Lund University

Master's Programme in Finance

Capturing time variation within systemic risk estimation

Assessing the impact of state variable selection on CoVaR estimates

By

Baltsar Lindgren

Christian Hovstadius

NEKN02

Master's Thesis

June 2023

Supervisor: Anders Vilhelmsson

Abstract

Systemic risk can be defined as the risk to the whole financial system. Financial institutions may contribute more or less to this risk, and measuring the systemic risk contributions of institutions is of central importance for regulators. This is important since it makes it possible to determine to what extent different institutions contribute to the overall systemic risk of the financial system and hence which ones are more or less systemically important. Adrian & Brunnermeier (2011) proposes the systemic risk measure CoVaR, which builds on the framework of Value-at-Risk (VaR). The definition of CoVaR is the q^{th}% VaR of institution j (or in the case of this essay, the European financial system) given that another institution i is at its q^{th}% VaR. ΔCoVaR measures the change in the VaR for institution j given that institution i is in distress (compared to its normal state), and estimates the marginal risk contribution for a given institution. To obtain time variation in the estimates, the authors suggests using state variables that condition the mean and volatility of the risk measure. This essay tries to answer the question whether the systemic risk estimates obtained by using the CoVaR methodology, and the systemic risk contribution rankings between banks, are sensitive to the selection of these state variables. Using equity price data for 141 European banks, and data for 20 state variables during the time period from 31^{st} of December 2002 to 30^{th} of September 2022, this essay estimates VaR, CoVaR, ΔCoVaR and $\Delta^{\$}$CoVaR using quantile regressions, following the methodology of Adrian & Brunnermeier (2011). Using four different state variable selection methods including the one suggested by Adrian & Brunnermeier, the supervised and unsupervised machine learning methods of Lasso regression and PCA, and a randomized method, systemic risk contributions over time are estimated and the banks are ranked according to these estimates. The results of this essay suggests that the CoVaR risk measure indeed is sensitive to the choice of state variables selection method, where both the estimates as well as rankings differs between the methods.

Keywords: VaR, CoVaR, State Variables, Quantile Regression, Systemic Risk, Lasso, PCA

Acknowledgements

To our supervisor Anders Vilhelmsson, who has offered us support and guidance throughout this process. Your expertise has been invaluable.

Contents

1. Introduction

The financial crisis of 2008 and similar economic events often result in detrimental effects on the financial system, society, and its members. What has become clear from these kinds of events and from previous research is that the risk financial institutions face in isolation is of little importance when wanting to assess the risk that the entire financial system faces (Montagna et al. 2020). Hence, the importance for regulators to measure this latter mentioned risk has, and will, continue to grow. This risk is the systemic risk, which can be defined in many ways, but what is central to the concept is that it reflects the risk faced by the financial system as a whole, and that this risk stems from connected components of the system being in distress. Because of the potential outcomes of a major disruption of the financial system, it is of great importance that regulators have proficient systemic risk measures at their disposal when quantifying and measuring the systemic risk in the economy. A potential suggestion for such a systemic risk measure is CoVaR which is extended to ΔCoVaR, first developed by Adrian & Brunnermeier (2011). One aspect of the ΔCoVaR methodology and the measurement of systemic risk in general is modeling and capturing its development over time, which aims toward giving us an understanding of the time-variation of systemic risk that the selected risk measure captures. Most of the previous research, including Adrian & Brunnermeier, has chosen to model the time variation of systemic risk by using state variables within the quantile regression framework. The selection of which state variables to use when obtaining time-varying estimates of ΔCoVaR has however been given little, if any, attention in previous research and in many cases the variables first suggested by Adrian & Brunnermeier have served as a benchmark. As far as the authors of this essay are concerned, the implications of different state variables selection methods on ΔCoVaR estimates have not been studied before. Hence, it is within this selection process this essay wishes to make a contribution.

The main objective of this essay is to investigate ΔCoVaR's expediency in capturing time variation in systemic risk. This is done by estimating the systemic risk in the European banking sector over time with different state variable selection methods following the estimation procedure of Adrian & Brunnermeier (2016). By doing this, the essay is able to investigate whether the systemic risk estimates show significant fluctuations or not, and hence assess the sensitivity of ΔCoVaR, more specifically, $\Delta^{\$}$CoVaR, to the choice of state variable selection method. The results of this essay will try to answer the following stipulated research questions:

> *Do different state variable selection methods change the systemic risk contribution estimates?*

> *Do different state variable selection methods change the systemic risk contribution rankings between banks?*

The first research question aims to answer whether different selections of state variables result in different values of the systemic risk estimates. More specifically, when answering the first research question this essay puts an emphasis on $\Delta^{\$}$CoVaR, which takes the size of a bank into account. This is further motivated in section six. The second question instead focuses less on the actual estimate levels, but more on if rankings of systemic risk contributions between banks are consistent across different state variable selection methods. Investigating how sensitive ΔCoVaR is to the choice of state variables is important since fluctuations depending on the choice of state variables introduce uncertainty in the systemic risk estimates. This could question the risk measures expediency to capture time variation in systemic risk, which is highly relevant for regulators wishing to estimate the risk the financial system faces. To assess the sensitivity of ΔCoVaR, this essay looks at the European banking sector and estimates VaR, CoVaR, ΔCoVaR and $\Delta^{\$}$CoVaR for a large sample of 141 European publicly traded banks in the period between December 2002 and October 2022. To determine the risk measures' dependency on the state variables, this essay uses a larger sample of state variables (20) and implements four types of selection procedures. First, systemic risk estimates are obtained by using Adrian & Brunnermeier's (2016) selected state variables with minor changes due to the differences in geographical coverage. Secondly, state variables are selected by using a supervised regularization method in form of Lasso regression, which is followed by an estimation conducted with the selected variables. Thirdly, an unsupervised method of dimensionality reduction is used, which is the Principal Components Analysis. Lastly, a randomized selection procedure is implemented. The estimations of VaR, CoVaR, ΔCoVaR and $\Delta^{\$}$CoVaR follow the methodology presented in Adrian and Brunnermeier (2016).

The remainder of the essay is structured as follows. The upcoming section introduces the theoretical framework that lies as a foundation for the work of this essay. More specifically, this contains sections on the concept of systemic risk, definitions of relevant risk measures, and regulation. Section three presents previous research and related literature within the field of systemic risk. In section four, the methodology of estimating CoVaR and ΔCoVaR and the state variables selection methods are explained. Section five describes the data used and the data collection process, and section six presents the empirical results. Section seven presents a discussion followed by section eight that summarizes the main conclusions of the essay and suggests areas of future research. Section nine contains the references.

2. Theoretical framework

2.1 Systemic risk

The concept of systemic risk has been extensively covered in the literature, but a universal definition of it has not yet been stipulated since different research emphasizes different aspects of the concept (W. Silva et al. 2017). However, it could be argued that most definitions share common characteristics.

According to the European Central Bank (ECB), a broad definition of systemic risk refers to "the risk that financial instability becomes so widespread that it impairs the functioning of a financial system to the point where economic growth and welfare suffer materially" (ECB 2009). Another definition is proposed by Adrian & Brunnermeier (2016), which focuses on distress (losses) spreading between institutions and how this can give rise to the endangering of the entire financial system. More specifically, the authors define systemic risk as the risk that the impairment of the functionality of the financial system results in potentially unfavorable consequences for the real economy. Kaufman & Scott (2003) argues for a similar definition and stipulate that systemic risk refers to "the risk or probability of breakdowns in an entire system", which is contrasted to events in separate parts of the system. Specific to the banking sector, the authors argue that systemic risk is showcased by a high correlation and concentration between bank failures. Acharya & Richardson (2009) defines systemic risk in similar terms by stating that systemic risk is the "joint failure of financial institutions and capital markets that considerably shorten the supply of capital to the real market" (W. Silva et al. 2017).

Evidently, different definitions of systemic risk are brought up within the literature, but a certain emphasis is on systemic risk as the risk to the whole financial system stemming from simultaneous or linked distresses of financial institutions. According to Smaga (2014), there are a number of aspects that are most often brought up within definitions of systemic risk. Firstly, systemic risk is concerned with a larger part of the financial system alternatively a notable number of financial institutions, and the disruption of the performance and functions of this larger part. Secondly, a fundamental component of systemic risk is the transference and spreading of disturbances (distress) between connected parts of the financial system, which in turn may have an adverse effect on the economy.

2.2 Definition of risk measures

For the purpose of this essay, the relevant risk measure must be defined. The systemic risk measure CoVaR, as indicated by its name, builds on the foundation of Value at Risk (VaR), which is mathematically defined as:

$$VaR_q = \min\{l : Pr(L > l) \leq 1 - q\} \tag{2.1}$$

The definition of VaR, according to equation 2.1 above, at the q confidence level is the smallest loss l, that satisfies the condition that the probability of suffering a loss L that is larger than l is less than or equal to $1 - q$. I.e., VaR is the loss level during a certain period of time that will not be exceeded by a probability of $1 - q\%$ (Hull 2018). Equation 2.1 shows the definition of VaR when the loss function is discrete. However, in the case of a continuous loss distribution, the definition is slightly altered:

$$VaR_q = Pr(L > VaR_q) = 1 - q \tag{2.2}$$

Equation 2.2 essentially states that VaR_q by definition is, at the q confidence level, the q^{th} quantile of the loss distribution, but the above-mentioned interpretation still applies. (ibid.) (JP Morgan 1996)

However, VaR measures and puts an emphasis on the risk that an institution faces in isolation which won't necessarily showcase an institution's contribution or connection to the systemic risk. This is empirically shown by Adrian & Brunnermeier (2016). They show, which is further mentioned in section three, that the connection between an institution's VaR and its systemic risk contribution is very weak. Hence, Adrian & Brunnermeier (2016) introduces the systemic risk measure CoVaR (Conditional Value at Risk). CoVaR for institution j is defined as the VaR for that same institution conditional on another institution i being in a certain state. Institution j is by the authors assumed to be the financial system and the state is described as when institution i is having a loss equal to or above the institutions' VaR for a specific quantile. CoVaR of institution j (or the system) is hence defined in the following way:

$$Pr(L^j|C(L^i) \leq CoVaR_q^{j|C(L^i)}) = q\% \tag{2.3}$$

Where $C(L^i)$ represents the conditioning on institution i being at a certain state. Similar to the above-mentioned definition of VaR, but now in the conditional setting, $CoVaR_q^{j|C(L^i)}$ is defined by the q% quantile of the probability distribution.

To calculate and measure a specific institution's systemic risk contribution, the authors define ΔCoVaR. This is the difference between the VaR of the financial system when firm i is having a loss equal to its VaR at a q% quantile and the VaR of the financial system when firm i is having a loss equal to its median state, which is VaR at the 50% quantile. ΔCoVaR is hence defined as:

$$\Delta CoVaR_q^{j|i} = CoVaR_q^{j|L^i=VaR_q^i} - CoVaR_{0.5}^{j|L^i=VaR_{0.5}^i} \tag{2.4}$$

What is clear from equation 2.4 is that this measure calculates the change in CoVaR when the state that is conditioned on changes from an adverse loss scenario to a more "normal" one. ΔCoVaR hence measures the tail-dependency between the financial system and a specific institution whereas a higher ΔCoVaR implies a larger systemic risk contribution and vice versa. ΔCoVaR however does not consider the size of institutions. This consideration is important if the objective is to compare systemic risk contributions across institutions that vary in size, which is the case for this essay. For this Adrian & Brunnermeier proposes $\Delta^{\$}CoVaR$, which measures the change in dollar amounts when the conditioning state changes accordingly to the above-described shift. $\Delta^{\$}CoVaR$ is defined and calculated by:

$$\Delta^{\$}CoVaR_q^{j|i} = MC^i * \Delta CoVaR_q^{j|i} \tag{2.5}$$

MC^i corresponds to the market value of equity of institution i.

