insignificant variables can better be set to a low value than to zero. It is remarkable that even if half of the values of the coefficient matrix to be estimated has been set to zero before rescaling, the Ridge performed better than the Alasso. Because the estimated coefficient by the Alasso should be

closer to the real coefficient in this setting, this result can only be explained by the covariance structure of the error term. This indicates that the resulting error term by using the coefficient matrix of the Ridge, has a covariance structure that is closer to the real covariance structure.

6 Empirical application

As mentioned in Section 1, the results of this research could be applicable for constructing macroeconometric models that incorporate multiple indicators for a lot of countries. In this section, the methods for constructing gIRFs will be applied to an emperical setting.

Consider Table 4. In this table the biggest import and export partners of the Netherlands are listed. The data used to generate this table has been found via the WorldBank (2024). Because these countries trade a lot with the Netherlands, they are assumed to be influential on the economic indicators of the Netherlands.

Table 4: Import and Export Partner Share with the Netherlands

Country	Import Share (%)	Export Share (%)
Germany	17.64	23.51
Belgium	10.19	10.48
China	8.71	2.04
US	8.08	4.30
UK	5.53	8.71
France	4.38	7.91
Russia	3.75	1.00
Norway	3.19	0.90
Italy	2.39	4.15
Poland	1.93	2.93
Japan	1.91	0.82
Spain	1.83	2.91
Malaysia	1.79	0.27
Sweden	1.69	1.85
Czechia	1.38	1.57
Ireland	1.37	0.82
Brazil	1.16	0.64
Vietnam	1.14	0.19
Hong Kong	1.02	0.54
Denmark	1	1.37

Import and Export Partner Share with the Netherlands, the top 20 countries in 2015.

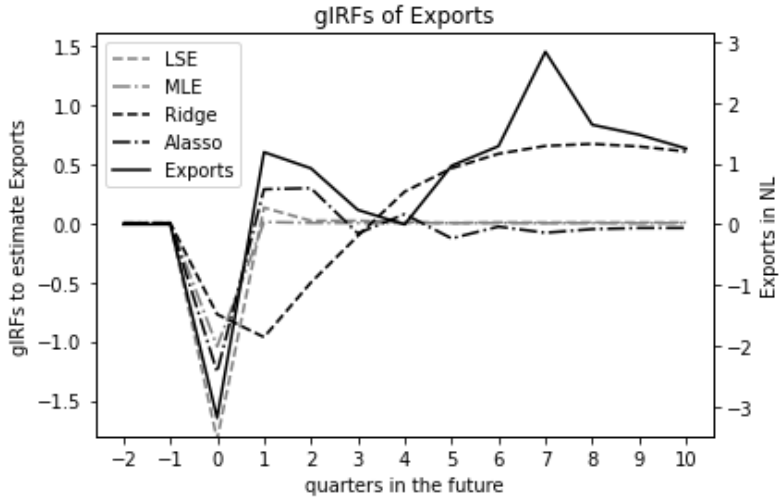
Data of four economic indicators have been collected for these countries, including the Netherlands, starting at the first quarter of 1995 up to and including the first quarter of 2015, represented per quarter. This data has been found via the Organisation for Economic Co-operation and Development (OECD, 2024). However, there was not (enough) data available for China, Russia, Malaysia, Brazil, Vietnam and Hong Kong. Therefore, these countries have not been added to the dataset. The other countries represent the columns of the dataset. Data concerning the growth rate of the following economic indicators has been gathered: the volume of the gross domestic product (GDP), the consumer price index (CPI), and the volume of exports and imports. Furthermore, the short term interest rates for these countries were added to the data as percent per annum. Because there are 81 quarters between 1995 and 2015 and five indicators have been taken into account regarding 15 countries, the dataset has length and dimensions (T, m, n) being equal to $(81, 5, 15)$.

