EÖTVÖS LORÁND UNIVERSITY

CORVINUS UNIVERSITY OF BUDAPEST

Paulik Kata

MSc in Actuarial and Financial Mathematics

Faculty of Quantitative Finance

Measuring causality with machine learning method: A Lasso- Bootstrap Granger Causality Approach

MSc Thesis

Supervisor:

Badics Milán Csaba

Corvinus University of Budapest

Department of Finance

Budapest, 2022

Acknowledgement

I would like to express my gratitude to my supervisor, Milán Badics for his time, useful advices, patience and for the many consultations and guidance I recieved whenever I turned to him.

I would also like to thank my family and friends, especially my mother and Marcell Vörös, who always supported me during my studies.

NYILATKOZAT

Név: Paulik Kata

ELTE Természettudományi Kar, szak: Biztosítási és pénzügyi matematika

NEPTUN azonosító: USAGST

Szakdolgozat címe:
Measuring causality with machine learning method: A Lasso- Bootstrap Granger Causality Approach

A **szakdolgozat** szerzőjeként fegyelmi felelősségem tudatában kijelentem, hogy a dolgozatom önálló szellemi alkotásom, abban a hivatkozások és idézések standard szabályait következetesen alkalmaztam, mások által írt részeket a megfelelő idézés nélkül nem használtam fel.

Budapest, 2022.05.30.

Paulik Kata

a hallgató aláírása

Contents

1 Introduction **1**

2 High-dimensional Granger causality testing **5**

2.1 Granger causality testing in VAR framework 6

3 Post-selection inference **8**

3.1 Penalized Maximum Likelihood estimation 8

3.2 Tuning parameter selection 9

3.3 Post-selection inference 10

3.4 Bootstrap Granger causality test 12

4 Simulation study **14**

4.1 Simulations 14

4.2 Results of the sensitivity analysis of the bootstrap Granger causality test for the number and length of the time series 17

4.3 Results of the sensitivity analysis of the bootstrap Granger causality test for the correlation of the error terms 19

4.4 Results of the sensitivity analysis of the bootstrap Granger causality test for the number and length of the time series and the correlation of the errors using DGP2 22

4.5 Results of the sensitivity analysis of the bootstrap Granger causality test for different values of λ_{ridge} 25
4.6 Results of the sensitivity analysis of the bootstrap Granger causality test for t-distributed errors 27

5 Conclusion **30**

Bibliography **32**

Appendix **36**

A Glossary **37**

B Summary of my thesis in Hungarian **43**

List of Figures

4.1 Size-power curves of the bootstrap Granger causality test (red line) and the standard Wald test (blue line) for every combination of $K = (10, 20, 40)$ and $T = (100, 200, 400)$. (Source: own construction.) 18

4.2 Size-power curves of the bootstrap Granger causality test (red line) and the standard Wald test (blue line) with $\rho = 0.3$ for $K = (10, 20, 40)$ and $T = 100$. (Source: own construction.) 20

4.3 Size-power curves of the bootstrap Granger causality test (red line) and the standard Wald test (blue line) with $\rho = 0.6$ for $K = (10, 20, 40)$ and $T = 100$. (Source: own construction.) 20

4.4 Size-power curves of the bootstrap Granger causality test (red line) and the standard Wald test (blue line) with $\rho = 0.9$ for $K = (10, 20, 40)$ and $T = 100$. (Source: own construction.) 21

4.5 Size-power curves of the bootstrap Granger causality test with $\rho = 0.2$ (red line), with $\rho = 0.4$ (blue line), with $\rho = 0.6$ (green line) and with $\rho = 0.8$ (black line) for every combination of $K = (10, 20, 40)$ and $T = (100, 200, 400)$. (Source: own construction.) 22

4.6 Size-power curves of the bootstrap Granger causality (red line) and the Wald test (blue line) for every combination of $K = (10, 20, 40)$ and $T = (100, 200, 400)$ with no correlation, simulated from DGP2. (Source: own construction.) 23

4.7 Size-power curves of the bootstrap Granger causality (red line) and the Wald test (blue line) for every combination of $K = (10, 20, 40)$ and $T = (100, 200, 400)$ with correlation = 0.7, simulated from DGP2. (Source: own construction.) 24

4.8 Size-power curve of the bootstrap Granger causality test with the setting of $K = 10$ and $T = 100$ with different values of λ_{ridge}. (Source: own construction.) 26

4.9 Size-power curves of the bootstrap Granger causality test with t-distributed error terms: $df = 3$ (red line), $df = 5$ (blue line) and $df = 10$ (green line) for every combination of $K = (10, 20, 40)$ and $T = (100, 200, 400)$. (Source: own construction.) 27

4.10 Size-power curves of the bootstrap Granger causality test with t-distributed errors, where $df = 10$ (red line) and with normally distributed errors (blue line) for every combination of $K = (10, 20, 40)$ and $T = (100, 200, 400)$. (Source: own construction.) 29

List of Tables

4.1 simulated sizes for the Wald and bootstrap Granger causality tests. (Source: own calculation.) 17

Chapter 1

Introduction

In recent years, researchers devoted more attention to high-dimensional time series modelling, as more data are available and more complex methods are needed to interpret them. Although we can access information easily, dimensionality has become a common issue. A model is high-dimensional if the number of time series is large relative to the length of the time series (*Wilms & Croux, 2016*).

In the case of high-dimensional networks, vector autoregressive (VAR) models and Granger causality analysis are applied in various papers. Important directions to handle dimensionality in VAR models are factor models, bayesian methods and regularized or penalized estimations of sparse VARs (*Hecq, Margaritella & Smeekes, 2020*).

Factor models (e.g. *Stock & Watson, 2002* and *Bernanke, Boivin & Elliasz, 2005*) and bayesian VAR models (e.g. *Banbura, Giannone & Reichlin, 2010*) are used in many papers, but their disadvantage is that the result of these models are difficult to interpret compared to penalized methods. Lately, regularized methods have gained attention, like the estimation of sparse VAR models. These are based on the lasso as a regression shrinkage method (*Tibshirani, 1996*) and the elastic net (*Zou & Hastie, 2005*), which is a data-driven method selecting a subset of the coefficients to zero in order to aim sparsity.

Methods which use regularized estimations have many advantages. On the one hand, these are easier to interpret, as the irrelevant variables' coefficient becomes exactly zero. On the other hand, shrinkage methods are useful for variance reduction too, which improves the forecast performance. Furthermore, sparse approaches can be used in case of high-dimensional datasets too, while standard, not penalized methods cannot be applied when the number of time series exceeds the length of the time series (*Wilms & Croux, 2016*).

A lot of progress has been made related to the theory of high-dimensional VAR models. Kock and Callot (*2015*) show for VAR models that under certain conditions, estimation of the non-zero coefficients with adaptive lasso is asymptotically equivalent to the oracle assisted least squares estimator. So the method finds the correct sparsity pattern and makes the same estimations for the coefficient of the relevant variables as the OLS, including only the relevant variables. Medeiros and Mendes (*2016*) extend the mentioned result. Additionally, they show that adaptive lasso has the oracle property even with non-Gaussian and conditionally heteroskedastic error terms. The method performs well with highly correlated regressors and t-distributed errors, which is common in financial and macroeconomic applications. Basu and Michailidis (*2015*) show that it is possible to estimate consistently in high-dimensional settings via $\mathscr{L}_1$-regularization under sparsity constraints for a large class of stable processes. Masini, Medeiros and Mendes (*2019*) demonstrate that in the case of weakly sparse VAR models, the lasso estimation has the oracle property with heavy tailed, weakly dependent innovations too. It is essential, because many volatility processes used in financial applications of VAR models satisfy the stated assumptions (*Hecq et al., 2020*).

High-dimensional VAR models became more popular, and various authors applied these models (*Song & Bickel, 2011*, *Audrino & Camponovo, 2018* and *Wong, Li & Tewari, 2020*). For example, Nicholson, Matteson and Bien (*2017*) introduced a vector autoregression model with exogenous variables, where unmodeled variables are also allowed to be included. Barbaglia, Croux and Wilms (*2020*) used a t-lasso VAR model to account for the fat-tailed distribution of the error terms.

Authors working with high-dimensional models often focus on forecasting and do not account for causality. The vector error correction model (e.g. *Lütkepohl, 2007*) is a tool for estimating and testing for cointegration relationships. Wilms and Croux (*2016*) developed a penalized maximum likelihood approach, designed for sparse estimation of cointegrating vectors and they compare it to Johansen's (*1988*) maximum likelihood method. The latter has many limitations if the number of time series is large compared to the time series length. If the number of time series exceeds the time series length, Johansen's test cannot even be used. The authors show that their method outperforms Johansen's approach significantly in high-dimensional settings.

If the model is used for causality testing instead of forecasting, performing inference is very important, but it is a non-trivial issue (*Hecq et al., 2020*). If we select the relevant variables with a lasso estimator, and then apply an OLS model with the variables with non-zero coefficients, we ignore the uncertainty of the first step selection procedure (*Leeb & Pötscher, 2005*). Nowadays, several authors developed post-selection inference methods, but Hecq et al. (*2020*) highlight some disadvantages of these models.

Skripnikov and Michailidis (*2019*) develop a joint regularized modelling framework to estimate multiple Granger causal networks. They use a shrinkage method in a VAR framework but focus on estimation instead of testing. Song and Taamouti (*2019*) investigate the causal structure in a multivariate time series. They test indirect and spurious causal effects with different statistical procedures. The paper is based on big data analysis, but they focus on factor models instead of penalized methods. Krampe, Kreiss and Paparoditis (*2018*) introduce bootstrap methods to infer the properties of sparse VAR models. This method uses a model-based bootstrap procedure that generates pseudo time series and de-sparsifies the VAR model. Chaudhry, Xu and Gu (*2017*) introduce an asymptotically unbiased Granger causality estimator with corresponding test statistics and confidence intervals. They also develop a false discovery rate control method that outperforms previous techniques related to power in multiple testing. Hecq et al. (*2020*) develop a post-double-selection procedure and present a valid post-selection Granger causality test in high-dimensional VAR framework. They take the approach of Belloni, Chernozhukov and Hansen (*2014b*) as a basis. With this post-double-selection method, one can get sharper conclusions than by applying standard low-dimensional VAR techniques. However, in the case of regularized estimations, the method of tuning parameter selection is very important. As the approach of Hecq et al. (*2020*) is very sensitive to the selection of the tuning parameter, I decided to examine the method of Wilms, Gelper and Croux (*2016*) in the simulation study.

In this paper, I apply a bootstrap Granger causality test in high-dimensional VAR model, based on adaptive lasso (*Zou, 2006*) and developed by Wilms et al. (*2016*). Chatterjee and Lahiri (*2011*) developed a residual bootstrap procedure for high-dimensional cross-section dataset. Wilms et al. (*2016*) extend the mentioned procedure to high-dimensional time series data. In the study, they identify those industry segments which have statistically significant predictive power for future macroeconomic developments. They identify these segments with a bootstrap test statistic based on adaptive lasso. They compare their method with the standard Wald test. Based on the simulation, their test statistic outperforms the Wald test statistic in high-dimensional settings.

I make a similar comparison of the mentioned bootstrap Granger causality test with Wald test, but using different data generating processes, stated in Chapter 4. The contribution of my paper to the study of Wilms et al. (*2016*) are the followings:

- The data generating processes I use are selected based on financial applications.
- I examine in detail the effect of the number of time series and the length of the time series on the results with every combination of three-three different settings.
- I study how sensitive the simulations are for different covariance matrices of the error terms. Based on Hecq et al. (*2020*) it is important to examine it.
- I analyse the effect of generating error terms from (fat-tailed) t-distribution, which is common in financial applications.