The method used by Adrian & Brunnermeier (and many others, see section three) to obtain ΔCoVaR estimates over time, involves the use of state variables. This estimation procedure is further described in section four. However, the theoretical interpretation of the state variables is of importance for the reader since they are at the center of the purpose of this essay, and hence some theoretical elaboration is needed. According to Adrian & Brunnermeier, the state variables used to estimate time-varying ΔCoVaR should not be seen as risk factors to the system. I.e., they should not be interpreted as factors driving the systemic risk in a certain direction. Instead, the state variables are *"variables that condition the mean and the volatility of the risk measure"*. As indicated by the name, the variables wishes to capture the state of the economy rather than describing a causal relationship with systemic risk. Hence they should be interpreted as conditioning variables rather than systemic risk factors.

What becomes clear from the included state variables in previous research and in this essay is that they reflect changes in market conditions that affect all the institutions participating in the market and hence control for economically meaningful events (Krygier 2014).

2.3 Regulation

In 2018, the Basel Committee on Banking Supervision (BCBS) provided an updated and revised methodology for assessing the systemic importance of globally operating banks (G-SIB: Global Systemically Important Banks). This update replaced the methodology formulated in 2013. In general, the BCBS communicates that the systemic importance of a bank should be measured by looking at how the global financial system and the economy are affected by the failure of that bank. More specifically, the definition and detection of such a bank relies on what is called an indicator-based approach, which uses indicators that are meant to reflect the size, interconnectedness, available substitutes or financial institution infrastructure, cross-jurisdictional activity, and complexity of the bank. In turn, these categories aim to assess the systemic importance of a bank. Each of the five categories is, in total, given equal weight but each category contains smaller sub-categories that differ in terms of weighting. A bank is then given an overall score that reflects its systemic importance. This score is calculated by taking the average of the banks' scores in the five categories and if the overall score is above a certain decided threshold, the bank is classified as a G-SIB. The framework then uses a so-called bucketing approach, where a higher overall score implies a higher bucket for the bank which gives the bank a higher loss absorbency requirement. The loss absorbency refers to Common Equity Tier 1 as a percentage of risk-weighted assets where the Common Equity Tier 1 is defined in the Basel III framework. The first bucket imposes an additional loss absorbency of 1% and the fifth, and highest bucket, imposes an additional loss absorbency of 3.5%. For a more detailed explanation of the framework and methodology, the reader is referred to the full documentation *"Global systemically important banks: revised assessment methodology and the higher loss absorbency requirement"* (Basel Committee on Banking Supervision 2018). The list of G-SIBS and which respective buckets they are allocated are supplied by the Financial Stability Board (FSB) (Financial Stability Board 2022) and the list as of November 2022 can be found in Appendix G. In section six, when the essays' empirical findings are presented, lists for previous years are used. These are also supplied by FSB on their website.

3. Literature review

Previous research within the field of systemic risk is extensive. However, research that puts emphasis on the modeling of time variation and the selection and usage of state variables does not seem to have been conducted. This section will hence put focus on previous research that has estimated systemic risk by using ΔCoVaR as its selected risk measure.

Adrian & Brunnermeier (2016) introduce the systemic risk measure CoVaR (conditional Value at Risk) and further ΔCoVaR. To estimate their proposed systemic risk measure, the authors use quantile regressions. This is further elaborated on and explained in section four of this essay. In Adrian & Brunnermeier's estimation of CoVaR and ΔCoVaR, they use data on 1823 publicly traded financial institutions within banking, security- brokerage, insurances, and real estate in the US. Their data covers the period between 1971Q1-2013Q2.

The estimation is conducted both conditionally and unconditionally, where the unconditional estimation results in a ΔCoVaR estimate that is time constant. The conditional estimation, which is of greater importance for this essay, is conducted with the ambition to capture time variation in the systemic risk measure, and hence the development of tail risk over time. This is performed by running quantile regressions including state variables of a macroeconomic character. The state variables selected and used are: *the change in the three-month treasury bill rate, the change in the slope of the yield curve, TED-spread, change in credit spread, market return from the S&P500, real estate sector return in excess of the market financial sector return, and equity volatility.* The selection of these variables are motivated by that they capture the variation over time in the conditional moments of asset returns, are liquid and easily tradable, as well as tractable. Additionally, Adrian and Brunnermeier propose a forward-looking systemic risk measure called forward-ΔCoVaR. These estimates are obtained by regressing ΔCoVaR on institution characteristics as well as state variables. The main institution characteristics that the authors consider are leverage, maturity mismatch, size, and a boom indicator, which indicates the number of quarters in a row an institution has been in the highest decile of the market-to-book ratio compared to other firms. Examples of additional institution characteristics are loan loss allowances, intangible assets, and interest-bearing core deposits. Regarding the forward-looking ΔCoVaR, the authors find that an increase in the mentioned main institutional characteristics significantly predicts ΔCoVaR and hence contribute to an increased systemic risk. The authors additionally assert that the forward-looking ΔCoVaR (2006Q4 value) "would have predicted more than one-third of realized ΔCoVaR during the 2007-2009 financial crisis" (Adrian & Brunnermeier, 2016). Another main conclusion from Adrian & Brunnermeier is that the connection between an institutions VaR and its systemic risk contribution (ΔCoVaR) seems to be weak in the cross-sectional dimension. The connection is however strong in the time series dimension. This is important since it implies that regulation cannot only focus on the VaR of institutions in isolation when trying to mitigate systemic risk. Hence, when aiming toward mitigating systemic risk, regulation should consider the connection between institutions. Otherwise, there is a risk of regulation overlooking the systemic dimension of the risk.

Following the publication of Adrian & Brunnermeier's paper, a lot of research using their systemic risk methodology has been conducted. Borri et al. (2012) looks at the European banking sector aiming to identify main predictors of systemic risk contribution. Their methodology follows the one presented by Adrian and Brunnermeier (2016) and they perform conditional estimation of ΔCoVaR on a sample of 233 publicly listed banks in the Eurozone area between 1999 and 2011. The included state variables are: *change in option implied DAX volatility, short-term liquidity spread, change in the slope of the yield curve, DAX return, and the three-month treasury change.* To identify which variables can significantly predict systemic risk contributions, the authors performed a pooled OLS-regression with their ΔCoVaR estimates as their dependent variable. Independent variables were selected with the ambition to account for bank's balance sheets, banking system, market-related factors, and risk characteristics. The authors find that ΔCoVaR is a good systemic risk measure because of its persistence. Additionally, they find that banks' balance sheets weakly predict systemic risk contributions while size and leverage are found to be significant predictors. Furthermore, if a bank has its headquarters located in a more concentrated banking system, it significantly contributes to an increase in the systemic risk. These conclusions indicate that regulation which only focuses on the size of banks will not capture the entire systemic risk contribution. Petrella et al. (2019) also studies the systemic risk in Europe but instead of looking at risk contributions at the institutional level, they look at the country level to estimate each country's systemic risk contribution. The authors also conduct two testing techniques called the dominance test and the significance test first proposed by Castro & Ferarri (2014). These tests are conducted to be able to distinguish partly if a country's risk contribution is statistically significant, and hence can be considered systemically relevant. Also, the dominance test wishes to distinguish if a certain country contributes to negative spillover effects more than other countries. The authors use return data on the EuroStoxx50 index between 2008 and 2017 and hence wish to attribute systemic risk contributions to the nations represented in the index. These are France, Germany, Spain, Netherlands, Italy, Belgium, Finland and Ireland. Following the methodology of Adrian and Brunnermeier, ΔCoVaR is conditionally estimated with a set of state variables to capture tail risk development over time including e.g. *the VSTOXX index (EuroStoxx50 volatility), MSCI real estate index returns, three-month EURIBOR, exchange rates, and more.* The authors conclude that all eight countries can be seen as systemically relevant and hence that impactful events tend to spread through the system beyond the respective country's domestic markets. Additionally, the authors find that France and Germany have the largest systemic risk contributions and that these two countries also are most likely to generate negative spillover effects and hence jeopardize the whole system.

Other examples of systemic risk research that implement Adrian and Brunnermeiers CoVaR methodology are Lopez-Espinoza et al. (2012), Bernal et al. (2014) and Ashgarian et al. (2022). In *Short-term wholesale funding and systemic risk: A global CoVaR approach*, Lopez-Espinosa et al. (2012) uses the CoVaR methodology to distinguish the main determinants of systemic risk contribution with a data set of 54 large international banks between 2001-2009. A particular geographical focus is not taken and the study wishes to contribute with an analysis of global systemic risk contributions. Time-varying ΔCoVaR is estimated by quantile regressions and estimates are obtained by using Adrian & Brunnermeier's US state variables. To test which institutional-specific factors drive systemic risk contributions, variables such as size, leverage, short-term borrowings to total assets ratio, total assets, etc are regressed on the ΔCoVaR estimates by using panel data regression methods. Conclusions are drawn by the study that offers nuance to the earlier-mentioned European studies. Short-term wholesale funding is found to be the most significant determinant of a bank's global systemic risk contribution. Banks with higher short-term funding ratios are shown to be more connected to other banks, among other features, which results in an increase in their systemic risk contribution. Additionally, the authors find, in contrast to earlier studies, that neither size nor leverage are main factors behind systemic risk in their data set. Another important conclusion drawn by the study is the role of asymmetric responses within the financial system. To be able to capture the role of asymmetries, the authors propose an extension called "Asymmetric CoVaR". For details regarding the econometric extensions, the reader is referred to the paper in question. The authors find that individual institutions' balance sheet contraction, which is defined as "(. . .) the propagation of distress associated with a decline in the market value of assets held by individual institutions" (Lopez-Espinoza, et al., 2012), significantly contributes to negative spillovers on the global index VaR. Hence, taking balance sheet contraction into account is important when trying to estimate and rank institutions' systemic risk contributions.

Bernal et al. (2014) studies systemic risk contribution on a sector level by looking at how distress within the financial sectors of banking, insurance, and ”other” financial services contributes to systemic risk. The authors wish to compare the three industries between the Eurozone and the US and use data spanning between 2004-2012. The two systems are defined by the Stoxx Europe 600 index excluding financial companies as well as the S&P 500 index with the same exclusion, and the three sectors by EuroStoxx 50 and Dow Jones sector indexes. ΔCoVaR is conditionally estimated, and the state variables used are in line with the ones in Adrian and Brunnermeier (2016). To be able to determine whether one financial sector has a significant systemic risk contribution, the authors also conduct a significance test (Kolmogorov-Smirnov (KS)), and to test if a sector has a more impactful systemic risk contribution than another, they conduct a stochastic dominance test which follows a bootstrapping strategy developed by Abadie (2002). The authors find that the sector that contributes most to the systemic risk in the Eurozone is the ”other” financial services sector and that the banking sector comes in second place. This result differs for the US, where the insurance industry is found to have the highest systemic risk contribution and the banking sector the lowest. However, all the included financial sectors are shown to have significant systemic risk contributions with ΔCoVaR estimates significantly different from zero. This is the case both for the Eurozone and the US.