The methods examined in this paper will be used to generate gIRFs as a result of shocking one of the indicators in one of the countries. Furthermore, these gIRFs will be compared to the real development of the response indicator in the response country. As has been mentioned in Section 5, the VAR model will not work in this situation because of the curse of dimensionality. Moreover, it is not feasible to simply gather more data, because already six countries had to be removed from the dataset because there was no data available for these countries. By gathering data for a longer time period, more countries will have to be removed because there is no data available for these countries over extended time periods. In addition, as mentioned in Section 1, it is not favorable to exclude economic indicators or relevant countries for the Netherlands because these variables are likely to be related (Pesaran et al., 2004). Therefore, this empirical application underlines the motivation of this paper. With the mentioned methodologies in this paper, more robust gIRFs could be conducted where all these economic indicators and relevant countries will be taken into account.

By looking at the data for the mentioned indicators and countries, some periods ahead of 2015, it stands out that between the first and second quarter of 2015 the GDP growth of Ireland is shocked negatively from 20.6 to minus 2. Therefore, using the methods of this study, the gIRFs regarding the exports in the Netherlands will be estimated, as a result of this shock, for ten periods ahead, using the described dataset. As mentioned, these gIRFs will be compared with the real development of the export volume of the Netherlands during this time period.

In Figure 4 the resulting gIRFs and the real development of exports in the Netherlands are shown over the corresponding time period. In this graph the vertical labels on the left correspond with the gIRFs, and those on the right with the exports value in the Netherlands. What stands out, is that the LSE method generated a more extreme direct response to the shock than the Ridge and Alasso, although Figures 1, 2 and 3 showed that in most situations this is likely to be the other way around.

Figure 4: gIRFs of the export value compared with its real adjustment



The gIRFs estimated by the different methods of the exports of the Netherlands resulting from a shock on Irelands GDP growth.

Furthermore, it comes to attention that the Alasso is able to generate the correct path of the exports volume in the Netherlands quite well until four quarters in the future.

There are multiple arguments why this result could be in line with, or contrary on, the results of this research. First, this application is in line with the results of this paper, because it has been shown that when the curse of dimensionality is intense, in most situations only the Ridge and Alasso were able to generate gIRFs that had extreme enough values, especially for a number of periods ahead beyond one. Therefore, it is also in line with Section 5 that the LSE and MLE generated too weak responses. However, this research showed that in most cases the Alasso performed the worst with the mentioned data structures and dimensions. This is not completely contrary on the results of this paper. This research shows that with the mentioned data structures and dimensions the other methods work better on average. However, this application proves that there are still situations where the Alasso works better than the other methods.

Furthermore, it can be seen that Ridge is (almost) unstable in this application, because its gIRF does not tend to go to zero for a lot of periods ahead. However, this result has only been seen a little bit in Section 4.3. This is probably because the curse of dimensionality is in this application more intense than in the scenarios of Section 4. Therefore, the Ridge was able to construct stable gIRFs before, but fails in this application. This underlines why the Ridge regression had solely been added to the simulation study to be compared with the Alasso, as mentioned in Section 1.

To dive deeper into the reasons behind the differences between the estimated gIRFs, the calculated inputs for the gIRF (see Section 3.4) will be compared to one another.

7 Conclusion

In this paper a potential method to construct a macroeconometric model capable of accurately representing multinational relationships by modelling high dimensional multivariate time series, the matrix autoregressive model, has been examined. In the existing literature, varying vector autoregressive models have been used to do this by generating impulse response functions whereby the time points have been treated as vector. An important drawback of this approach is that these models are vulnerable for the curse of dimensionality, which weakens their performance to generate impulse response functions. The matrix autoregressive model seemed promising because it needs less data to calculate its autoregressive coefficient, as a result of which it should be more robust against the curse of dimensionality. To test whether the matrix autoregressive model works better to generate impulse response functions, a simulation study has been done. In this study multiple variations of the vector autoregressive model, the VAR, Ridge and Alasso, where compared to two estimations techniques to estimate the matrix autoregressive model, the least squares and maximum likelihood estimation, regarding constructing generalized impulse response functions.