The remainder of this paper is structured as follows. Chapter 2 introduces the high-dimensional VAR models and Granger causality tests. In chapter 3 I propose the estimation and the inferential framework and describe the bootstrap Granger causality test. Chapter 4 establishes the simulation study and reports the results. Chapter 5 concludes.

Chapter 2

High-dimensional Granger causality testing

Primarily, it is needed to define Granger causality. Let Ω be a given information set and let X and Y be a variable. Both X and Ω are observed prior to Y. If we add X to Ω and therefore the conditional distribution of Y alters, thus X improves the predictability of Y, or in other words, X Granger cause Y with respect to Ω (*Granger, 1969, 1980*).

When building a model, we have to pay attention to spurious Granger causality. It occurs when a variable Z Granger causes variable X and variable Y too, but we omit Z from Ω and, therefore, X seems to Granger cause Y. To remedy this problem, when we previously determine which variables should be in Ω, we should select all of the potentially relevant variables. Thus, the more potentially relevant variable is included in Ω, the better, although it can easily cause a high-dimensional dataset (*Hecq et al., 2020*).

In this paper I apply a bootstrap Granger causality approach in a VAR framework. If we increase the number of variables in a VAR model, the number of parameters increases quadratically with the number of time series included in the model. Considering an unrestricted VAR(p) model, where p is the lag-length and the number of time series included is denoted with K, we have to estimate K^2p coefficients. In empirical applications, the time series length is usually relatively small. In this case, standard least squares and maximum likelihood methods can easily overfit the data. In addition, standard statistical methods cannot be used if the number of time series exceeds the length of the time series.

One possible solution is to apply penalized approaches (*Hecq et al., 2020*).

2.1 Granger causality testing in VAR framework

Let $y_1, ..., y_T$ be a K-dimensional multiple weakly stationary time series process. I assume that $y_t = (y_{1,t}, ..., y_{K,t})'$ is generated by a VAR(p) process, where p is the lag-length

$$y_t = A_1 y_{t-1} + ... + A_p y_{t-p} + u_t, \quad t = p+1, ..., T, \tag{1}$$

where I assume that all of the time series are mean centered, so intercept is not included. $A_1, ..., A_p$ are $K \times K$ parameter matrices, and u_t is a martingale difference sequence of error terms.

Assumption 1. The VAR model in (1) satisfies the followings:

- $\{u_t\}_{t=1}^T$ is a weakly stationary martingale difference sequence with respect to $\mathcal{F}_t = \sigma(y_t, y_{t-1}, y_{t-2}, ...)u_t$ such that $\mathbb{E}(u_t|\mathcal{F}_{t-1}) = 0$ for all t and $\Sigma_u = \mathbb{E}(u_t u_t^{'})$ is a positive definite.
- All roots of $\det(I_K - \Sigma_{j=1}^p A_j z^j)$ lie outside the unit disc, such that the lag polynomial is invertible (*Hecq et al., 2020*).

To make it more simple, we can write model (1) in matrix notation (*Wilms et al., 2016*) as

$$y = \beta X + u, \tag{2}$$

where y is the $(KT \times 1)$ vector $(y_{1,1}, ..., y_{K,1}, y_{1,2}, ..., y_{K,2}, ..., y_{1,T}, ..., y_{K,T})'$. The matrix $X = (\underline{X}_1, ..., \underline{X}_p)$, where $\underline{X}_j$ is $(KT \times K)$, contains the values of the time series at lag j in its columns, for $1 \leq j \leq p$. β is the $(pK \times K)$ matrix of the coefficients $A_1, ..., A_p$.

I would like to test if the time series in the set J Granger cause the time series in the set I in mean, conditional on the other variables in the VAR model (2), where $J, I \subset \{1, ..., K\}$ and $J \cap I = \emptyset$. Let $N_I = |I|$ and $N_J = |J|$ denote the number of time series in the sets I and J, respectively. The researcher can determine N_J and N_I based on the application, and often $N_J = N_I = 1$, so both sets consist of one time series (*Hecq et al., 2020*). In the simulation study, I also determine $N_J = N_I = 1$.

It can be assumed that the lag-length (p) is small since, in the case of univariate regressions or small systems, the need for a large p is generally caused by omitted variables (*Hecq, Laurent & Palm, 2016*). As I mentioned, the omitted variable bias can be prevented by adding all of the possibly relevant variables to the model. Thus, a smaller value for p is realistic. It can result in a high-dimensional VAR model, but one can handle it with regularized estimators. Although the value of p is usually unknown in practice, one can give a small upper bound on p, which bound depends on the application. Then it is possible to find the appropriate value of p with an algorithm. Otherwise, p has to be estimated (*Hecq et al., 2020*).

Chapter 3

Post-selection inference

This chapter introduces the bootstrap Granger causality approach of Wilms et al. (*2016*), but applied in a VAR model instead of AR framework. First, I present the Penalized Maximum Likelihood estimator of the coefficients. Then, I discuss the post-selection problems related to inference.

3.1 Penalized Maximum Likelihood estimation

If $Kp > T$, the Maximum Likelihood estimator is not computable, but we still can use the Penalized Maximum Likelihood estimator. It is an important assumption that β is sparse. Thus, it can be estimated with a coefficient vector, which has a significant portion of the coefficients equal to zero. Since we assume sparsity, we can reduce the dimension of the model with regularized methods without ruining the predictability (*Hecq et al., 2020*).

We can obtain the Penalized Maximum Likelihood estimator of the regression coefficient β by minimizing the negative log-likelihood with a penalization on the elements of β:

$$\hat{\beta}_\lambda = \underset{\beta}{argmin}\left(\frac{1}{T}(y - \beta X)'(y - \beta X) + \lambda \sum_{i=1}^{p(1+k)} \hat{w}_i |\beta_i|\right), \tag{3}$$

where $\lambda > 0$ is a non-negative sparsity parameter determining the strength of the penalty, and $\hat{w}_i$ are (non-negative) weights that belong to the i-th β (*Wilms et al., 2016*). In case of standard lasso, the weights would be one for the penalized and zero for the not penalized parameters. There are more methods to determine which weights should be zero and which should be one. Hecq et al. (*2020*) use information criteria to determine it. As I use adaptive lasso (*Zou, 2006*), the weights are parameter specific and can take other values too. Chatterjee and Lahiri (*2011*), show that with a bootstrap procedure, the distribution and variance of the adaptive lasso estimator can be estimated consistently, so I use adaptive lasso in the simulation. Although, in this application, the oracle property is not so important because I do not need to identify all of the unnecessary variables. The aim is to eliminate the effect of other unnecessary variables on the relation between the variables tested for Granger causality (*Hecq et al., 2020*). Based on Wilms et al. (*2016*) we can get the weights as

$$\hat{w}_i = \frac{1}{\left|\hat{\beta}_i^{ridge}\right|},$$

where the Ridge estimator is the following:

$$\hat{\beta}_\lambda^{ridge} = \underset{\beta}{argmin}\left(\frac{1}{T}(y-\beta X)'(y-\beta X) + \lambda_{ridge}\sum_{i=1}^{p(1+k)}\beta_i^2\right),$$

The notation $\hat{\beta}_\lambda$ in (3) highlights that the minimization problem's solution depends on the parameter λ. I select the sparsity parameter λ of (3) using the Bayesian information criterion (BIC), and I also summarize the methods for tuning parameter selection in Section 3.2.

3.2 Tuning parameter selection

There are more possibilities for selecting the appropriate λ and it is essential to achieve a good model. Hecq et al. (*2020.*) compare some popular methods of tuning parameter selection. They highlight three methods. One of these is the cross-validation. Although it is a very popular method, it is not always effective in the case of time series, because of the inherent serial correlation and the potential non-stationarity of the data (*Bergmeir, Hyndman & Koo, 2018*).

Another method that I summarize briefly is to use estimates of theoretically optimal values (e.g. *Belloni & Chernozhukov, 2013*). For example, Hecq et al. (*2020*) estimate the value of σ (the variance of the error term) based on Belloni et al. (*2012*). Firstly, they get an initial least squares estimation of y (the response vector), and then, they update the estimation iteratively.

The third method is to minimize an information criterion (IC) in order to determine the appropriate λ, which is also a data-driven method. Hecq et al. (*2020*) compare three IC in a simple way. Let $\hat{S}(\lambda) = \{m \in \{1, ..., K_p\} : |\hat{\beta}_m(\lambda)| > 0\}$ denote the set of variables with non-zero coefficient for given λ in the solution of lasso. If y is the generic response vector and X is the predictor matrix, the value of λ^{IC} can be written as

$$\lambda^{IC} = \underset{\lambda}{argmin} \ln\left(\frac{1}{T}\left\|y - X\hat{\beta}(\lambda)\right\|_2^2\right) + \frac{C_T}{T}\left|\hat{S}(\lambda)\right|,$$

where C_T is the penalty specific to each IC. In case of the Akaike information criterion (AIC) (*Akaike, 1974*), $C_T = 2$, in case of the Bayesian information criterion (BIC) (*Schwarz, 1978*), $C_T = \ln(T)$, and in case of the Extended Bayesian information criterion (EBIC) (*Chen & Chen, 2008*), $C_T = \ln(T) + 2\gamma\ln(Kp)$, where $\gamma = 0.5$ based on (*Chen & Chen, 2012*). They show that AIC and BIC cannot select the correct variables in high-dimensional systems while EBIC remains consistent.

In the simulation study, I select the sparsity parameter λ using BIC. Although it is not consistent in high-dimensional settings, I use it because I compare the Penalized Maximum Likelihood method with the Wald test, which is only applicable in low-dimensional settings. I solve (3) over a range of values for λ, and I select the one with the minimal BIC value.

3.3 Post-selection inference

In this section, I explain why post-selection inference is needed, and I summarize some possible solutions. If we only apply a lasso or adaptive lasso procedure and set the weights of the Granger causing variables to zero and then test whether the coefficient of these variables is equal to zero (estimated again the model by least squares method on the selected variables), we do not take account of the fact that the final model depends on the data, so it is an overfitted model. Thus the selection step depends on the data, and it can occur that we omit weakly relevant variables that are difficult to detect, but cause too large omitted variable bias that cannot be ignored asymptotically, so the inference is not uniformly valid (*Hecq et al., 2020*). The post-selection estimators converge only point-wise to the normal distribution, instead of uniformly (*Leeb & Pötscher, 2005*). Post-selection based on oracle properties does not cause inferential problems in the case of ruling out beta-min conditions before the estimation. By applying beta-min conditions (which refer to the size of the coefficients), we can exactly separate the non-zero coefficients from the zero ones (*Van de Geer & Bühlmann, 2011*). However, it is difficult to justify in empirical analysis (*Hecq et al., 2020*).

Recently, researchers developed a couple of methods to valid post-selection inference (e.g. *Berk et al., 2013*, *Lee et al., 2016* and *Van de Geer et al., 2014*). Hecq et al. (*2020*) extend the approach developed by Belloni, Chernozhukov and co-authors (*2014a*) to dependent data.

Uniform inference for treatment effects in the case of partially linear, penalized models with high-dimensional controls can be obtained using a post-double-selection approach developed by Belloni, Chernozhukov and Kato (*2015*). It consists of two steps. They estimate the coefficients of both the outcome and the treatment variables on all of the control variables. Then they apply a post-selection least squares estimation of the outcome on the treatment variable. All the control variables are selected at least once in the two steps. This approach considerably reduces the omitted variable bias, and in the final model, the errors are orthogonal with respect to the treatment variable. The procedure is valid in the case of heteroskedastic and non-Gaussian error terms too.