In *Systemic risk and centrality: The role of interactions*, Vilhelmsson et al. (2022) use CoVaR in combination with constructed centrality measures to study how much an institution's systemic risk contribution depends on its centrality in financial networks. More specifically, the paper investigates whether institutional characteristics have a stronger or weaker impact on systemic risk contribution when an institution's centrality in a financial network is taken into account. The paper uses stock return data and state variables in line with Adrian & Brunnermeier (2016) spanning between 1995-2016 to obtain time-varying ΔCoVaR estimates. To calculate the centrality measures, syndicated loan data is used. This enables the authors to measure the centrality of a bank in the syndicated loan market, which in turn serves as a proxy for the connection between banks. The constructed financial network includes 7740 banks and other financial firms. In total, the authors construct six centrality measures using network theory as well as analysis and adjacency matrix. For a more detailed explanation of the construction of the financial network and the calculation of the author's centrality measures, see the paper in question. Additionally, the paper uses data on accounting variables that are determinants of systemic risk contribution. Examples of these are size, leverage, probability of default, stock market beta (systemic risk), VaR, and more. One key conclusion presented in the paper is that banks' centrality matter for systemic risk contributions. However, centrality is found to be important by its effect on other risk measures rather than by its own direct effect in being a determinant of systemic risk contribution. E.g., the authors find that the systemic risk contribution of the probability of default increases with the centrality of an institution. This highlights that current regulation, since it only considers centrality as an isolated component, does not consider that the importance of institutional characteristics is dependent on the centrality of the institution. The authors further find that an institution's beta and its VaR complement the previously established factors for determining systemic risk contributions.

Other research that in some way implements the CoVaR methodology presented in Adrian & Brunnermeier (2016) can be found in Girardi & Ergün (2013), Roengpitya & Rungcharoenkitkul (2011) and Wen et al. (2020)

Previous research shows CoVaR's usefulness for estimating systemic risk contributions of different kinds of entities, the interconnectedness between financial institutions, and negative spillovers. However, neither of them considers the implications of different state variables selection methods and how this might affect the estimates of systemic risk contributions. Simply using the state variables proposed by Adrian & Brunnermeier (2016) neglects the question of what implications different selections might have.

Although, there exists some research that further proposes a method for the selection of variables included in the CoVaR methodology. In *Interconnectedness and systemic risk network of Chinese financial institutions: A LASSO-CoVaR approach*, Xu et al. (2019) wishes to study the interconnectedness between financial institutions in China between 2010-2017. To study the interconnectedness and the systemic risk contribution of different institutions, the authors construct a network of 50 publicly listed financial institutions. To be able to construct the network, the authors introduce a method incorporating the least absolute shrinkage and selection operator (LASSO) into the CoVaR methodology. More specifically, the LASSO-CoVaR approach proposed by the authors combines the CoVaR estimation methodology with LASSO and models CoVaR as a function of institutional balance-sheet characteristics, state variables, and the VaRs of other financial institutions selected by LASSO. Hence, by selecting the significant VaRs (the relevant institutions) for the respective institutions' CoVaR estimation, the LASSO enables the authors to construct a network that showcases the interconnectedness between the included institutions. Furthermore, the authors calculate connectivity measures and later rank institutions by their systemic risk contributions in four different sub-periods. The state variables used in the paper are in line with the ones used in Adrian & Brunnermeier (2016), except that the real estate sector return is not in excess of the financial sector. Institutional characteristics include e.g. leverage, size, and market-to-book value. A central conclusion drawn by the authors is that the impact of other institutions is important when estimating CoVaR and that the interconnectedness should be considered in CoVaR estimation. Additionally, the authors find that interconnectedness in the system tends to be the highest when the system is in distress. This is shown to be especially visible during China's stock market crash in 2015-2016.

4. Methodology

4.1 Systemic risk estimation

4.1.1 Quantile regression

Regressions are used to quantify a relationship between a dependent variable with one or more independent variables, with the most common type of regression being the Ordinary Least Squares (OLS) regression. The OLS-regression models the conditional mean of the dependent variable, where the goal is to minimize the distance between the observed values and the values that the regression line predicts. This essay uses quantile regressions, which instead uses weights that are assigned to the distances between the actual values and the predicted values from the regression line, and the method then tries to minimize the weighted distance (Le Cook and Manning 2013). Hence, OLS tells us about the effect on the *mean* of the dependent variable, given a change in an independent variable (while controlling for other independent variables), while the quantile regression tells us the similar effects but at different points on the distribution (Porter 2015). Appendix B provides a visual representation of the difference between OLS and quantile regression estimate.

Quantile regressions dates back to the late 1970s and was first presented by Koenker & Bassett (1978). Given a real-valued random variable X, with a distribution function of $F(x) = P(X \leq x)$, the quantile τ is given by:

$$F^{-1}(\tau) = \inf\{x : F(x) \geq \tau\} \tag{4.1}$$

Where inf stands for infimum, i.e., the smallest value that x can take on which satisfies the given inequality. Note that by definition, a quantile must lie within the range of 0 to 1 (although it is often mentioned in terms of 0 to 100). In some cases, finding a specific quantile can be relatively straight forward. E.g., if a distribution exhibits a sample size of N = 100 and we want to find the 25^{th} quantile, the sample is first ordered from lowest to smallest and then the 25^{th} observation is our 25^{th} quantile. However, Koenker & Bassett (1978) showed that a specific sample q^{th} quantile could be found using the following equation, where the quantile is the value of β that yields the minimum value for y groups of observations (Porter 2015, p.340):

$$\sum_{i:y_i \geq \beta}^{N} q|y_i - \beta| + \sum_{i:y_i < \beta}^{N} (1-q)|y_i - \beta| \tag{4.2}$$

This optimization approach can be readily extended to find the quantile regression estimator, $\hat{\beta}$, as given by below equation. The quantile regression estimator is again the value of β that yields the minimum value for y groups of observations, where $\mathbf{x_i'}$ is a matrix with independent variables being used in the quantile regression, and $\boldsymbol{\beta}$ is a vector of regression coefficients (ibid., p.341).

$$\sum_{i:y_i \geq \mathbf{x_i'}\boldsymbol{\beta}}^{N} q|y_i - \mathbf{x_i'}\boldsymbol{\beta}| + \sum_{i:y_i < \mathbf{x_i'}\boldsymbol{\beta}}^{N} (1-q)|y_i - \mathbf{x_i'}\boldsymbol{\beta}| \tag{4.3}$$

The method of minimizing is called the *least absolute deviations estimator* (as compared to *least squares estimator* in the OLS regression), and since the equation is not differentiable, it needs to be estimated using linear programming methods (Cameron and Trivedi 2005).

Quantile regressions are often called non-parametric regression since there is no requirement of a specific assumption of the distribution to be able to estimate the parameters. Hence, the usual assumption of normality can effectively be relaxed, making quantile regressions more appealing when using stock market returns (due to the often non-normal nature of equity returns). Furthermore, compared to e.g. OLS, quantile regression is more robust to outliers (Brooks 2019). This makes intuitive sense, since large outliers will affect the mean more than it will affect the median given a large sample size. One of the most common assumptions of quantile regression is that the response variable is independently distributed and homoscedastic, meaning that the variance is constant. Although Brooks (ibid.) argues that this assumption could be relaxed at the cost of additional model complexity.

4.1.2 Estimation of VaR, CoVaR, ΔCoVaR and $\Delta^{\$}$CoVaR

CoVaR and ΔCoVaR can, as previously mentioned, be estimated both unconditionally and conditionally. The unconditional estimation gives a CoVaR estimate for each institution that is constant over time while the conditional estimation yields time-varying

CoVaR estimates. The conditional estimation outlined in Adrian & Brunnermeier (2016), and the one that this essay follow, is conducted by conditioning on a set of state variables.

To be able to estimate the unconditional CoVaR and ΔCoVaR (no time-variation), the first step is to estimate VaR for each financial institution. Using basic historical simulation, VaR_q^i for financial institution i and quantile q was obtained by running a quantile regression on losses of firm i as the dependent variable and a constant as the independent variable. The same was done for the system losses (section five describes this essays procedure for defining the system). The equations are seen below where q is the quantile, and L is the losses for institution i (or system):

$$L_q^i = \alpha_q^i + \epsilon_q^i \tag{4.4}$$

$$VaRq^i = \hat{\alpha_q^i} \tag{4.5}$$

$$L_q^{system} = \alpha_q^{system} + \epsilon_q^{system} \tag{4.6}$$

$$VaRq^{system} = \alpha_q^{\hat{system}} \tag{4.7}$$

Furthermore, estimations of CoVaR were carried out by running a quantile regression with system losses on a constant and the losses of firm i. The estimated $\hat{\alpha}$ and $\hat{\beta}$ from equation 4.8 is used in the estimation of CoVaR (eq 4.9) together with the VaR obtained in equation 4.5. It is important to bear in mind that the obtained β-coefficient can be interpreted as the tail dependency between the financial institution i and the system, i.e., how much the two covary when the institution is in distress.

$$L_q^{system} = \alpha_q^i + \beta_q^i L_q^i + \epsilon_q^i \tag{4.8}$$

$$CoVaR_q^i = VaR_q^{system|L^i=VaR_q^i} = \hat{\alpha_q^i} + \hat{\beta_q^i} VaR_q^i \tag{4.9}$$

To obtain the ΔCoVaR, all equations above were being run two times, using the desired quantile, *0.99*, and the median quantile, *0.50*. This results in equation 2.4 turning into the following equation:

$$\Delta CoVaR_q^i = \hat{\beta}_q^i (VaR_q^i - VaR_{0.5}^i) \tag{4.10}$$

However, since different institutions vary in size, it is more appealing to observe the $\Delta^{\$}$CoVaR, which was calculated by using equation 2.5 from section 2.2.

To allow for time-varying effects of the CoVaR, Adrian & Brunnermeier (2016) suggests, as previously mentioned, using a vector of lagged state variables, $\mathbf{S_{t-1}}$. I.e., VaR and CoVaR become a function of specific state variables. Mendonça & Silva (2018) estimates CoVaR using both lagged and non-lagged state variables and finds no significant difference in the estimates obtained. Therefore, this essay uses state variables lagged one time period, which corresponds to one month. The following equations use the same notations as above but with an added subscript t, which indicates that the estimates vary over time. In the first two regressions, losses of firm i is used as the dependent variable and the vector of state variables together with an intercept acts as the independent variables, seen in equation 4.11. As for equation 4.12, the system losses act as dependent variable and intercept, vector of state variables and losses of firm i acts as independent variables.

$$L_t^i = \alpha_q^i + \gamma_q^i \boldsymbol{S_{t-1}} + \epsilon_{q,t}^i \tag{4.11}$$

$$L_t^{system|i} = \alpha_q^{system|i} + \gamma_q^{system|i} \boldsymbol{S_{t-1}} + \beta_q^{system|i} L_q^i + \epsilon_{q,t}^{system|i} \tag{4.12}$$

The estimated $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\gamma}$ were obtained for each institution, i, and time, t. By including these estimates in the following two equations, the time varying VaR and CoVaR was obtained:

$$VaR_{q,t}^i = \hat{\alpha_q^i} + \hat{\gamma_q^i} \boldsymbol{S_{t-1}} \tag{4.13}$$

$$CoVaR_{q,t}^{system|i} = \alpha_q^{sys\hat{t}em|i} + \gamma_q^{sys\hat{t}em|i} \boldsymbol{S_{t-1}} + \beta_q^{sys\hat{t}em|i} VaR_{q,t}^i \tag{4.14}$$

As with the unconditional case, all quantile regressions were run two times using $q = 0.99$ and $q = 0.50$ in order to calculate the conditional ΔCoVaR from:

$$\Delta CoVaR_{q,t}^{system|i} = \hat{\beta_q^{system|i}}(VaR_{q,t}^i - VaR_{0.5,t}^i) \quad (4.15)$$

The $\Delta^{\$}$CoVaR was obtained by once again using equation 2.5 for the specific firm at each specific time. The above procedure was done in the same way for our different selection methods (A&B, Lasso and random), however, for PCA, the created principal components were instead used as the state variables. Note that for simplicity, this essay uses the term *variable selection* (as in the research questions) for describing the selection of different state variables and the usage of Principal Component Analysis. Hence, variable selection refers to both the actual selection of different state variables, but also dimensionality reduction, since PCA is not a variable selection method but rather a method for reducing dimensionality.