It turned out that in general the matrix autoregressive model estimated with the least squares estimation constructed the most robust generalized impulse response functions in settings where the curse of dimensionality becomes more pronounced, especially when the autoregressive coefficient to be estimated possesses a Kronecker structure. Only when the curse of dimensionality was less noticeable, the VAR turned out to construct more accurate generalized impulse response functions in some settings. However, Pesaran et al. (2004) showed that in economic theory, economic indicators are highly related to a lot of other (international) economic indicators, meaning that to create a macroeconometric model, the curse of dimensionality will be noticeable.

Although the matrix autoregressive model turned out to generate robust generalized impulse response functions, its main drawback is that it generates to weak responses for a time period further ahead. It would be interesting for further research to dive into the reasons why the matrix autoregressive model generates to weak responses and how this could be solved. This drawback has also been emphasized in the emperical application, where the Alasso turned out to generate the most accurate generalized impulse response function, because it performed better for some periods ahead.

References

Barigozzi, M. and Conti, A. M. (2018). On the stability of euro area money demand and its implications for monetary policy. *Oxford Bulletin of Economics and Statistics*, 80(4):755–787.

Chen, R., Xiao, H., and Yang, D. (2021). Autoregressive models for matrix-valued time series. *Journal of Econometrics*, 222(1):539–560.

Dey, S. R. and Tareque, M. (2020). External debt and growth: role of stable macroeconomic policies. *Journal of Economics, Finance and Administrative Science*, 25(50):185–204.

Ewing, B. T. (2002). The transmission of shocks among s&p indexes. *Applied Financial Economics*, 12(4):285–290.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 96(456):1348–1360.

Fan, J. and Peng, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters.

Golub, G. H. and Van Loan, C. F. (2013). *Matrix computations*. JHU press.

Hafidh, A. A. (2021). Responses of islamic banking variables to monetary policy shocks in indonesia. *Islamic Economic Studies*, 28(2):174–190.

Hammer, P. (1962). Adaptive control processes: a guided tour (r. bellman).

Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer.

Higham, N. J. (1988). Computing a nearest symmetric positive semidefinite matrix. *Linear algebra and its applications*, 103:103–118.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Inoue, A. and Rossi, B. (2021). A new approach to measuring economic policy shocks, with an application to conventional and unconventional monetary policy. *Quantitative Economics*, 12(4):1085–1138.

Koop, G., Pesaran, M. H., and Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of econometrics*, 74(1):119–147.

Lanne, M., Meitz, M., and Saikkonen, P. (2017). Identification and estimation of non-gaussian structural vector autoregressions. *Journal of Econometrics*, 196(2):288–304.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.

Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso.

Montes-Rojas, G. (2022). Estimating impulse-response functions for macroeconomic models using directional quantiles. *Journal of Time Series Econometrics*, 14(2):199–225.

OECD (2024). OECD Data Explorer: Short-term Economic Statistics. https://data-explorer.oecd.org/?lc=en. Accessed: 2024-07-17.

Pesaran, H. H. and Shin, Y. (1998). Generalized impulse response analysis in linear multivariate models. *Economics letters*, 58(1):17–29.

Pesaran, M. H., Schuermann, T., and Weiner, S. M. (2004). Modeling regional interdependencies using a global error-correcting macroeconometric model. *Journal of Business & Economic Statistics*, 22(2):129–162.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, pages 1–48.

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

Van Loan, C. F. (2000). The ubiquitous kronecker product. *Journal of computational and applied mathematics*, 123(1-2):85–100.

Van Loan, C. F. and Pitsianis, N. (1993). *Approximation with Kronecker products*. Springer.

Wiesen, T. F. and Beaumont, P. M. (2024). A joint impulse response function for vector autoregressive models. *Empirical Economics*, 66(4):1553–1585.

WorldBank (2024). Netherlands trade profile: Trade summary for netherlands 2015. Accessed: 2024-07-16.

Zhang, R., Zhao, T., Lu, Y., and Xu, X. (2022a). Relaxed adaptive lasso and its asymptotic results. *Symmetry*, 14(7):1422.

Zhang, X., Zhou, J., and Du, X. (2022b). Impact of oil price uncertainty shocks on china's macro-economy. *Resources Policy*, 79:103080.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429.