The method of Belloni et al. (*2015*) is extended by Chernozhukov et al. (*2020*) with (weak) temporal and cross-sectional dependency. They apply penalization iteratively in the system, and they choose the overall penalty by a block multiplier bootstrap procedure. The procedure's oracle property and the bootstrap consistency are proved, and they obtain simultaneous valid inference.

The method of Hecq et al. (*2020*) is similar to the approach of Chernozhukov et al. (*2020*), but there are several differences. First, it can be applied faster and does not consist of bootstrap methods. Second, it is developed specifically for VAR models and Granger causality testing, while the method of Chernozhukov et al. (*2020*) is developed for general systems of equations. Third, as Hecq et. al (*2020*) focus on applications related to financial econometrics, they consider different assumptions to establish a valid method.

3.4 Bootstrap Granger causality test

If $X_{t,j}$ is the vector of the j-th time series for $t = 1, ..., p$, with the coefficient at lag i of $a_{i,j}$, where $i = 1, ..., p$. As I already explained at the beginning of Chapter 2, the multivariate time series $X_{t,j}$ Granger cause Y_t, if the former has incremental predictive power for the latter. If the coefficients on all lags of $X_{t,j}$ are equal to zero, $X_{t,j}$ does not Granger cause Y_t. Thus $a_{1,j} = ... = a_{p,j} = 0$.

The adaptive lasso estimator in (3) is sparse if it has elements equal to zero beside the non-zero ones. For larger values of λ the estimator is sparser (*Wilms et al., 2016*). Based on the "Granger lasso selection" approach (e.g. *Bahadori & Liu, 2013*), a time series $X_{t,j}$ Granger causes Y_t if it has at least one non-zero estimated coefficient. j refers to the j-th time series. The approach of Wilms et al. (*2016*) differs from it because they infer Granger causality relations from a bootstrap testing procedure.

The null hypothesis is that $X_{t,j}$ not Granger causes Y_t:

$$H_0 : R_j\beta = 0, \tag{4}$$

where R_j is a suitable $pj \times p(1 + K)$ matrix. The elements of R_j can be zero or one. If the j-th element of R_j is equal to one, it means that the mentioned element is an autoregressive parameter of $a_{1,j}, ..., a_{p,j}$. The corresponding Wald test statistic is given by

$$Q = (R_j\hat{\beta})'(R_jCov(\hat{\beta})R_j')^{-1}(R_j\hat{\beta}). \tag{5}$$

I use a residual bootstrap procedure to bootstrap this test statistic, which consists of the following three main steps (*Kreiss & Lahiri, 2012*):

1. Estimate model (1) under the null hypothesis with the time series $X_{t,j}$ removed and compute the centred residuals $\hat{u}_t$, for $t = 1, ..., T$.

2. For $b = 1, ..., B$, where $B = 500$ is the number of bootstraps:

(a) Let y_t^* the bootstrap time series. Construct y_t^* from model (1) with the estimated coefficients of the previous step and with bootstrap errors $u_t^* = \hat{u}_{U_t}$ with U_t for $t = 1, ..., T$ an independent and identically distributed (i.i.d.) sequence of discrete random variables uniformly distributed on $1, ..., T$.

(b) Apply the PML estimator of equation (3) to the bootstrap sample and denote the coefficients estimated with the bootstrap procedure by $\hat{\beta}_b^*$.

(c) Compute the bootstrap test statistic $Q_b^* = (R_j \hat{\beta}_b^*)'(R_j Cov(\hat{\beta}) R_j')^{-1}(R_j \hat{\beta}_b^*)$.

3. Compute mid p-Value which is equal to $\frac{1}{B}\sum_{b=1}^{B}\big(I(Q_b^* < Q) + \frac{1}{2}I(Q_b^* = Q)\big)$, with Q_b^* B independent bootstrap test statistics, for $b = 1, ..., B$. $I(.)$ is an indicator function. If its argument is true, it takes on the value one, and zero otherwise (*Wilms et al., 2016*).

Chapter 4

Simulation study

This chapter provides important information about the simulations and presents the results. First, I introduce the data generating processes I used in this paper. Second, I present the effect of the length of the time series and the number of time series to the results. Third, I show the results of the sensitivity analysis of the simulations with different correlation of the error terms. Fourth, I present the results of the sensitivity analysis of the simulations with different values of λ_{ridge}. Last, I present the effect of generating error terms from t-distribution.

4.1 Simulations

By means of the simulation study, I evaluate the finite-sample performance of the proposed bootstrap Granger causality test. I evaluate the performance with a size and power analysis (e.g. *Wilms et al., 2016* and *Hecq et al., 2020*). The null hypothesis is that the second time series does not Granger cause the response. I test the H_0 and compare the performance of the proposed bootstrap Granger lasso test to the standard Wald test, where the second one is computed from the standard Maximum Likelihood (ML) estimator, similarly to Wilms et al. (*2016*).

First, I study the size of the test statistic. After simulating $N = 200$ time series under the H_0, I compute the simulated size, thus I compute the proportion of simulations where the null hypothesis is rejected with the following formula:

$$\frac{1}{N}\sum_{j=1}^{N} I(p_j^{H_0} < \alpha),$$

where $p_j^{H_0}$ for $j = 1, ..., N$ is the mid p-Value. It is obtained in the j-th simulation run. α is the significance level, which is pre-specified, and I consider $\alpha = 0.01$ and $\alpha = 0.05$.

Second, I study the power of the test statistic with size-power curves. After constructing two empirical distribution functions, I followed the next three steps:

1. Simulate $N = 200$ time series under the H_0. Compute the mid p-Value $p_j^{H_0}$ for the j-th simulation run, where $j = 1, ..., N$. Calculate the empirical distribution function of the mid p-Values with the following formula:

$$\hat{F}^{H_0}(x_i) = \frac{1}{N}\sum_{j=1}^{N} I(p_j^{H_0} \leq x_i),$$

   for a grid of values between zero and one of x_i, for $i = 1, ..., m$.

2. Simulate $N = 200$ time series under the alternative hypothesis H_A. Compute the mid p-Value $p_j^{H_A}$ for the j-th simulation run, where $j = 1, ..., N$. Calculate the empirical distribution function of the mid p-Values with the following formula:

$$\hat{F}^{H_A}(x_i) = \frac{1}{N}\sum_{j=1}^{N} I(p_j^{H_A} \leq x_i).$$

3. Plot $\hat{F}^{H_0}(x_i)$ against $\hat{F}^{H_A}(x_i)$ for x_i, where $i = 1, ..., m$ (*Wilms et al., 2016*).

I consider two Data Generating Processes (DGPs) inspired by Hecq et al. (*2020*) and Kock and Callot (*2015*).

$$DGP1 : y_t = \begin{bmatrix} 0.5 & 0 & \dots & 0 \\ 0 & 0.5 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0.5 \end{bmatrix} y_{t-1} + \epsilon_t,$$

$$DGP2: y_t = \begin{bmatrix} A & 0 & \dots & 0 \\ 0 & A & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & A \end{bmatrix} y_{t-1} + \epsilon_t, \text{with} \underbrace{A}_{5x5} = \begin{bmatrix} 0.15 & 0.15 & \dots & 0.15 \\ 0.15 & 0.15 & \dots & 0.15 \\ \dots & \dots & \dots & \dots \\ 0.15 & 0.15 & \dots & 0.15 \end{bmatrix}.$$

As in the study of Hecq et al. (*2020*), DGP1 is a diagonal VAR and DGP2 is a block-diagonal system. Both respect the sparsity assumptions. Thus both have zero elements besides the non-zero ones. In the case of DGP1, each time series Granger cause only themselves, while DGP2 is a block-diagonal system. This structure is motivated by sectoral relations or quarterly macroeconomic models. For both designs, I consider a DGP under H_0 and under H_A. While DGP1 satisfies the null hypothesis, that is, unit 2 does not Granger cause unit 1, DGP2 does not satisfy. For the power analysis of DGP1, I set the coefficients of (2, 1) equal to 0.2. For DGP2, for the size analysis, I set the same coefficient equal to zero instead of 0.15, which was previously set. Therefore, I can evaluate the performance with a size and power analysis in the case of both DGPs.

I examine the case of having a single variable of interest for both the size and the power test, so I choose $I = \{2\}$ and $J = \{1\}$. As I already mentioned, I determine $p = 1$ lag, which is the same as in the DGPs ($j = 1$), since, in the case of univariate regressions or small systems, the need for large p is generally caused by omitted variables (*Hecq et al., 2016*). Thus, the equation is the following:

$$y_{2,t} = \beta_{GC} y_{1,t-1} + \sum_{j=2}^{K} \beta_j y_{j,t-1} + \epsilon_{2,t},$$

where the index GC refers to Granger causing. For both DGPs I test $H_0 : \beta_{GC} = 0$ against $H_A : \beta_{GC} \neq 0$ using the bootstrap Granger causality and the standard Wald test.

4.2 Results of the sensitivity analysis of the bootstrap Granger causality test for the number and length of the time series

Correlation = 0							
	T	100		200		400	
	K	Wald	Bootstrap	Wald	Bootstrap	Wald	Bootstrap
Size	10	0.035	0.01	0.025	0.01	0.05	0.005
1%	20	0.02	0.025	0.025	0.01	0.055	0.01
	40	0.02	0.025	0.04	0.02	0.04	0.01
Size	10	0.15	0.065	0.08	0.04	0.13	0.035
5%	20	0.1	0.07	0.095	0.045	0.155	0.08
	40	0.09	0.08	0.095	0.085	0.095	0.06
Power	10	0.65	0.635	0.91	0.88	0.985	0.985
	20	0.54	0.58	0.875	0.875	0.995	0.995
	40	0.35	0.53	0.775	0.875	0.995	0.995

Table 4.1: simulated sizes for the Wald and bootstrap Granger causality tests. (Source: own calculation.)

Table 4.1 shows the size and power of the bootstrap Granger causality test and the Wald test for $N = 200$ by using all of the nine combinations of the number of time series $K = (10, 20, 40)$ and time series length $T = (100, 200, 400)$. The lag-length is fixed, $p = 1$. The burn-in period contains 100 observations. The simulated sizes of the two tests are similar to the results of Wilms et al. (*2016*) but not so close to the nominal size α in every case as I expected, and somewhere the values do not increase with the number of time series monotonically. Maybe because I only simulated $N = 200$ instead of $N = 1000$, as in the study of Wilms et al. (*2016*), in order to reduce the computational time.

Although, almost for every given K and T value, the bootstrap Granger causality test is closer to the nominal α than the Wald test, except for $T = 100$ and $K = (20, 40)$. I can observe a deterioration in the results with a fixed value of K and decreasing values of T and also with a fixed value of T and increasing values of K, but not in every case. Both tests have really similar results for power for $T = 400$, but the smaller the value of T, the bigger the difference between the bootstrap Granger causality test and the standard

Wald test. The difference is even bigger if the value of K increases, as I expected. Thus, the bootstrap Granger causality test achieves larger power than the Wald test.

Fig. 4.1 reports the size-power curves of the bootstrap Granger causality test and the standard Wald test. In this figure, I demonstrate the effect of the length of time series and the effect of the number of time series on the performance of the bootstrap Granger causality and the standard Wald test. The larger the difference between the size-power curve and the 45° line, the more power the test has (*Wilms et al., 2016*).