4.2 Selection of state variables

4.2.1 Adrian & Brunnermeier

In the final version of the paper on CoVaR (Adrian and Markus K. Brunnermeier 2016), the authors uses, as previously mentioned, seven state variables to capture the time-variation in the systemic risk estimation. They gathered data from the US for *change in 3m treasury yield, change in slope of yield curve, short term TED-spread, change in credit spread, S&P500 market return, real estate sector return and equity volatility.* For this essay, data on the European equivalents of these seven variables was collected and used in the conditional CoVaR estimation (more in section 5.1). Since these state variables are the ones used by Adrian & Brunnermeier and a lot of other systemic risk research, this essay finds it highly reasonable to include these variables as a state variable selection method.

4.2.2 Regularization: Lasso

The *Least Absolute Shrinkage and Selection Operator*, also called Lasso, is a statistical method used for regularization and variable selection, which was first proposed by Tibshirani (1996). The Lasso is chosen as one of the methods in this paper due to it being one of the most established machine learning selection methods and its characteristics of being a selection operator, unlike Ridge regression, which only shrinks the coefficients (James et al. 2021, p. 237). Given a set of independent variables, the Lasso shrinks some coefficients and sets others equal to zero. I.e., it tries to emphasize the variables that explain the most variation in the response variable. This can be particularly helpful when dealing with large dimensionality.

Given a vector of independent variables, $\mathbf{x^i} = (x_{i,1}, x_{i,2} \dots x_{i,j})^T$ for $i = 1, 2, \dots, N$, they are first standardized such that equation 4.16 and equation 4.17 holds.

$$\sum_{i=1}^{N} \frac{x_{i,j}}{N} = 0 \tag{4.16}$$

$$\sum_{i=1}^{N} \frac{x_{i,j}^2}{N} = 1 \tag{4.17}$$

Now if we let a vector of coefficients be $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2, ...\hat{\beta}_j)^T$, then the Lasso parameters will be given by equation 4.18 (ibid.).

$$(\hat{\alpha}, \hat{\beta}) = argmin\{\sum_{i=1}^{N}(y_i - \alpha - \sum_{j=1}^{p} \beta_j x_{i,j})^2\} + \lambda \sum_{j=1}^{p} |\beta_j| = RSS + \lambda \sum_{j=1}^{p} |\beta_j| \tag{4.18}$$

RSS is the residual sum of squares, which measures the level of variance in the residuals of the regression model. Furthermore, the second term is the *penalty term* where if the tuning parameter, λ, is sufficiently large, some coefficient will be set to zero.

Hence, an important matter still remains, which is the selection of the optimal tuning parameter. James, et al, (ibid.) argues that cross-validation provides a simple way to tackle the problem. The first step is to randomly split all data into a training and a validation set. The model is then fitted on the training data and the dependent variable is predicted using the estimated parameters and compared to the *true* dependent variable in the validation set. A loss function is used to calculate the error between the two. This procedure is done for a range of different values of the tuning parameter, and the optimal parameter is the one resulting in lowest value of the loss function.

Using Lasso as a variable selection method in this essay, all possible state variables were included as independent variables and system losses as the dependent variable. In essence, the underlying calculations are based on an OLS regression, and not a quantile regression.

This is due to the fact that an econometric framework for incorporating the quantile regression into the Lasso regression is not, to the author's knowledge, fully developed. The number of cross-validations was set to five, the loss function was *mean squared error*, and the parameter was allowed to be in the range of 0.0001 and 2. A validation check of the search interval was conducted to make sure that one of the endpoints was not chosen as the *best-fit parameter*. Furthermore, the non-zero Lasso variables were then used in the calculation of conditional CoVaR, as outlined in section 4.1.2. More specifically, the state variables selected by the Lasso regression, are then used in equation 4.11 and onward.

4.2.3 Principal component Analysis

One of the more popular ways of deriving a low-dimensional set of variables from a large set of variables is to use the Principal Component Analysis, PCA (James et al. 2021, p. 252-253). Due to these characteristics, it is a useful additional method for the research questions. Since there are 24 state variables (20 after accounting for high correlation, more in section 5.2), the PCA is able to extract the most useful information while keeping dimensionality low. Hence, even though PCA is not a variable selection method it is relevant for this essay because it produces another set of time-varying estimates of CoVaR while also reducing the risk of overfitting.

Using a table of data, the unsupervised machine learning method aims to fulfill four goals, which are: to extract the most important information from the data table, suppress the size of the data set by keeping only this important information, simplify the description of the data set and analyze the structure of the observation and the variables (Abdi and Williams 2010).

Assuming a set of k explanatory variables denoted by $x_1, x_2 \dots x_k$, the PCA uses these to create k new uncorrelated variables, $p_1, p_2, \dots p_k$:

$$\begin{aligned} p_1 &= a_{11}x_1 + a_{12}x_2 + \dots + a_{1k}x_k \\ p_2 &= a_{21}x_1 + a_{22}x_2 + \dots + a_{2k}x_k \\ &\dots \\ p_k &= a_{k1}x_1 + a_{k2}x_2 + \dots + a_{kk}x_k \end{aligned} \tag{4.19}$$

The requirement of the factor loadings, a, is that the sum of the squares equal to one, i.e.:

$$\sum_{j=1}^{k} a_{i,j}^2 = 1 \quad \forall \quad 1, ..., k \tag{4.20}$$

The components are ordered in descending importance; hence, a specific number of principal components can be chosen that explain a large part of the variance.

Before doing the Principal Component Analysis, standardization was done on all the state variables followed by a method of choosing the number of components. One common method is selecting a specific threshold for how much variance the components should explain. In this paper, a threshold of approximately 95% explained variance was chosen, which corresponded to five explanatory variables, or principal components. Using the same methodology as previously stated, the conditional CoVaR was calculated using these five components. James, et al., (2021) emphasizes that while PCA reduces the dimension of explanatory variables, it is not a method of variable selection, since all principal components are a linear combination of all original variables.

4.2.4 Randomization

Lastly, a state variable selection method using randomization was implemented. Five, seven, and nine random state variables was drawn and after each random draw, the estimation of conditional CoVaR was done following the same methodology as the other methods. To make the randomization more robust, the randomization was done five times for each number of state variables and an average of the conditional CoVaR was calculated.

5. Data

5.1 Data collection

The first step in the data collection was to choose which banks to include in the data set. Since this paper estimates systemic risk contributions in the European banking sector, banks that at least offer both retail and corporate banking services from 31 European countries were gathered. In total, equity prices of 141 banks were gathered on a monthly basis ranging from 31^{st} of December 2002 to 30^{th} of September 2022. The start and end date was set due to several state variables not having data before or after these dates. A summary of each bank and its ticker, weight in system and market capitalization can be found in Appendix A. Using equations 5.1 and 5.2, prices for firm i, at time t, were converted to returns, which was further multiplied with -1 to obtain losses, L. Returns were used due to their stationary nature and prices being nominated in several different currencies.

$$r_{i,t} = \frac{P_{i,t}}{P_{i,t-1}} - 1 \tag{5.1}$$

$$L_i^t = -1 * r_{i,t} \tag{5.2}$$

Furthermore, monthly market capitalization data for all banks were gathered during the same time period, all being nominated in euros. Equity and market capitalization data were downloaded through Bloomberg. For some institutions and for specific dates, both price data and market capitalization data were missing. If the missing data were between two data points this was handled using linear interpolation following the methodology of Xu, et al., (2019). The same was done for GDP, which is released on a quarterly basis, while this paper uses a monthly data frequency. The equation for interpolation can be seen below, where m is the number of months in a quarter and s is each time step:

$$P_{t+\frac{s}{m}} = (1 - \frac{s}{m})P_t + \frac{s}{m}P_{t+1} \tag{5.3}$$

If the missing data were at the end of a data series, and linear interpolation could not be performed, returns were set to zero. This is because the principal component analysis could not be performed without consistency in the number of data points per bank. It is worth mentioning that these cases were few and for banks with relatively small system weights. Hence, the conducted methods for handling the missing data are believed to have a non-significant effect on the results of this essay.

By weighing all banks at each time period, t, with their respective markets capitalizations (equation 5.4), a system (representing the banking sector in Europe) was generated. Furthermore, system losses were obtained by multiplying the weight of firm i at time t with the loss of firm i at time t, from equation 5.5 (MC representing market capitalization).

$$w_{i,t} = \frac{MC_{i,t}}{\sum_{i=1}^{n} MC_{i,t}} \tag{5.4}$$

$$L_t^{System} = \sum_{i=1}^{n} w_{i,t} * L_t^i \tag{5.5}$$

Following Adrian & Brunnermeier (2016), to obtain time-variation in the estimates of VaR, CoVaR, ΔCoVaR and $\Delta^{\$}$CoVaR, data was gathered for 24 state variables. Except for the seven state variables used in Adrian & Brunnermeier (2016), the percentage change in Eurozone inflation was also added following the works of (Hanif et al. 2020). Furthermore, other macroeconomic variables such as GDP, CPI and industrial production index was added as argued by Roengpitya & Rungcharoenkitkul (2011). Petrella, et al. (2019), suggests using exchange rate as state variables, and due to this, the return of EUR/USD was added. Furthermore, several other macro-economic variables that account for market specific information as well as captures and controls for the conditional moments of asset returns were gathered. A comprehensive list with a description as well as a motivation of each of them can be found in Appendix C. Even though all variables are believed to be important, the selection methods are still used to reduce the risk of overfitting the data. The state variable data was obtained for Europe, and if no such data existed, data for Germany was gathered as a proxy. The assumption that data for Germany is representative for Europe is believed to be reasonable due to the spillover from the German economy to the rest of Europe as well as the high correlation between German and non-German state variables (Borri et al. 2012). For one of Adrian & Brunnermeiers state variables, *change in credit spread*, no data for an European investment grade index were available, and hence, data for the US equivalent was gathered. Data for the state variables were collected through a combination of Bloomberg, FRED (Federal Reserve Economic Data) and ECB. Several banks that went out of business (or were merged with another bank) during the time period were included in the data set, hence the data does not suffer from survivorship bias.

As mentioned, Adrian & Brunnermeier (2016) uses weekly stock return data. This paper however uses monthly data. This is due to Bloomberg only supplying ten-year historical data for a weekly frequency. The implication of this is that some variation occurring between two data points is left out, e.g., if there are two very volatile weeks in between two monthly observations, these two weeks will not affect the result. However, if a weekly frequency were to be used and ten years of historical data were obtained, a lot of variation over time in the selected state variables would be lost. For example, the financial crisis would not be part of the data set. Hence, there is a trade-off to be considered. Ultimately, a time span of just under 20 years was chosen on a monthly frequency due to this essays' focus on the time series dimension of the CoVaR estimation.