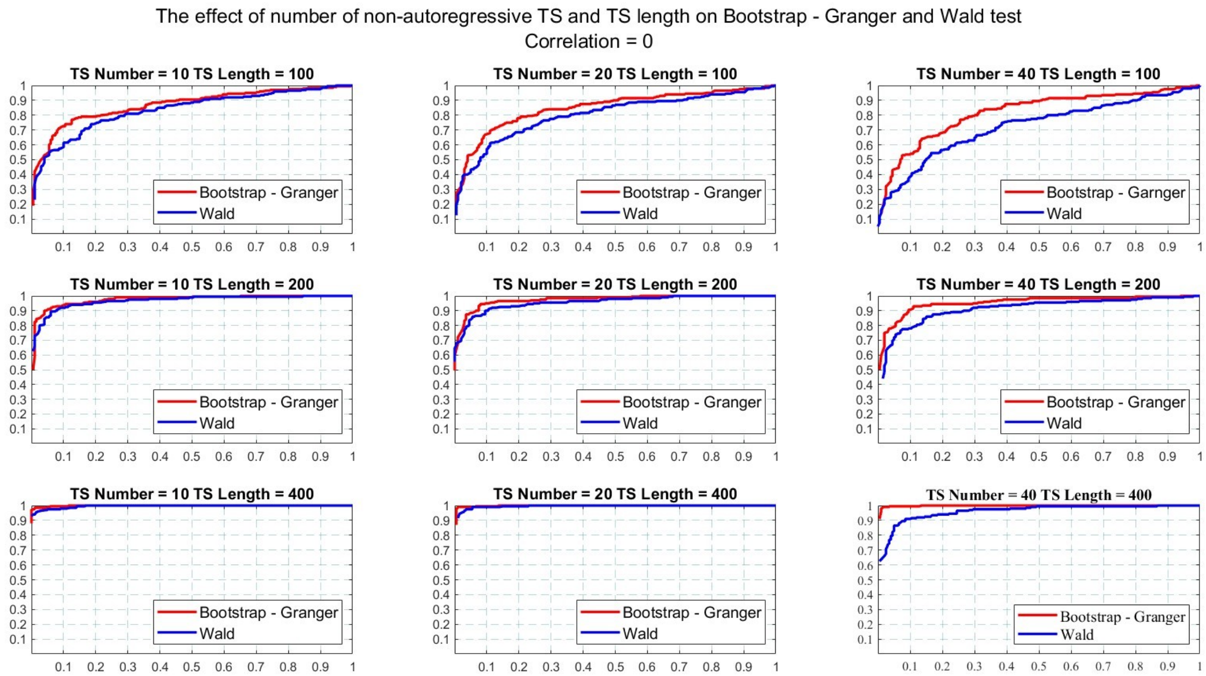

Figure 4.1: Size-power curves of the bootstrap Granger causality test (red line) and the standard Wald test (blue line) for every combination of $K = (10, 20, 40)$ and $T = (100, 200, 400)$. (Source: own construction.)

I can observe that for a fixed value of K and increasing values of T, both tests have better results and also for a fixed value of T and decreasing values of K, as I expected. For the most low-dimensional settings, thus for $T = 400$ and $K = 10$ and for $T = 400$ and $K = 20$ both tests perform very well. In these cases, both curves lie almost on the left side and the top of the diagram, which would mean perfect performance without any mistakes. Although, in these two cases, the methods have similarly good results, the bootstrap Granger causality test outperforms the Wald test in these settings too. For the most high-dimensional setting, thus, for $T = 100$ and $K = 40$ none of the two tests perform as well.

When the time series length is large relative to the number of time series, the difference between the two tests is small. Although, the larger the number of time series relative to the length of the time series, the bigger the difference between the power of the two tests for the benefit of the bootstrap Granger causality test. Based only on the prior results, the bootstrap Granger causality test is a better option than the standard Wald test, mainly in high-dimensional settings but in low-dimensional settings too. Until this point, I made a similar analysis as Wilms et al. (*2016*), but more in detail, and I got really similar results.

4.3 Results of the sensitivity analysis of the bootstrap Granger causality test for the correlation of the error terms

Table 4.1 and Fig. 4.1 report results with correlation equal to zero. Although, covariance matrices of the error terms have an impact on the efficiency of the lasso. Thus I compare the power of the bootstrap Granger causality test in ten cases for the correlation. I calculate the covariance matrix with a Toeplitz-version as $\Sigma_{i,j} = \rho^{|i-j|}$ as in the study of Hecq et al. (*2020*) by using ten cases of correlation: $\rho = \{0, 0.1, 0.2, 0.3, ..., 0.9\}$. The first case corresponds to the results of Table 4.1 and Fig. 4.1, that is no correlation. It is equivalent to set $\Sigma = I_K$. In the case of other values of ρ, the bigger the distance from the diagonal of the covariance matrix, the smaller the value we get.

Fig. 4.2, fig. 4.3 and fig. 4.4 shows that the bigger the correlation of the error terms, the worse the bootstrap Granger causality test's performance relative to the Wald test's performance. For the lower-dimensional settings, thus for $T = 100$ and $K = 10$ and $K = 20$, the standard Wald test outperforms the bootstrap Granger causality test even for $\rho = 0.3$, and the difference between the efficiency of the two tests becomes larger for larger values of ρ. For the higher-dimensional setting, thus for $T = 100$ and $K = 40$, the bootstrap Granger causality test outperforms the Wald test when $\rho = 0.3$, but when

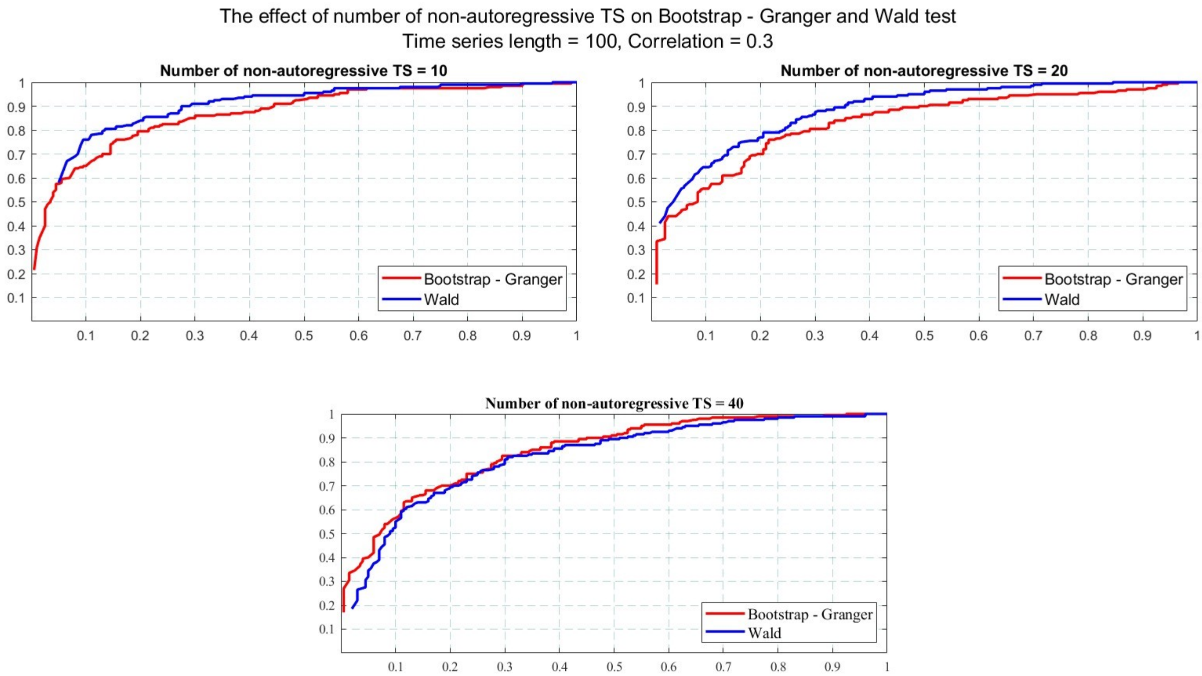

Figure 4.2: Size-power curves of the bootstrap Granger causality test (red line) and the standard Wald test (blue line) with $\rho = 0.3$ for $K = (10, 20, 40)$ and $T = 100$. (Source: own construction.)

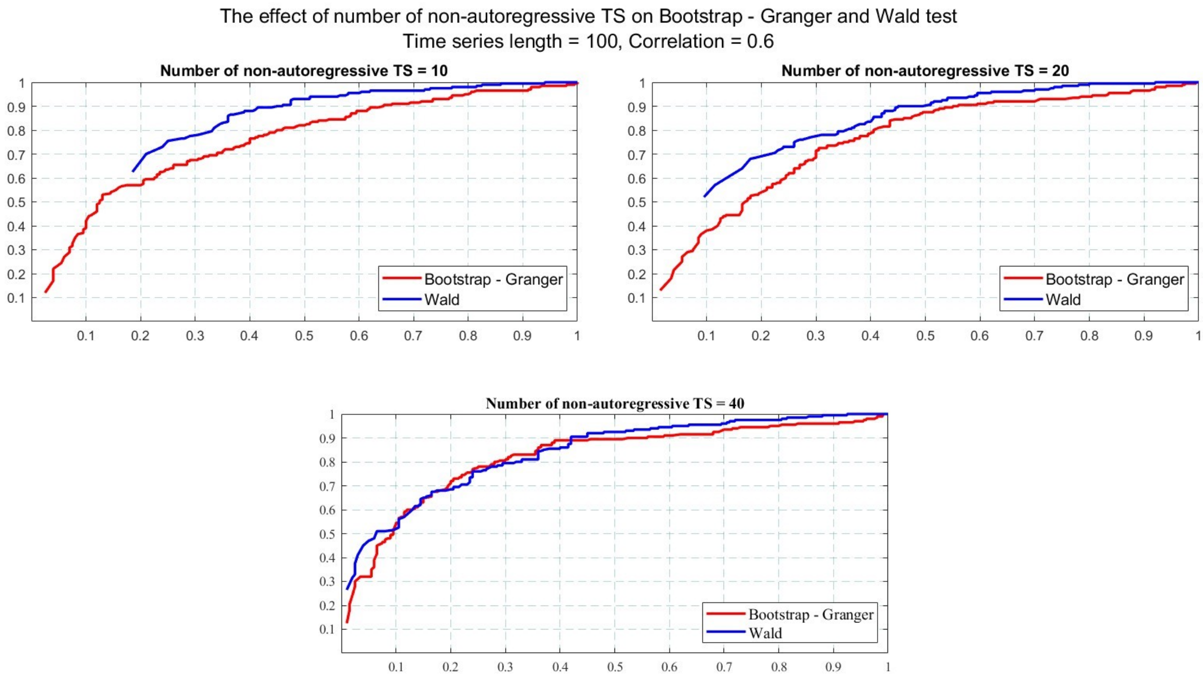

Figure 4.3: Size-power curves of the bootstrap Granger causality test (red line) and the standard Wald test (blue line) with $\rho = 0.6$ for $K = (10, 20, 40)$ and $T = 100$. (Source: own construction.)

$\rho = 0.6$, it is hard to decide which test performs better based on the size-power curves. Therefore, based on comparing the bootstrap Granger causality test with the Wald test for the bigger correlation of the error terms, I would not use the bootstrap Granger causality test in every application. For example, in low-dimensional settings, where the correlation of the error terms is supposedly high, the Wald test can be a better option.

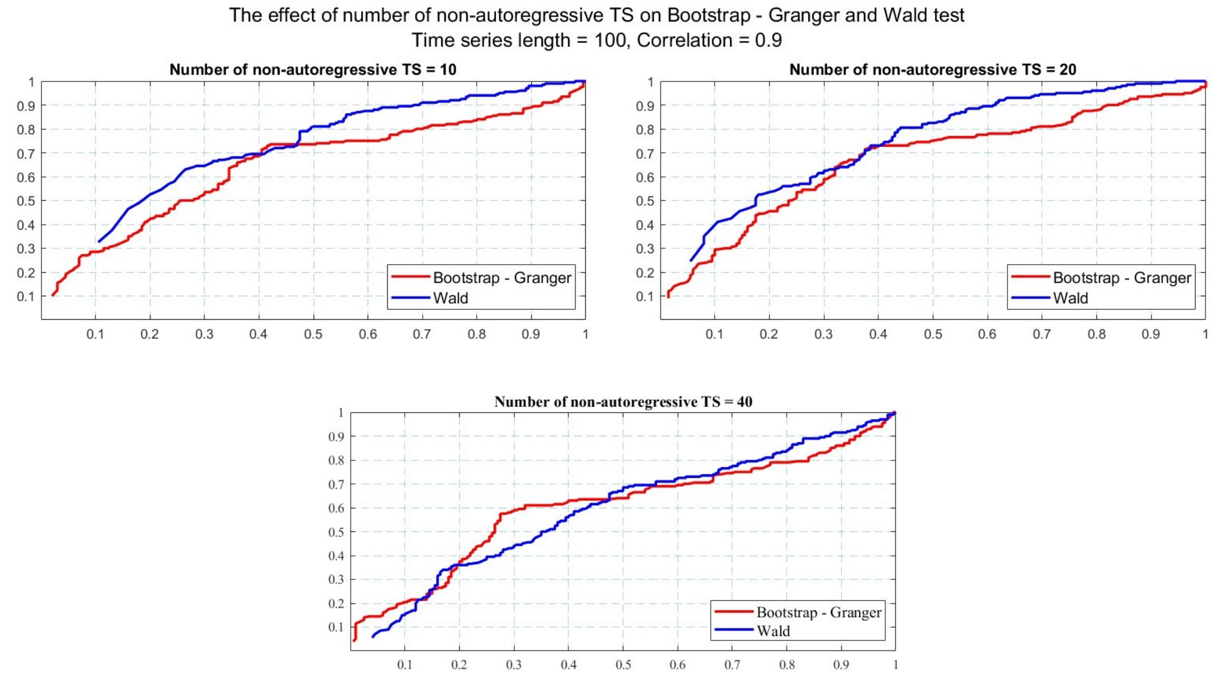

Figure 4.4: Size-power curves of the bootstrap Granger causality test (red line) and the standard Wald test (blue line) with $\rho = 0.9$ for $K = (10, 20, 40)$ and $T = 100$. (Source: own construction.)

Fig. 4.5 shows the difference between the power of the bootstrap Granger causality test with $\rho = \{0.2, 0.4, 0.6, 0.8\}$ for every combination of $T = \{100, 200, 400\}$ and $K = \{10, 20, 40\}$. The sensitivity analysis of the bootstrap Granger causality test for the correlation of the error terms shows that for bigger correlation, the test performs worse in every setting of T and K with some exceptions. Maybe there are some exceptions because I only used 200 simulations. For the most low-dimensional settings, the bootstrap Granger causality test with $\rho = 0.8$ performs relatively bad compared to other, smaller settings of the correlation. The figure shows that the test has worse performance for a fixed value of

K and decreasing values of T. For a fixed value of T and increasing values of K, the test has worse performance too.

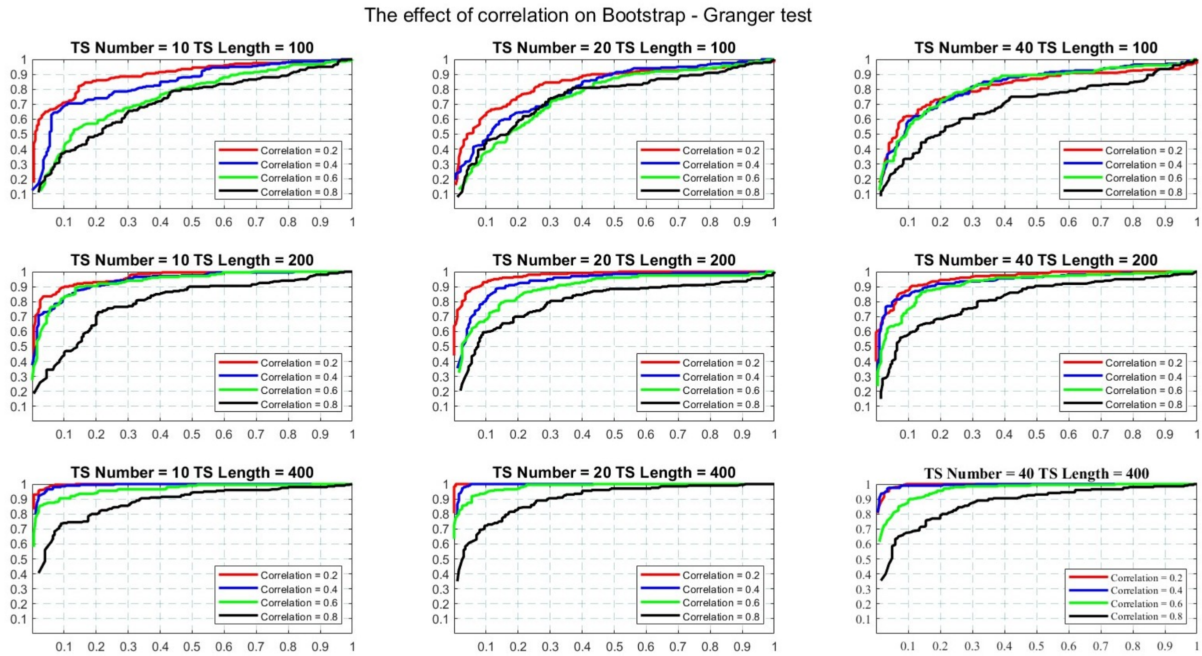

Figure 4.5: Size-power curves of the bootstrap Granger causality test with $\rho = 0.2$ (red line), with $\rho = 0.4$ (blue line), with $\rho = 0.6$ (green line) and with $\rho = 0.8$ (black line) for every combination of $K = (10, 20, 40)$ and $T = (100, 200, 400)$. (Source: own construction.)

4.4 Results of the sensitivity analysis of the bootstrap Granger causality test for the number and length of the time series and the correlation of the errors using DGP2

Fig. 4.6 reports the size-power curves of the bootstrap Granger causality test and the Wald test with no correlation. While until this point I only examined DGP1, here the data was simulated using DGP2. I have the same observations as in the case of Fig. 4.1, where we can see the same curves but simulated with DGP1.

As I expected, for a fixed value K and increasing values of T and also for a fixed value of T and decreasing values of K both tests have better results. For the most low-dimensional settings both tests perform very well. For every combination of $T = (200, 400)$ and $K = (10, 20)$, and for $T = 100$ and $K = 10$, the two tests' performance is really similar. For higher-dimensional settings, the bootstrap Granger Causality test outperforms the Wald test, so I can observe that the difference between the two tests increases for the benefit of the bootstrap Granger Causality test. For the most high-dimensional setting, thus for $T = 100$ and $K = 40$ none of the two tests performs as well, the size-power curves are close to the 45° line. The only difference compared to Fig. 4.1, where I simulated with DGP1 is that both tests perform worse in all of these nine combinations of T and K values without exception.

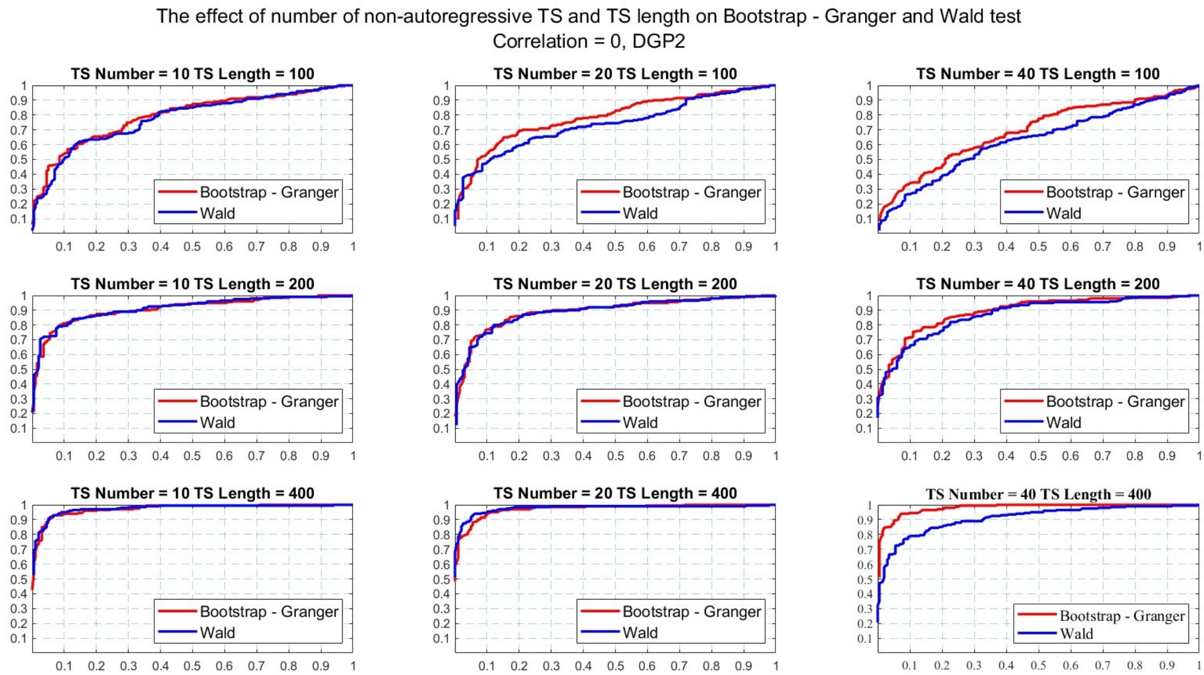

Figure 4.6: Size-power curves of the bootstrap Granger causality (red line) and the Wald test (blue line) for every combination of $K = (10, 20, 40)$ and $T = (100, 200, 400)$ with no correlation, simulated from DGP2. (Source: own construction.)

Fig. 4.7 shows the effect of the number of time series and length of time series on bootstrap Granger causality and Wald test simulating with DGP2, as Fig. 4.6, but the correlation of the error terms is 0.7. Both tests have a deterioration of their performance for a fixed value of T and increasing values of K and also for a fixed value of K and decreasing values of T. Maybe the case of $K = 20$ of the bootstrap Granger causality test is an exception if I look at the curves made with the three settings where $T = 400$, but it can be caused by the low number of simulations.