5.2 Descriptive statistics

Table 5.1 shows descriptive statistics for the losses including all banks, the 40 largest banks, the 40 smallest banks and the system. Since it is shown in losses, a negative loss is equivalent to a gain, and vice versa. Also, important to notice is that the system consists of losses for all banks, weighted after their market capitalization. Since large banks tend to have lower losses (larger gains) than the smaller banks and also higher weight, the system will exhibit lower losses than the aggregate of all banks (where they are equally weighted). For all banks and for the 40 smallest banks, the maximum loss is 100%, which means that the bank went out of the public market (bankruptcy, buyout etc.). A distribution of the system losses is shown in figure 5.1 below.

Descriptive statistics, Losses

	Count	Mean	Median	Std Dev	Min	Max
All banks	28693	-0.32%	-0.14%	10.66%	-182.53%	100.00%
40 largest	9058	-0.36%	-0.41%	10.93%	-169.43%	70.31%
40 smallest	7578	-0.25%	0.00%	10.17%	-121.01%	100.00%
System	**238**	**-0.42%**	**-0.96%**	**6.99%**	**-35.99%**	**26.51%**

Table 5.1: The table shows the descriptive statistics for losses, calculated as the negative of percentage returns. The table shows the number of data points, mean, median, standard deviation, minimum and maximum value of four different samples; all banks, 40 largest banks, 40 smallest banks (based on market capitalization per March 28^{th} 2023) and system.

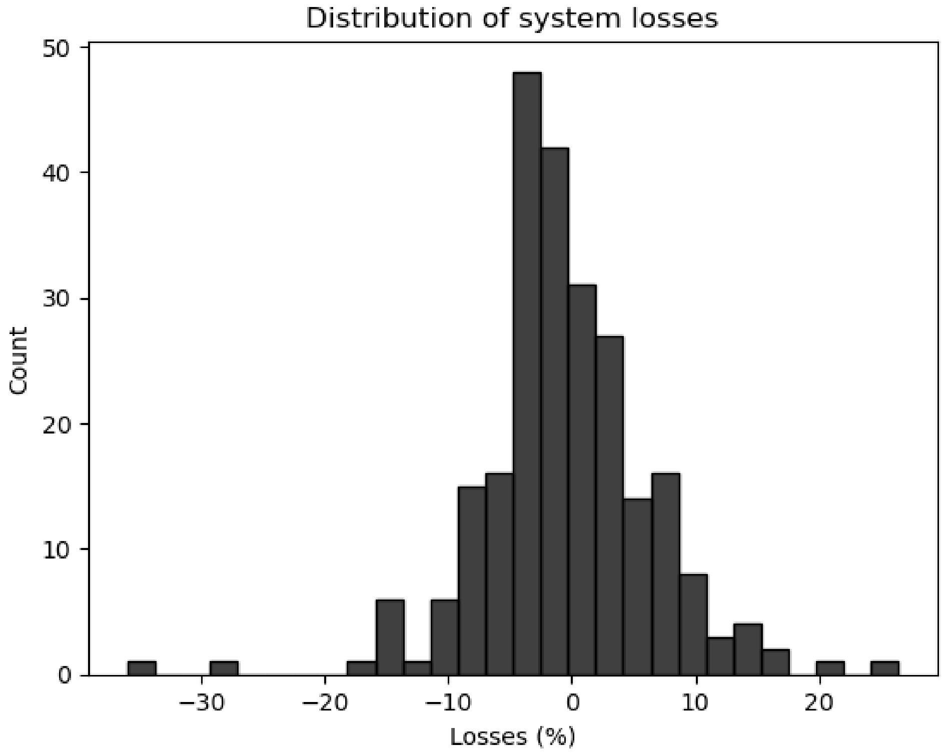


Figure 5.1: The figure shows the distribution of system losses, calculated as the negative of percentage returns. In total, there are 238 observations.

To ensure stationarity, all variables were expressed in percentage changes, and an augmented Dickey-Fuller, as well as KPSS test for stationarity was conducted. Since a large number of hypothesis were tested for stationarity, problems connected to multiple hypothesis testing had to be considered. When conducting 24 tests using a significance level of 5%, it is probabilistically possible that stationarity is falsely rejected in at least one of the cases. To account for this, a Bonferroni correction was used, which basically shrinks the p-value for hypothesis testing (Bonferroni 1935). The results of the stationarity tests showed that all variables was stationary.

Furthermore, a test for multicollinearity was conducted, called the VIF-test. A value greater than ten usually implies that there is a problem with multicollinearity. There was no indication of multicollinearity in the test results, since all variables had a VIF-value well below ten (highest VIF-value in the sample was 3.15). Even though the VIF-test shows no obvious problems, there can still be high correlation between variables, which can be problematic in the estimation (especially in the randomization method). Therefore, a correlation matrix was calculated and a threshold of 0.7 was set. Since four variables had a correlation greater than 0.7, these were dropped from the data set. The affected state variables were RBANK, VBANK, VOLB and VRE. A correlation matrix with the remaining variables can be found in Appendix E. Hence, after the data had been processed and cleaned, the data set had 141 banks and 20 state variables, which in total amounts to 33 453 datapoints. Table 5.2 below shows descriptive statistics for the remaining 20 state variables, and a density plot of each variable is found in Appendix D. Although the majority of the state variables are denoted in percentage change, it is worth noting that the state variables within the CoVaR methodology are primarily used to obtain time varying systemic risk estimates. As previously mentioned, the variables are conditioning variables rather than systemic risk factors. Hence, the unit measures and the coefficients are not of central importance.

Descriptive statistics, State variables

	Count	Mean	Median	Std Dev	Min	Max
TMGB	238	-0.0095	0.0035	0.1830	-1.8390	0.7300
CSYC	238	-0.0030	-0.0100	0.2385	-0.7140	1.6610
TED	238	0.3932	0.2935	0.5097	-0.3673	3.5301
CCS	238	0.0016	-0.0095	0.2489	-0.7710	1.7220
R600	238	0.0037	0.0095	0.0419	-0.1480	0.1373
REER	238	0.0040	0.0073	0.0565	-0.1976	0.1462
V600	238	16.3630	13.6600	9.2044	5.5800	69.4600
VOL	238	0.0168	-0.0151	0.2128	-0.4187	1.1127
GDP	238	0.0010	0.0016	0.0058	-0.0337	0.0300
CPI	238	0.0015	0.0011	0.0040	-0.0103	0.0198
ECBA	238	0.0103	0.0062	0.0332	-0.0878	0.2895
RECI	238	0.3151	0.0000	0.4655	0.0000	1.0000
EXCH	238	0.0003	-0.0009	0.0273	-0.0972	0.1010
UNEM	238	-0.0092	0.0000	0.1102	-0.3000	0.5000
OIL	238	0.0114	0.0198	0.1146	-0.5424	0.8838
CONF	238	-0.0002	0.0000	0.0021	-0.0100	0.0051
M2	238	0.0049	0.0043	0.0058	-0.0116	0.0318
INDU	238	0.0824	0.1000	2.2129	-19.3000	14.2000
GOLD	238	0.0083	0.0035	0.0487	-0.1689	0.1301
RRE	238	0.0523	-0.0005	0.3596	-0.7504	2.2645

Table 5.2: The table shows descriptive statistics for the state variables. The state variables are expressed in different units, e.g., basis points changes, percentage changes and changes in percent. For more detailed description of each state variable, see Appendix C

6. Empirical results

6.1 Comparison of estimates

This essay chooses to present the empirical results in terms of $\Delta^{\$}$CoVaR (more specifically $\Delta^{€}$CoVaR since the market capitalization of a bank is denoted in euro) and not ΔCoVaR because of one main reason. The indicator-based approach used by regulators described in section 2.3 acknowledges size as an important factor for determining systemically important banks, and by focusing on $\Delta^{\$}$CoVaR, this essay also emphasizes the size difference between banks. To present and rank the banks with the highest ΔCoVaR would, according to the authors, be uninformative since a very small bank can have a larger ΔCoVaR than a G-SIB but obviously, the G-SIB is more systemically important.

	A&B	**Lasso**	**PCA**	**5 random**	**7 random**	**9 random**
HSBC	10,676	9,407	12,377	12,573	10,042	9,420
UBSG	4,945	8,041	3,617	5,757	5,721	5,420
BNP	7,504	7,685	4,989	8,213	7,378	6,679
SAN	7,134	10,193	4,975	8,222	7,065	7,691
ISP	4,959	4,338	2,286	4,829	3,892	4,385
INGA	5,933	6,390	2,903	6,120	5,992	5,941
BBVA	4,878	5,828	2,739	4,980	4,828	4,665
LLOY	3,863	6,174	2,148	5,709	4,569	4,761
NDA	1,946	3,252	1,588	2,887	2,276	2,467
UCG	3,182	4,405	1,240	3,797	3,151	2,965
ACA	3,089	3,507	1,728	2,991	2,921	2,811
NWG	3,902	5,649	975	4,908	3,114	3,883
CABK	1,591	1,940	977	1,747	1,537	1,574
KBC	1,509	1,611	661	1,611	1,203	1,414
DNB	1,633	1,892	1,309	1,558	1,574	1,806
BARC	3,928	4,368	2,155	4,548	3,641	3,684
SEB	1,790	1,826	1,132	1,883	1,442	1,606
STAN	2,761	2,705	1,528	3,639	2,969	2,637
DBK	1,902	2,018	1,390	2,929	2,515	2,815
SWEDA	1,668	1,828	863	1,581	1,247	1,376

Table 6.1: The table shows the calculated mean of the $\Delta^{\$}$CoVaR estimates in millions of euros for the 20 largest banks, as of March 28^{th} 2023 for each of the methods. The mean is calculated by taking the average of the monthly estimates of $\Delta^{\$}$CoVaR during the time period 2002-12-31 to 2022-09-30.

Table 6.1 shows $\Delta^{\$}$CoVaR estimates in MEUR generated by the different variable selection methods. The banks presented are the twenty banks with the highest market capitalization as of 28^{th} of March 2023 and the estimates are a mean of the obtained monthly $\Delta^{\$}$CoVaR estimates during the essays' selected time period. A bold bank indicates that the bank is included in the G-SIB list as of November 2022. Appendix F shows the selected state variables used in each of the estimations. [1]

What can be seen in table 6.1 is that the estimates are dependent on the selected state variable selection method and that the $\Delta^{\$}$CoVaR estimates clearly differs between the methods chosen to capture time variation. Looking at the table, it is not unusual that estimates differ by more than 100% between methods. This is the case when comparing the PCA with other selection methods since using PCA generates noticeably lower $\Delta^{\$}$CoVaR estimates for all banks in the table except for HSBC. By calculating the difference between the highest and lowest estimates generated by the methods for a specific bank and then expressing this number as an percentage of the average estimate for the same bank showcases the estimate dispersion. The average of this calculation for all banks in table 6.1 including all methods is 131.5%. Estimates generated by the methods excluding PCA are more in line with each other but visible differences still appear. The same calculation described above yields an average of 67.9% when excluding PCA. Regarding estimate differences between methods, it can also be observed in table 6.1 that the selection methods A&B, five random, seven random and nine random seem to showcase the same general trend in the estimates and the dispersion measure including only these four methods yields an average of 24%. It should however be mentioned that the differences in the estimates seem to be somewhat dependent on the bank, where some banks showcases larger differences between methods (Santander, Lloyds and NatWest Group) while other banks have estimates with smaller differences (SEB and DNB).

Except for PCA generating the lowest estimates, some additional patterns can be observed in table 6.1. For more than half of the included banks, Lasso generates the highest $\Delta^{\$}$CoVaR. More specifically, this happens for 12 out of 20 banks.

[1] The estimation results for all 141 banks are available upon request.
E-mail: baltsar.lindgren@gmail.com, christian@hovstadius.com

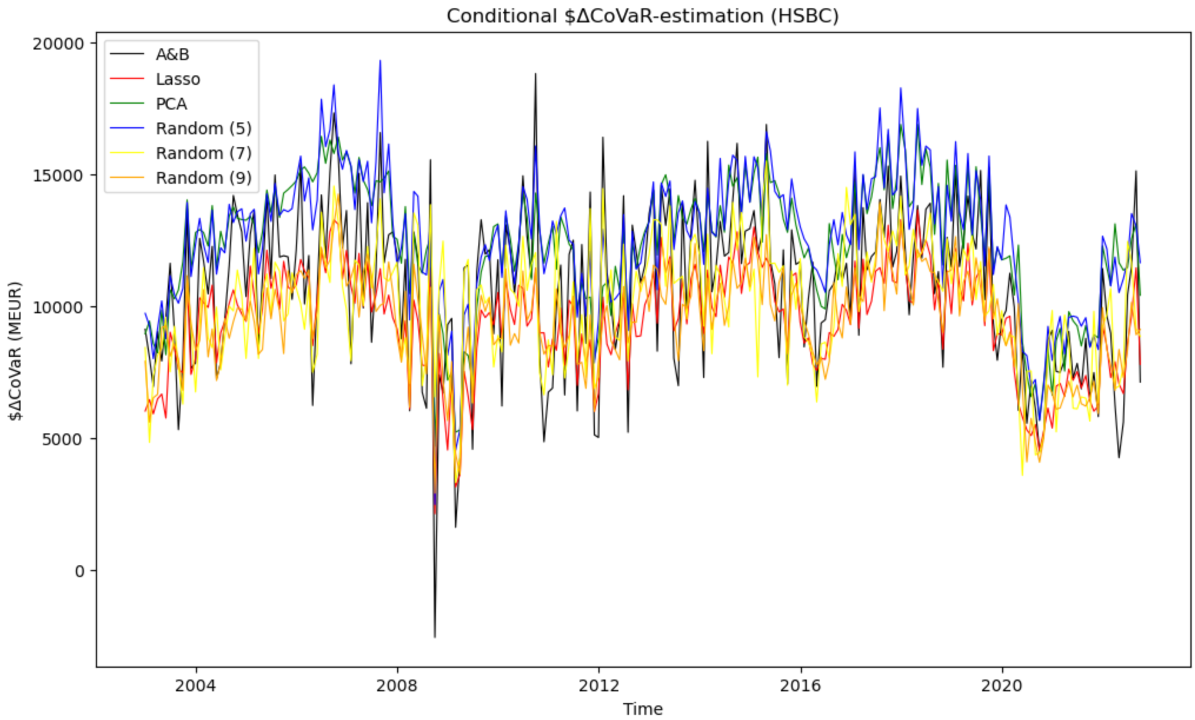


Figure 6.1: The figure plots HSBC's systemic risk contribution measured as monthly $\Delta^{\$}$CoVaR during the time period 2002-12-31 to 2022-09-30.

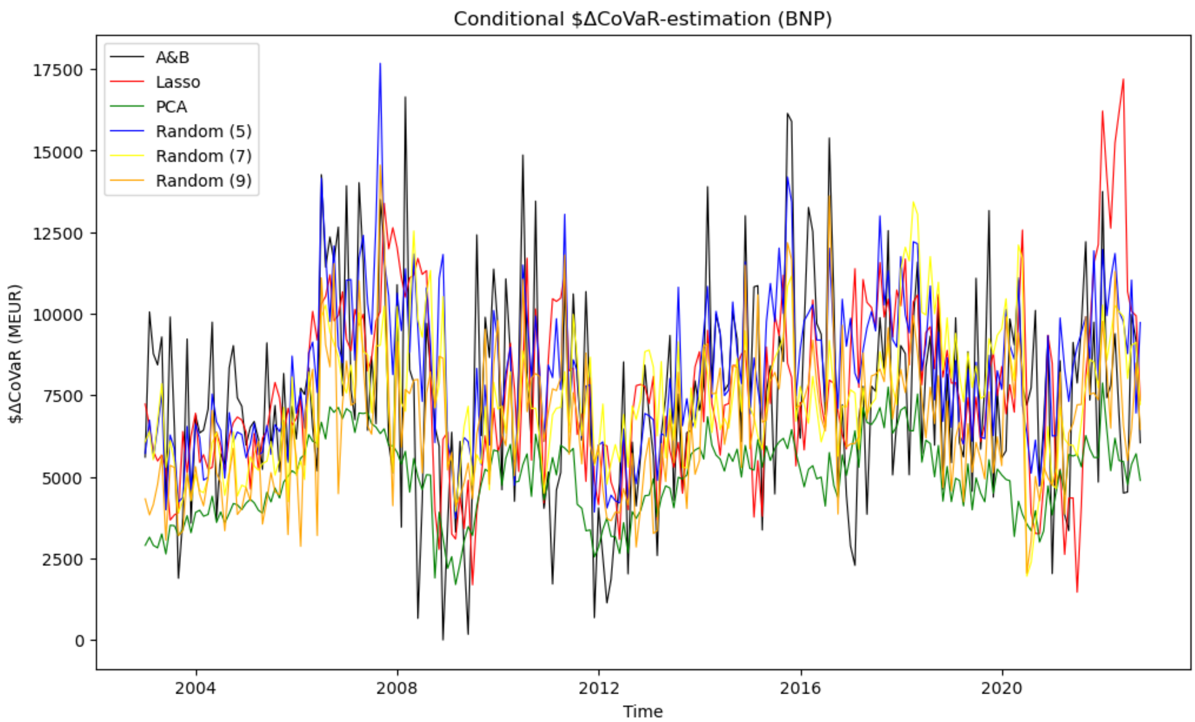


Figure 6.2: The figure plots BNP Paribas' systemic risk contribution measured as monthly $\Delta^{\$}$CoVaR during the time period 2002-12-31 to 2022-09-30.

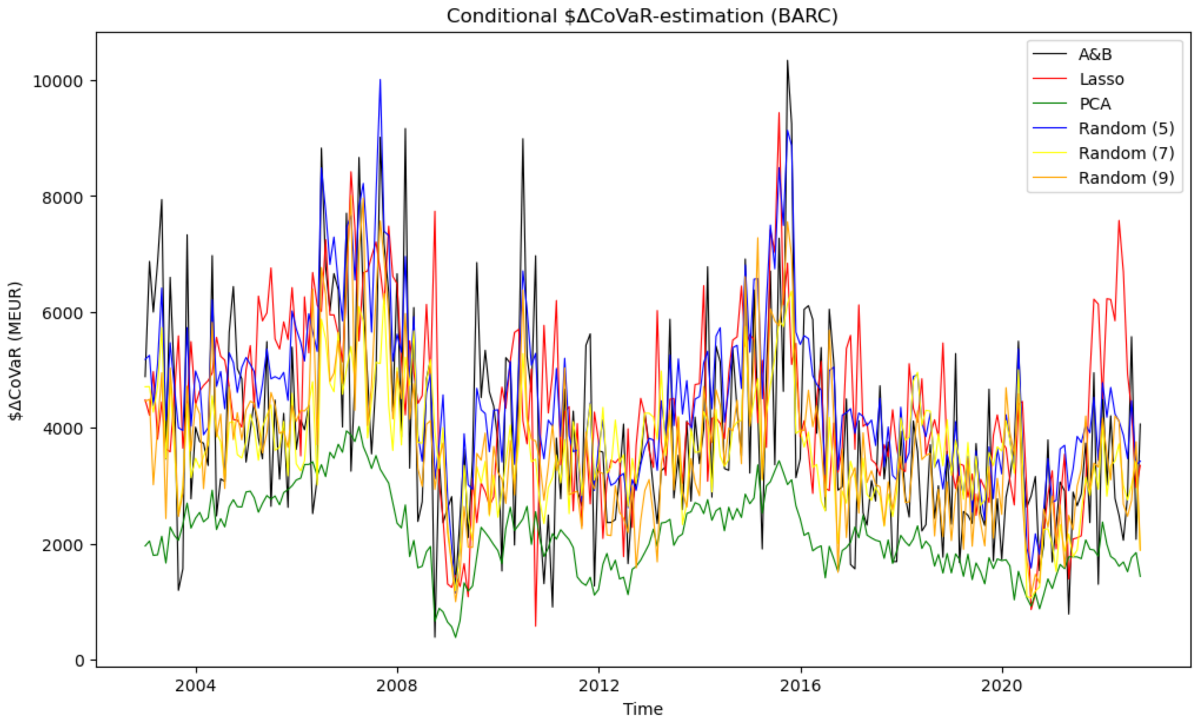


Figure 6.3: The figure plots Barclays' systemic risk contribution measured as monthly $\Delta^{\$}$CoVaR during the time period 2002-12-31 to 2022-09-30.

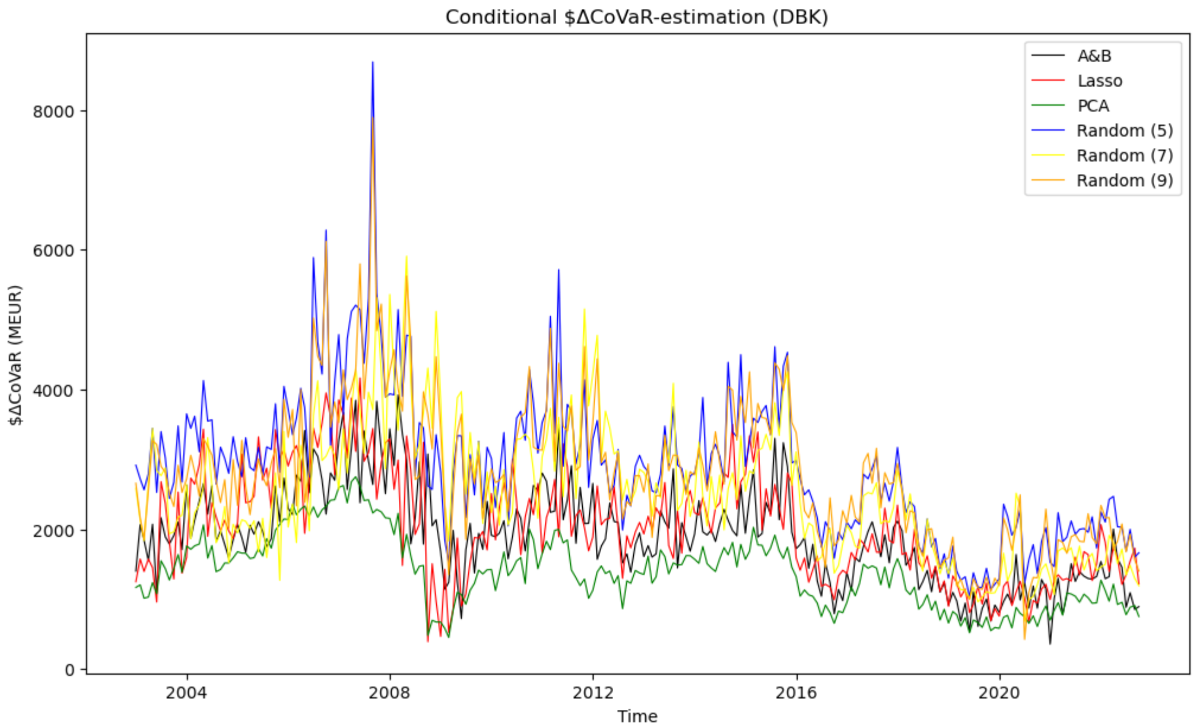


Figure 6.4: The figure plots Deutsche Bank's' systemic risk contribution measured as monthly $\Delta^{\$}$CoVaR during the time period 2002-12-31 to 2022-09-30.