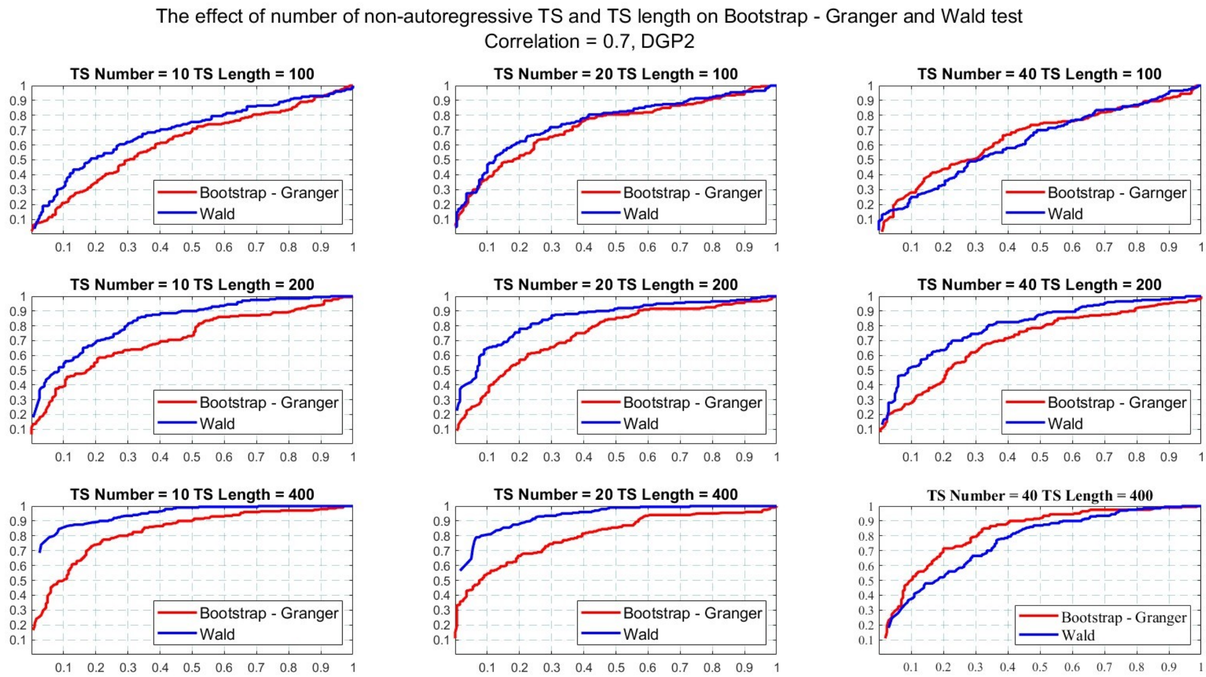

Figure 4.7: Size-power curves of the bootstrap Granger causality (red line) and the Wald test (blue line) for every combination of $K = (10, 20, 40)$ and $T = (100, 200, 400)$ with correlation $= 0.7$, simulated from DGP2. (Source: own construction.)

I can observe that almost in every case of the nine size-power curves, both tests have worse performance than in Fig. 4.6, where there was no correlation. The two exceptions are the case of $T = 100$ and $K = (20, 40)$, so the most high-dimensional cases. In these two cases, the tests have very similar results as in the case of $\rho = 0$.

If I compare the performance of the bootstrap Granger causality test to the Wald test, the curves show that in every case where $K = (10, 20)$, so in lower-dimensional cases, the

Wald test outperforms the bootstrap Granger causality test. Maybe an exception is where $T = 100$, where the two tests have similar results. Although, in higher-dimensional settings it is not obvious which test performs better, because in these simulations, the two tests' performance is quite similar in the case of $T = 100$ and $K = 40$, the Wald test outperforms the bootstrap Granger causality test in case of $T = 200$ and $K = 40$, and finally, the bootstrap Granger causality test outperforms the Wald test in the $T = 400$ and $K = 40$ setting. Thus, in case of strong correlation, I would rather use the Wald test in lower-dimensional settings than the bootstrap Granger causality test, but in higher dimensional settings the bootstrap Granger causality test can be a good choice too.

Based on the previous analysis, in case of no correlation of the error terms, the bootstrap Granger causality test outperforms the standard Wald test in every combination of $T = (100, 200, 400)$ and $K = (10, 20, 40)$ with both data generating processes. Although, I can observe a deterioration of the performance of the bootstrap Granger causality test compared to the Wald test if I set bigger correlation of the error terms, which was not examined in the study of Wilms et al. (*2016*). Another observation is that the bigger the correlation of the error terms, the worse the performance of the bootstrap Granger causality test (thus, not just compared to the Wald test).

4.5 Results of the sensitivity analysis of the bootstrap Granger causality test for different values of λ_{ridge}

While I simulate the data, the value of the λ parameter (used for the second step of the adaptive lasso) is optimized in each simulation. Although, the value of the λ_{ridge} parameter (used for the first step of the adaptive lasso) is set as 0.1, as in the original code, but it can be optimized between 0 and 10. I examine whether it improves the performance of the bootstrap Granger causality test if I set other values for the λ_{ridge} parameter. I simulate time series $N = 200$ times, setting $T = 100$ and $K = 10$ and examine the performance of the test with $\lambda_{ridge} = (0.1, 0.5, 1, 5, 10)$. Thus, with four values besides the value 0.1, which is the original setting.

Fig. 4.8 shows the effect of setting different values of λ_{ridge} on bootstrap Granger causality test for $T = 100$, $K = 10$ and $\rho = 0.7$. The performance of the test is similar in every setting. It is difficult to decide, but maybe the test performs better with $\lambda_{ridge} = (0.5, 1)$ than with $\lambda_{ridge} = (0.1, 5, 10)$, although I do not see a big difference. Based on this result, the bootstrap Granger causality test is not so sensitive to the value of λ_{ridge} (which is used for the first step of the adaptive lasso), and this result corresponds to my previous expectations. Also based on this result, I might exclude that the worse performance of the bootstrap Granger causality test in case of stronger correlation of the error terms was caused by using a wrong value for the λ_{ridge} parameter.

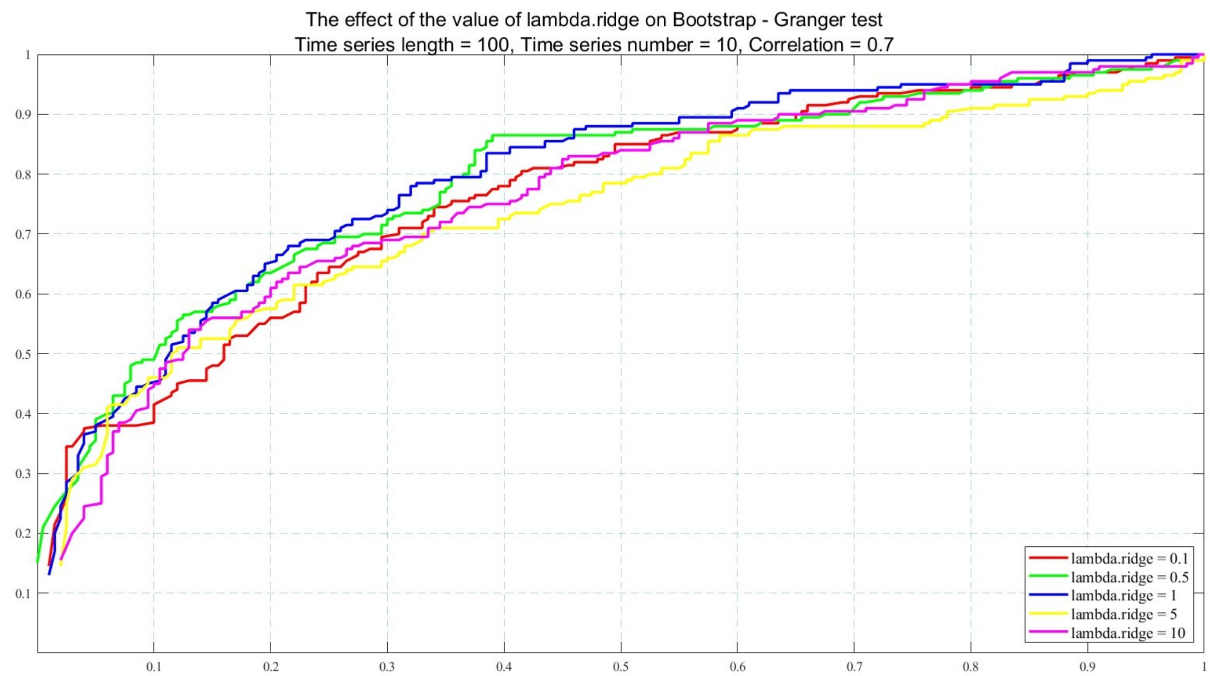

Figure 4.8: Size-power curve of the bootstrap Granger causality test with the setting of $K = 10$ and $T = 100$ with different values of λ_{ridge}. (Source: own construction.)

4.6 Results of the sensitivity analysis of the bootstrap Granger causality test for t-distributed errors

Adaptive lasso has the oracle property not just with Gaussian but also with non-Gaussian and conditionally heteroskedastic error terms. Thus, the method performs well with highly correlated regressors and t-distributed errors, which is common in financial applications (*Medeiros & Mendes, 2016*).

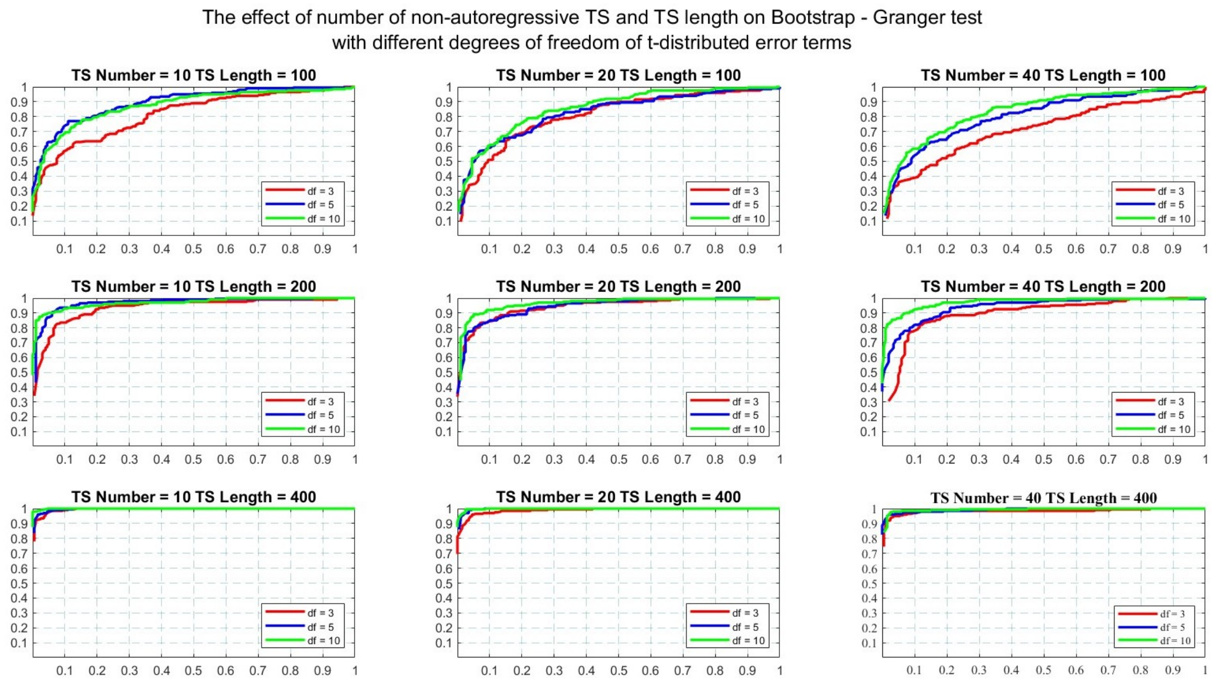

Figure 4.9: Size-power curves of the bootstrap Granger causality test with t-distributed error terms: $df = 3$ (red line), $df = 5$ (blue line) and $df = 10$ (green line) for every combination of $K = (10, 20, 40)$ and $T = (100, 200, 400)$. (Source: own construction.)

Until this point, the error terms were simulated from multivariate normal distribution. Now the error terms are simulated from multivariate t-distribution with different values of degrees of freedom. These are the followings: $df = (3, 5, 10)$. The smaller the value of df, the more heavy-tailed the distribution of the error terms. The bigger the value of df, the closer the distribution of the error terms to the normal distribution. For $df \to \infty$ the distribution is multivariate normal (*Barbaglia et al., 2020*).