Figures 6.1 - 6.4 plots the $\Delta^{\$}$CoVaR estimates, once again in MEUR, generated by the different state variable selection methods over time. The plots show this for the four European banks that are placed in any of the three highest buckets on the G-SIB list as of November 2022. Much of what can be seen in the above table 6.1 is further showcased by looking at the development of systemic risk contributions over time for the four banks. The estimates obtained by the PCA once again seem to be at noticeably lower levels than the other estimates, with HSBC being the exception. Looking at figure 6.1, one can also observe that after 2008, the Adrian & Brunnermeier state variables produce a $\Delta^{\$}$CoVaR estimate below zero and hence a negative systemic risk contribution. This implies, in theory, that during this time period the risk of the financial system decreased rather than increased when HSBC was in distress.

What is also shown by figures 6.1 - 6.4 is that even though the different methods vary in their level of estimates, in general, they seem to be responsive in the same ways and follow the same trends. For all methods for the four banks, there are clear build-ups in systemic risk contributions before 2008 (Financial crisis), 2015 (Chinese stock market crash), and 2020 (Covid-19 outbreak), followed by drastic decreases in the $\Delta^{\$}$CoVaR estimates. However, the PCA once again is seen as a clear exception, and in periods of substantial increases in systemic risk contributions, the PCA is not as responsive, producing less drastic increases. For HSBC, the PCA however seems to be more in line with the other methods. Although, it should be mentioned that all the methods follow similar trends seems to be the case in only very broad general terms, at least when looking at BNP Paribas, Barclays, and Deutsche Bank. It is not unusual to observe that methods differ in the direction of systemic risk contribution changes, which implies that different methods can estimate positive or negative changes for the same period. E.g., this is visible around 2020 for both BNP Paribas and Barclays where all methods seem to estimate increases while the PCA yields a decrease in its estimate. Other clear disagreements can also be found after 2020 for HSCB, between the Adrian & Brunnermeier estimates and some of the other methods. Also, for BNP Paribas, the Adrian & Brunnermeier estimates before 2012 and after 2016 disagree with the other methods. These are just examples, but the main result to be emphasized is that the state variable selection methods can yield different signs of systemic risk contribution changes during the same estimation period for the same bank.

What probably already has become apparent for the reader is the potential impact the change in market capitalization has on the $\Delta^{\$}$CoVaR estimates. E.g., in the Deutsche Bank graph, it's clear that the bank's systemic risk contribution has been steadily decreasing, which could be due to its market capitalization decreasing by approximately 31% during this essay's chosen time period. However, this must not be the only explanation. BNP Paribas' market capitalization decreased more in percentage terms than Deutsche Bank's during the same time period, but the $\Delta^{\$}$CoVaR estimates does not show the same decreasing trend.

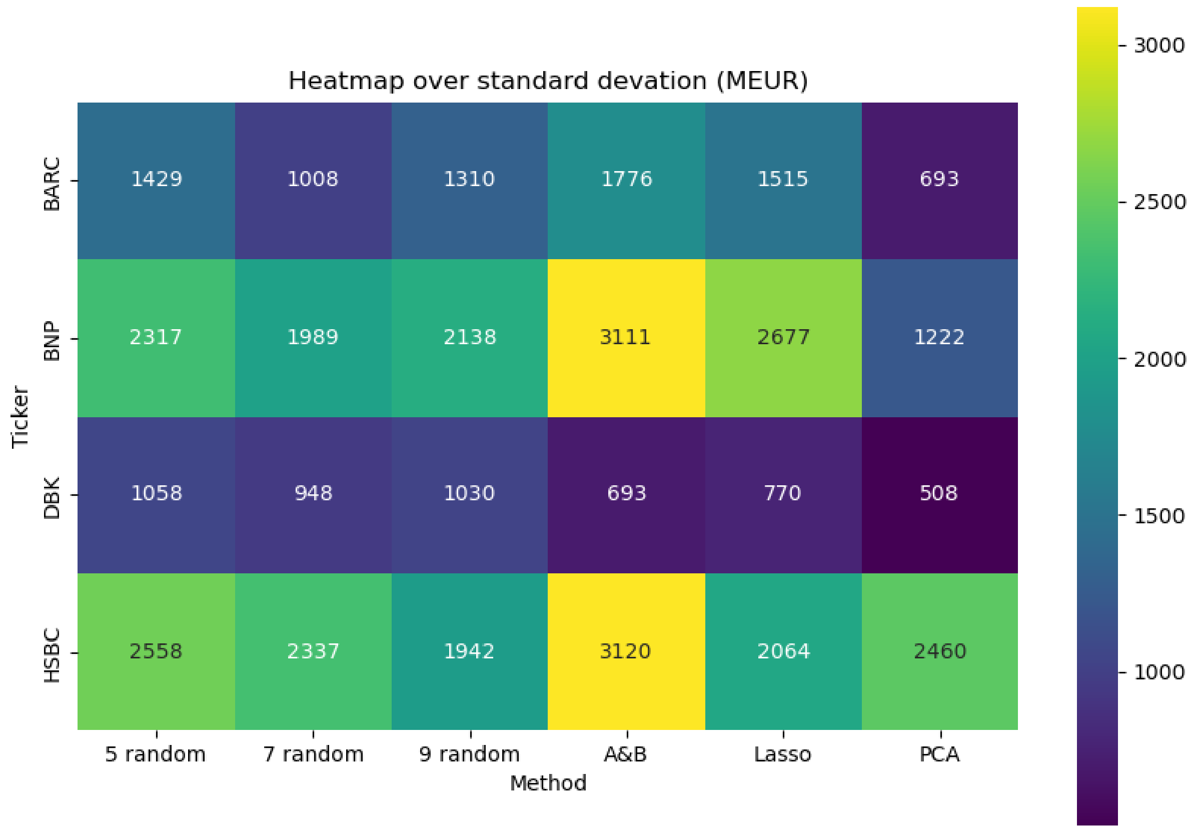
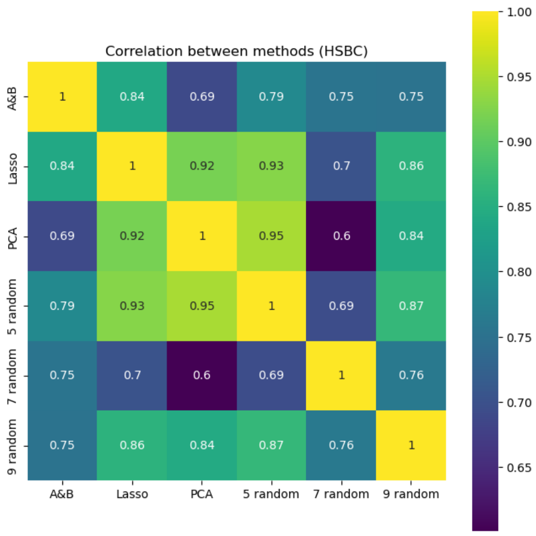


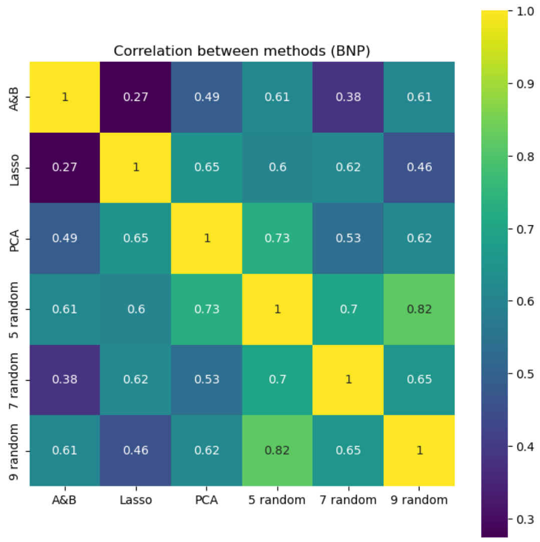
Figure 6.5: The figure shows the standard deviation in millions of euros generated by the different state variables selection methods for the four banks HSBC, BNP Paribas, Barclays and Deutsche Bank.

Figure 6.5 shows the standard deviation across the methods in MEUR for the four selected banks. This gives an understanding of the volatility of the estimates generated by the different state variable selection methods and wishes to quantify the volatility that can be visually inspected in figure 6.1 - 6.4. Looking at the standard deviation across measures for the same bank, the PCA yields the lowest standard deviation for all banks except for HSBC, while the estimates obtained by the Adrian & Brunnermeier state variables show the highest standard deviation for all the selected banks except for Deutsche

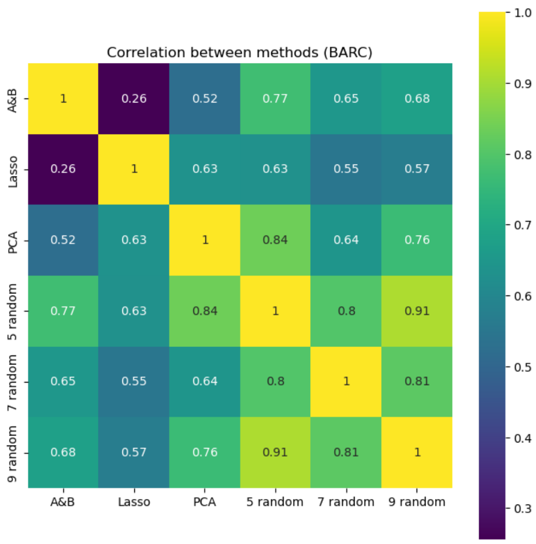
Bank. These numbers can be visually observed in figure 6.1 - 6.4 where the PCA clearly showcases lower volatility while Adrian & Brunnermeier generates a more volatile time series. This trend stays visible when looking at the method standard deviation for the twenty largest banks, shown in Appendix H. Looking at these banks, the PCA has the lowest volatility for all banks except for HSBC, while the Adrian & Brunnermeier estimated volatility is the highest for nine out of twenty banks.