Fig. 4.9 shows that the effect of the number of time series and the effect of length of time series on the bootstrap Granger causality test is the same with t-distributed error terms as with normally distributed errors. For a fixed value of K and increasing values of T the test has better results, and also for a fixed value of T and decreasing values of K. The bigger the value of degrees of freedom, the better performance the test has. It corresponds to my previous expectations since for $df \to \infty$ the distribution of the error terms is multivariate normal. The smaller the value of degrees of freedom, the more heavy-tailed the distribution of the errors is. In the three lower-dimensional cases, where $T = 400$, the test performs very well with every setting of the degrees of freedom. The difference between the performance of the test with the different values of degrees of freedom is bigger in the higher dimensional cases, for example, when $T = 100$ and $K = 40$.

Fig. 4.10 reports really similar results as the previous figures, although it shows a piece of additional information. It compares the performance of the bootstrap Granger causality test with t-distributed errors, where $df = 10$ with the performance of the test with normally distributed errors. The curves are quite similar in every combination of the values of T and K. Thus, for bigger degrees of freedom than 10, the bootstrap Granger causality test's performance is very similar to the test's performance with normally distributed errors. Maybe almost the same, and the differences are caused by the simulations.

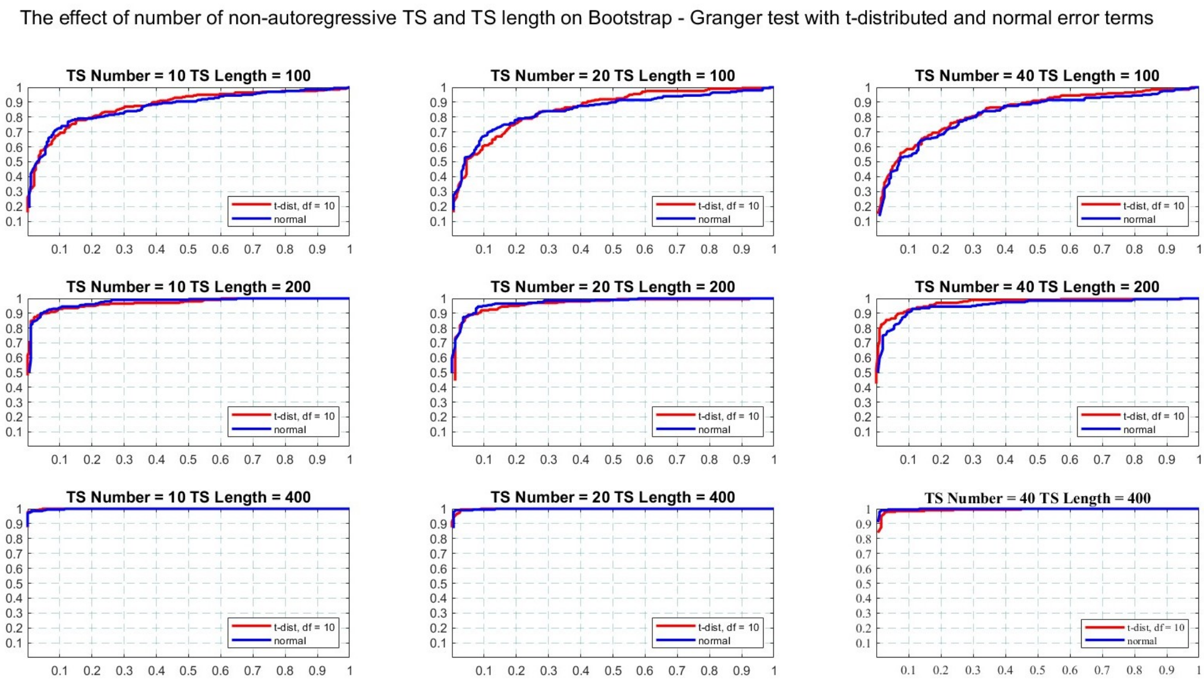

Figure 4.10: Size-power curves of the bootstrap Granger causality test with t-distributed errors, where $df = 10$ (red line) and with normally distributed errors (blue line) for every combination of $K = (10, 20, 40)$ and $T = (100, 200, 400)$. (Source: own construction.)

Chapter 5

Conclusion

In this study, firstly, I present the relevance of the topic. As more data are available, dimensionality has become a common issue nowadays. VAR models and Granger causality analysis are applied in various papers. Standard methods, such as the Wald test, cannot be applied in high-dimensional settings, but we can handle dimensionality with regularized methods. Although, an important property of the regularization is that the parameters become biased. Thus, we cannot calculate the confidence interval in the traditional way. A possible solution is to apply the bootstrap Granger causality method.

Then, I summarize the improvements related to the theory of high-dimensional VAR models. Under certain conditions, the estimation of the non-zero coefficients with adaptive lasso is asymptotically equivalent to the oracle assisted least squares estimator, with Gaussian and also with non-Gaussian and conditionally heteroskedastic error terms. Thus, the method performs well with correlated regressors and t-distributed errors, which is common in financial applications.

I compare a bootstrap Granger causality test with the standard Wald test. While the first one can be applied in high-dimensional settings, the Wald test exists only in low-dimension. I employ a bootstrap procedure using the adaptive lasso as a penalized estimator to select the relevant variables. With the bootstrap procedure, I reduce the omitted variable bias, thus, the inference is uniformly valid.

In the simulation study, I evaluate the performance of the bootstrap Granger causality test compared to the standard Wald test in low-dimensional settings, as the Wald test cannot be applied in the case of high-dimensional data. I use data generating processes selected based on financial applications, and the results are the following.

- The bootstrap Granger causality test outperforms the Wald test in every combination of $K = (10, 20, 40)$ and $T = (100, 200, 400)$ based on the size-power curves when the error terms are not correlated. This result corresponds to the results of Wilms et al. (*2016*).
- I also compare the bootstrap Granger causality test with correlated error terms too, which is not examined in the study of Wilms et al. (*2016*). In the case of $T = 100$ and $K = (10, 20)$, the Wald test outperforms the bootstrap Granger causality test even with $\rho = 0.3$, which is not a strong correlation.
- I examine in detail the effect of the correlation of the error terms on the bootstrap Granger causality test. The stronger the correlation of the errors, the worse the performance of the test.
- If the data generating process has a block-diagonal structure (DGP2), both tests perform worse than with DGP1. The effect of the correlation of the errors is similar to the case of DGP1.
- The test is not very sensitive to the selection of the λ_{ridge} parameter (which is used for the first step of the adaptive lasso).
- The effect of the number of time series and time series length on the bootstrap Granger causality test is the same with t-distributed errors as with normally distributed errors. The smaller the value of degrees of freedom, the more heavy-tailed the distribution of the errors is. The more heavy-tailed the distribution is, the worse performance the test has. Although, for $df \geq 10$ the results of the simulations are almost the same as with normally distributed errors.

There are some questions that I have not addressed yet. Further research is needed to examine the performance of the test in high-dimensional settings. It would also be interesting to compare the test with other regularized methods, as the approach of Hecq et al. (*2020*). Still, I think that I show important new results about the bootstrap Granger causality test that were not examined before, or at least not in detail.

Bibliography

[1] Akaike, H. (1974). A new look at the statistical model identification. IEEE transactions on automatic control, 19(6), 716-723.

[2] Audrino, F., & Camponovo, L. (2018). Oracle Properties, Bias Correction, and Bootstrap Inference for Adaptive Lasso for Time Series M-Estimators. Journal of time series analysis, 39(2), 111-128.

[3] Bahadori, M. T., & Liu, Y. (2013, May). An examination of practical granger causality inference. In Proceedings of the 2013 SIAM International Conference on data Mining (pp. 467-475). Society for Industrial and Applied Mathematics.

[4] Barbaglia, L., Croux, C., & Wilms, I. (2020). Volatility spillovers in commodity markets: A large t-vector autoregressive approach. Energy Economics, 85, 104555.

[5] Basu, S., & Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models. The Annals of Statistics, 43(4), 1535-1567.

[6] Belloni, A., Chen, D., Chernozhukov, V., & Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. Econometrica, 80(6), 2369-2429.

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

[8] Belloni, A., Chernozhukov, V., & Hansen, C. (2014a). High-dimensional methods and inference on structural and treatment effects. Journal of Economic Perspectives, 28(2), 29-50.

[9] Belloni, A., Chernozhukov, V., & Hansen, C. (2014b). Inference on treatment effects after selection among high-dimensional controls. The Review of Economic Studies, 81(2), 608-650.

[10] Belloni, A., Chernozhukov, V., & Kato, K. (2015). Uniform post-selection inference for least absolute deviation regression and other Z-estimation problems. Biometrika, 102(1), 77-94.

[11] 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.

[12] Berk, R., Brown, L., Buja, A., Zhang, K., & Zhao, L. (2013). Valid post-selection inference. The Annals of Statistics, 802-837.

[13] Bernanke, B. S., Boivin, J., & Eliasz, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. The Quarterly journal of economics, 120(1), 387-422.

[14] Chatterjee, A., & Lahiri, S. N. (2011). Bootstrapping lasso estimators. Journal of the American Statistical Association, 106(494), 608-625.

[15] Chaudhry, A., Xu, P., & Gu, Q. (2017, July). Uncertainty assessment and false discovery rate control in high-dimensional Granger causal inference. In International Conference on Machine Learning (pp. 684-693). PMLR.

[16] Chen, J., & Chen, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. Biometrika, 95(3), 759-771.

[17] Chernozhukov, V., Härdle, W. K., Huang, C., & Wang, W. (2021). Lasso-driven inference in time and space. The Annals of Statistics, 49(3), 1702-1735.

[18] Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. Econometrica: journal of the Econometric Society, 424-438.

[19] Granger, C. W. (1980). Testing for causality: A personal viewpoint. Journal of Economic Dynamics and control, 2, 329-352.

[20] Hecq, A., Laurent, S., & Palm, F. C. (2016). On the univariate representation of BEKK models with common factors. Journal of Time Series Econometrics, 8(2), 91-113.

[21] Hecq, A., Margaritella, L., & Smeekes, S. (2019). Granger causality testing in high-dimensional VARs: a post-double-selection procedure. arXiv preprint arXiv:1902.10991.

[22] Johansen, S. (1988). Statistical analysis of cointegration vectors. Journal of economic dynamics and control, 12(2-3), 231-254.

[23] Krampe, J., Kreiss, J. P., & Paparoditis, E. (2021). Bootstrap based inference for sparse high-dimensional time series models. Bernoulli, 27(3), 1441-1466.

[24] Kreiss, J. P., & Lahiri, S. N. (2012). Bootstrap methods for time series. In Handbook of statistics (Vol. 30, pp. 3-26). Elsevier.

[25] Kock, A. B., & Callot, L. (2015). Oracle inequalities for high dimensional vector autoregressions. Journal of Econometrics, 186(2), 325-344.

[26] Lee, J. D., Sun, D. L., Sun, Y., & Taylor, J. E. (2016). Exact post-selection inference, with application to the lasso. The Annals of Statistics, 44(3), 907-927.

[27] Leeb, H., & Pötscher, B. M. (2005). Model selection and inference: Facts and fiction. Econometric Theory, 21(1), 21-59.

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

[29] Masini, R. P., Medeiros, M. C., & Mendes, E. F. (2019). Regularized estimation of high-dimensional vector autoregressions with weakly dependent innovations. Journal of Time Series Analysis.

[30] Medeiros, M. C., & Mendes, E. F. (2016). l (1)-regularization of high-dimensional time-series models with non-gaussian and heteroskedastic errors.

[31] Nicholson, W. B., Matteson, D. S., & Bien, J. (2017). VARX-L: Structured regularization for large vector autoregressions with exogenous variables. International Journal of Forecasting, 33(3), 627-651.

[32] Schwarz, G. (1978). Estimating the dimension of a model. The annals of statistics, 461-464.

[33] Skripnikov, A., & Michailidis, G. (2019). Joint estimation of multiple network Granger causal models. Econometrics and Statistics, 10, 120-133.

[34] Song, S., & Bickel, P. J. (2011). Large vector auto regressions. arXiv preprint arXiv:1106.3915.

[35] Song, X., & Taamouti, A. (2019). A better understanding of granger causality analysis: A big data environment. Oxford Bulletin of Economics and Statistics, 81(4), 911-936.

[36] Stock, J. H., & Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. Journal of Business & Economic Statistics, 20(2), 147-162.

[37] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267-288.

[38] Van de Geer, S., Bühlmann, P., Ritov, Y. A., & Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. The Annals of Statistics, 42(3), 1166-1202.

[39] Van de Geer, S., Bühlmann, P., & Zhou, S. (2011). The adaptive and the thresholded Lasso for potentially misspecified models (and a lower bound for the Lasso). Electronic Journal of Statistics, 5, 688-749.

[40] Wilms, I., & Croux, C. (2016). Forecasting using sparse cointegration. International Journal of Forecasting, 32(4), 1256-1267.

[41] Wilms, I., Gelper, S., & Croux, C. (2016). The predictive power of the business and bank sentiment of firms: A high-dimensional Granger Causality approach. European Journal of Operational Research, 254(1), 138-147.

[42] Wong, K. C., Li, Z., & Tewari, A. (2020). Lasso guarantees for β-mixing heavy-tailed time series. The Annals of Statistics, 48(2), 1124-1142.

[43] Zou, H. (2006). The adaptive lasso and its oracle properties. Journal of the American statistical association, 101(476), 1418-1429.

[44] Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the royal statistical society: series B (statistical methodology), 67(2), 301-320.

Appendix A

Glossary

- adaptive lasso (adaptive least absolute shrinkage and selection operator): adaptív lasso (adaptív legkisebb abszolút zsugorító és szelekciós operátor)
- AIC (Akaike information criterion): AIC (Akaike információs kritérium)
- alternative hypothesis: alternatív hipotézis
- argument: argumentum (pl. függvényé)
- asymptotically equivalent: aszimptotikusan ekvivalens
- autoregressive: autoregresszív
- bayesian: bayes-i
- beta-min condition: minimum béta feltétel
- biased: torzított
- BIC (Bayesian information criterion): BIC (bayes-i információs kritérium)
- bootstrap: bootstrap
- burn-in: burn-in (első n generált minta elvetése)
- causality: (grangeri értelemben vett) okság

- coefficient: koefficiens, béta, paraméter
- cointegration: kointegráció
- conditionally heteroskedastic: feltételesen heteroszkedasztikus
- confidence interval: konfidencia intervallum
- consistent: konzisztens
- control variable: kontrollváltozó
- covariance matrix: kovarianciamátrix
- cross-section: keresztmetszeti
- cross-validation: keresztvalidáció
- data-driven method: adatvezérelt módszer
- data generating process: adatgeneráló folyamat
- diagonal: diagonális
- EBIC (Extended Bayesian information criterion): EBIC (kiterjesztett bayes-i információs kritérium)
- elastic net: rugalmas háló
- empirical distribution function: empirikus eloszlásfüggvény
- error term: hibatag
- exogenous variable: exogén (külső) változó
- factor models: faktormodellek
- Granger causality: Granger-okság
- heavy-tailed distribution: vastag szélű eloszlás
- high-dimensional: magas dimenziós

- hypothesis: hipotézis
- independent and identically distributed (i.i.d.): független azonos eloszlású
- indicator function: indikátor függvény
- inference: hipotézisvizsgálat
- information criterion: információs kritérium
- intercept: ordinátatengely-metszet
- invertible: invertálható
- lag-length: késleltetéshossz
- lasso (least absolute shrinkage and selection operator): lasso (legkisebb abszolút zsugorító és szelekciós operátor)
- linear: lineáris
- log-likelihood: log likelihood
- low-dimensional: alacsony dimenziós
- martingale difference sequence: martingál differencia sorozat
- matrix: mátrix
- mid p-Value: közép p-érték
- ML (maximum likelihood) method: ML (legnagyobb valószínűség) módszer
- multiple Granger causal network: többszörös Granger-oksági hálózat
- multivariate: többváltozós
- non-Gaussian: nem-Gauss eloszlású
- null hypothesis: nullhipotézis
- normal distribution: normális eloszlás

- OLS (ordinary least squares) model: OLS (legkisebb négyzetek) modell
- omitted variable bias: kihagyott változók okozta torzítás
- oracle asisted estimator, oracle property: ”jós tulajdonságú” becslés (előre ismerjük a releváns változókat)
- outcome variable: eredményváltozó
- overfit: túlilleszt
- partial: parciális
- penalization: penalizáció, regularizáció, büntetés
- point-wise: pontonként
- positive definite: pozitív definit
- post-double-seleczion procedure: dupla szelekció utáni eljárás
- post-selection: változószelekció utáni
- power: statisztikai erő
- predictor matrix: magyarázóváltozó mátrix
- p-value: p-érték
- regression: regresszió
- regressor: regresszor, változó
- regularization: regularizáció, penalizáció, büntetés
- residual: maradéktag, reziduum
- response vector: eredményváltozó vektor
- sensitivity analysis: érzékenységvizsgálat
- set: halmaz

- shrinkage methods: zsugorító módszerek
- significance level: szignifikanciaszint
- significant: szignifikáns
- size: a teszt mérete (elsőfajú hiba valószínűsége)
- sparse: ritka (pl. mátrixnál: vannak nulla elemei is)
- sparsity parameter: zsugorító paraméter
- spurious causal relation: hamis oksági kapcsolat
- stable process: stabilis folyamat
- stationary: stacionárius
- t-distribution: t-eloszlás
- test statistic: teszt statisztika
- time series: idősor
- treatment variable: magyarázóváltozó
- tuning parameter: hiperparaméter
- unbiased: torzítatlan
- uniform distribution: egyenletes eloszlás
- uniformly: egyenletesen
- unit disc: egységkör, egység sugarú kör
- univariate: egyváltozós
- upper bound: felső korlát
- valid: érvényes
- VAR (vector autoregressive) model: VAR (vektor autoregresszív) modell

- variance reduction: variancia csökkentés
- VECM (vector error correction model): VECM (vektor hibakorrekciós modell)
- vector: vektor
- weakly stationary: gyengén stacionárius

Appendix B

Summary of my thesis in Hungarian

Napjainkban egyre több adat áll rendelkezésünkre, aminek következtében a dimenzionalitás általános problémává vált. Számos cikkben alkalmazzák a VAR modellt és vizsgálják, hogy van-e Granger-okság. A standard módszereket - mint például a Wald tesztet - magas dimenziós adatokon nem lehet alkalmazni, de a dimenzionalitást többek közt regularizációs módszerekkel tudjuk kezelni. Azonban a regularizáció egy fontos tulajdonsága, hogy torzítottak lesznek a paraméterek, ami miatt nem lehet hagyományos módon konfidencia intervallumot számolni hipotézisvizsgálattal. A problémára egy lehetséges megoldás a bootstrap Granger-okság módszer alkalmazása.

A magas dimenziós VAR modellek elméletében fontos előrelépést jelentettek a következő állítások. Bizonyos feltételek mellett a nemnulla koefficiensek adaptív lassoval történő becslése aszimptotikusan ekvivalens a "jós tulajdonságú" OLS becsléssel. Ez nem csak normális eloszlású, hanem nem-Gauss eloszlású és feltételesen heteroszkedasztikus hibatagok esetén is fennáll. A módszer tehát jól működik akkor is, ha a változók korrelálnak és t-eloszlásúak a hibatagok, ami gyakori pénzügyi alkalmazások esetén.

Dolgozatomban összehasonlítom a bootstrap Granger-okság tesztet a standard Wald teszttel. Míg az első magas dimenziós adatok esetén is alkalmazható, a második csak alacsony dimenzió esetén. A releváns változókat egy bootstrap módszerrel választom ki, ahol a regularizáció adaptív lassoval történik. A lasso kiszűri az irreleváns változókat, majd a releváns változókra OLS futtatásával kapok t-statisztikákat. Azonban így nem venném figyelembe az első lépésbeli bizonytalanságot. A végső modell függene az adatoktól, túlillesztene. Kihagyhatnék gyengén releváns változókat, amik túl nagy kihagyott változók okozta torzítást eredményeznek, amit aszimptotikusan nem lehet figyelmen kívül hagyni. Ezáltal a szelekció utáni becslés nem egyenletesen, hanem csak pontonként konvergálna a normális eloszláshoz. A bootstrap procedúrával azonban csökkentem a kihagyott változók okozta torzítást és így kapok egyenletesen érvényes hipotézisvizsgálatot.

A bootstrap Granger-okság és a Wald teszt teljesítményét szimulációkkal hasonlítom össze. A szimulációkhoz úgy választottam meg az adatgeneráló folyamatokat, hogy megfeleljenek a pénzügyi alkalmazásoknak és a következő eredményeket kaptam.

- Ha a hibatagok nem korrelálnak, akkor a bootstrap Granger-okság teszt teljesítménye jobb mint a Wald teszté minden lehetséges $K = (10, 20, 40)$ és $T = (100, 200, 400)$ beállítás esetén, amit size-power görbékkel szemléltetek. Ez az eredmény megfelel Wilms et al. (*2016*) eredményének.
- A két módszer teljesítményét korreláló hibatagokkal is összehasonlítom, amit Wilms et al. (*2016*) tanulmányában nem vizsgáltak. $T = 100$ és $K = (10, 20)$ kombinációi esetén a Wald teszt teljesítménye jobb mint a bootstrap Granger-okság teszté még $\rho = 0.3$ korreláció esetén is, ami nem mondható erős korrelációnak.
- Részletesen megvizsgálom, hogy hogyan hat a hibatagok korrelációja a bootstrap Granger-okság tesztre. Minél erősebb a korreláció, annál rosszabb a teszt teljesítménye.
- Ha az adatgeneráló folyamat blokkos szerkezetű (mint DGP2), akkor mindkét teszt teljesítménye rosszabb, mint a DGP1 esetén. A hibatagok korrelációjának hasonló hatása van a bootstrap Granger-okság teszt teljesítményére mindkét adatgeneráló folyamat esetén.
- A teszt nem túl érzékeny a λ_{ridge} paraméter megválasztására (amit az adaptív lasso első lépésénél használok).
- t-eloszlású hibatagok esetén is ugyanolyan hatással van a teszt teljesítményére az idősorok száma és az idősorok hossza, mint normális eloszlású hibatagok esetén. Minél kisebb a szabadságfok, annál vastagabb szélű a hibatagok eloszlása. Minél

vastagabb szélű az eloszlás, annál rosszabb a teszt teljesítménye. Azonban ha a szabadságfok legalább 10, akkor már nagyon hasonló eredményt kapok, mint normális eloszlású hibatagok esetén.

Felmerül néhány kérdés, amiket a dolgozatomban nem válaszoltam meg. További kutatásban érdemes lenne megvizsgálni a bootstrap Granger-okság teszt teljesítményét magas-dimenziós adatokon is. Emellett érdekes lenne összehasonlítani a tesztet más regularizációs módszerekkel, például Hecq et al. (*2020*) tanulmányában alkalmazott módszerrel. Ennek ellenére úgy gondolom, hogy fontos új eredményeket kaptam a bootstrap Granger-okság tesztről amiket eddig még nem vizsgáltak, vagy legalábbis nem ennyire részletesen.