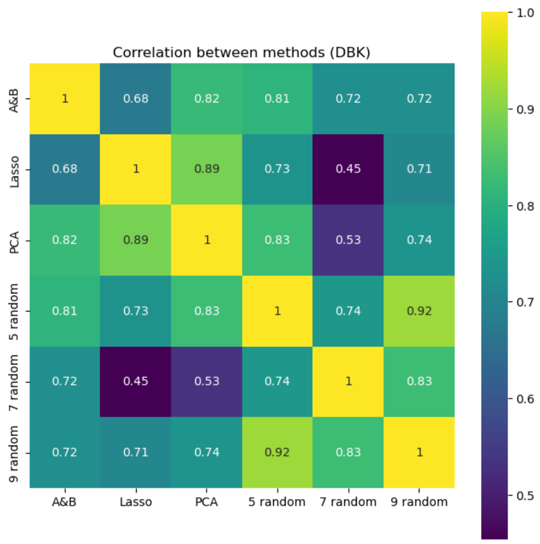
(a) HSCB



(b) BNP Paribas



(c) Barclays



(d) Deutsche Bank

Figure 6.6: The figures shows the Pearson correlation coefficients between the $\Delta^{\$}$CoVaR estimates generated by the different state variables selection methods. This is shown for the four banks HSCB, BNP Paribas, Barclays and Deutsche Bank.

Figures 6.6a - 6.6d above shows the Pearson correlation coefficient for the time series of the

$\Delta^{\$}$ CoVaR for each of the methods for each of the four G-SIB banks. The interpretation is that a high correlation between two methods, e.g., lasso and PCA, means that the estimated conditional $\Delta^{\$}$ CoVaR tend to be similar over the whole time period. In general, the three random selection methods are fairly highly correlated with a correlation often between 0.7 and 0.9. For HSBC, it can be seen that lasso, PCA and 5 random exhibit a high correlation between 0.92 and 0.95. For both BNP Paribas (BNP) and Barclays (BARC), Lasso and Adrian & Brunnermeier's method have the lowest correlation of 0.27 and 0.26 respectively. Apart from the random selection methods, there is no clear visible trend over what methods tend to have a high or low correlation. Note that these correlations offer some nuance to the presented mean estimates in table 6.1 above, which suggested that the methods A&B, five, seven and, nine random generate estimates in the same ball park. However, in some cases, the correlation of the whole time series does not suggest similar results, at least when looking at the correlation between A&B and the random methods. An example of this is the correlation between A&B and seven random for Barclays, where the correlation coefficient is at 0.65 while the mean estimate has a difference of approximately 10%. For BNP Paribas, the same observation can be made, where A&B and seven random generates highly similar mean estimates, while at the same time showcasing a correlation of 0.38. The same inconsistency between the mean estimate and the time series correlation can be observed in the other direction as well. Looking at Deutsche Bank, the correlation coefficient between A&B and five random is 0.81 while the five random estimate is 50% larger than the A&B estimate.

6.2 Comparison of rankings

For each of the selection methods, the banks were ranked after their estimated $\Delta^{\$}$ CoVaR (calculated as the mean of the time-series for each bank). By taking four different subsets of the time period, four different sets of rankings were obtained. This section starts by showing the rankings using the whole time period for calculation, followed by period one using January 1^{st} 2005 to December 31^{st} 2006, period two using January 1^{st} 2013 to December 31^{st} 2014 and period three using January 1^{st} 2020 to December 31^{st} 2021. The second table for each period shows the 20 largest banks (based on March 28^{st} 2023 market capitalization) and how their ranking changes for each method. For all periods, the G-SIB list at the end of the respective period was used, and the banks included in the list are shown in bold. However, the list was first published by the Financial Stability Board in 2011 and hence, no G-SIB list for period 1 (2005-2006) was available. Therefore, the G-SIB list as of 2011 was used for period one, G-SIB list as of 2014 for period two, G-SIB list as of 2021 for period three and G-SIB list as of 2022 for the whole sample period.

Rank	**A&B**	**Lasso**	**PCA**	**5 random**	**7 random**	**9 random**
1	**HSBC**	**SAN**	**HSBC**	**HSBC**	**HSBC**	**HSBC**
2	**BNP**	**HSBC**	**BNP**	**SAN**	**BNP**	**SAN**
3	**SAN**	**UBSG**	**SAN**	**BNP**	**SAN**	**BNP**
4	**INGA**	**BNP**	**UBSG**	**INGA**	**INGA**	**INGA**
5	ISP	**INGA**	**INGA**	**UBSG**	**UBSG**	**UBSG**
6	**UBSG**	LLOY	BBVA	LLOY	BBVA	LLOY
7	BBVA	BBVA	ISP	BBVA	LLOY	BBVA
8	**GLE**	NWG	**BARC**	NWG	**CSGN**	**CSGN**
9	**CSGN**	**CSGN**	LLOY	ISP	ISP	ISP
10	**BARC**	**UCG**	**CSGN**	**BARC**	**BARC**	NWG
11	NWG	**BARC**	**ACA**	**CSGN**	**GLE**	**BARC**
12	LLOY	ISP	**GLE**	**GLE**	**UCG**	**GLE**
13	**UCG**	**ACA**	NDA	**UCG**	NWG	**UCG**
14	**ACA**	GLE	**STAN**	**STAN**	**STAN**	**DBK**
15	**STAN**	NDA	**DBK**	**ACA**	**ACA**	**ACA**
16	NDA	**STAN**	DNB	**DBK**	**DBK**	**STAN**
17	**DBK**	**DBK**	**UCG**	NDA	NDA	NDA
18	SEB	CABK	SEB	SEB	DANSKE	DNB
19	DANSKE	DNB	DANSKE	CABK	DNB	SEB
20	SWEDA	SWEDA	ABN	KBC	CABK	CABK

Table 6.2: The table shows the top 20 rankings based on estimated mean $\Delta^{\$}CoVaR$ for each method. The time period 2002-12-31 to 2022-09-30 was used in the estimation. The bold banks are the ones included in the G-SIB list as of 2022.

Ticker	A&B	Lasso	PCA	5 random	7 random	9 random
HSBC	1	2	1	1	1	1
UBSG	6	3	4	5	5	5
BNP	2	4	2	3	2	3
SAN	3	1	3	2	3	2
ISP	5	12	7	9	9	9
INGA	4	5	5	4	4	4
BBVA	7	7	6	7	6	7
LLOY	12	6	9	6	7	6
NDA	16	15	13	17	17	17
UCG	13	10	17	13	12	13
ACA	14	13	11	15	15	15
NWG	11	8	23	8	13	10
CABK	22	18	22	19	20	20
KBC	23	21	25	20	24	22
DNB	21	19	16	23	19	18
BARC	10	11	8	10	10	11
SEB	18	21	18	18	21	19
STAN	15	16	14	14	14	16
DBK	17	17	15	16	16	14
SWEDA	20	20	24	22	23	23

Table 6.3: The table shows the ranking overview for the 20 largest banks (as of 2023-03-28) using time period 2002-12-31 to 2022-09-30 for calculation. The banks in bold are included in the G-SIB list as of 2022.

Observing table 6.2, there are twelve banks included in the G-SIB list and they are all in the top 20 ranking for each of the methods. For HSBC, SAN, BNP, INGA and UBSG, they are all ranked within the top five for all methods, except for UBSG that falls to a 6^{th} place ranking using Adrian & Brunnermeiers state variables. However, the mutual ranking within these five changes for each method, except for 5 random and 9 random, in which case they are the same. It is also noticeable that some banks fall in and out of the top 20 rankings, for example SEB which is included in the top 20 using A&B, PCA, 5 random and 9 random but not using Lasso or 7 random. Table 6.3 further gives a quick overview of how the ranking changes between methods for each of the top 20 banks. UCG, which is included in the G-SIB has the largest dispersion of ranking within the method, whereas it is ranked 10 using lasso and 18 using PCA. The bank with the most stable ranking is HSBC, which is ranked 1^{st} for all methods except Lasso, where it is ranked 2^{nd}.

Rank	**A&B**	**Lasso**	**PCA**	**5 random**	**7 random**	**9 random**
1	**HSBC**	**UBSG**	**HSBC**	**HSBC**	**HSBC**	**HSBC**
2	**BNP**	**HSBC**	**UBSG**	**SAN**	**UBSG**	**INGA**
3	**INGA**	**SAN**	**SAN**	**BNP**	**BNP**	**SAN**
4	**SAN**	NWG	**BNP**	**UBSG**	**SAN**	**UBSG**
5	**UBSG**	**INGA**	**INGA**	**INGA**	**INGA**	**CSGN**
6	GLE	**BNP**	**BBVA**	NWG	**CSGN**	**BNP**
7	**CSGN**	**CSGN**	**CSGN**	**CSGN**	**BBVA**	NWG
8	**UCG**	**BBVA**	**BARC**	GLE	**UCG**	**BBVA**
9	NWG	**UCG**	GLE	**BBVA**	NWG	**BARC**
10	**BARC**	LLOY	**ACA**	**BARC**	**BARC**	**UCG**
11	**BBVA**	**BARC**	LLOY	LLOY	GLE	GLE
12	ISP	GLE	**DBK**	**UCG**	LLOY	LLOY
13	**ACA**	**ACA**	ISP	**DBK**	**ACA**	**DBK**
14	LLOY	ISP	**UCG**	ISP	ISP	ISP
15	**DBK**	**DBK**	NWG	**ACA**	**DBK**	**ACA**
16	**STAN**	**NDA**	**STAN**	**STAN**	**STAN**	**STAN**
17	DANSKE	**STAN**	**NDA**	**NDA**	**NDA**	**NDA**
18	SWEDA	KBC	DANSKE	KBC	DANSKE	KBC
19	MB	MB	DNB	DANSKE	ETE	DANSKE
20	SEB	ETE	KBC	SEB	KBC	ETE

Table 6.4: The table shows the top 20 rankings based on estimated mean $\Delta^{\$}CoVaR$ for each method. In the estimation, time period 2005-01-01 to 2006-12-31 was used. The bold banks are the ones included in the G-SIB list as of 2011.

Ticker	A&B	Lasso	PCA	5 random	7 random	9 random
HSBC	1	2	1	1	1	1
UBSG	5	1	2	4	2	4
BNP	2	6	4	3	3	6
SAN	4	3	3	2	4	3
ISP	12	14	13	14	14	14
INGA	3	5	5	5	5	2
BBVA	11	8	6	9	7	8
LLOY	14	10	11	11	12	12
NDA	23	16	17	17	17	17
UCG	8	9	14	12	8	10
ACA	13	13	10	15	13	15
NWG	9	4	15	6	9	7
CABK	N/A	N/A	N/A	N/A	N/A	N/A
KBC	22	18	20	18	20	18
DNB	21	27	19	27	25	25
BARC	10	11	8	10	10	9
SEB	20	24	21	20	26	22
STAN	16	17	16	16	16	16
DBK	15	15	12	13	15	13
SWEDA	18	26	23	25	30	29

Table 6.5: The table shows the ranking overview for the 20 largest banks (as of 2023-03-28) using time period 2005-01-01 to 2006-12-31 for calculations. The banks in bold are included in the G-SIB list as of 2011.

The above two tables show the ranking using the years 2005 and 2006 as time period for calculation. As seen in table 6.4, the different methods yield widely different rankings, although most of the same banks stay in the top 20. HSBC is again ranked number one in all methods except for lasso where it is ranked second after USBG. Again, table 6.5 tell us that HSBC is the bank with the least dispersion between ranking for the different methods, where it is ranked either 1^{st} or 2^{nd} in all methods. Swedbank (SWEDA) further has the highest dispersion where it is ranked 18^{th} using Adrian & Brunnermeier state variables and 30^{th} using seven random state variables. Caixa Bank (CABK) did not exist in 2005-2006, and hence, it did not have a ranking.

Rank	**A&B**	**Lasso**	**PCA**	**5 random**	**7 random**	**9 random**
1	**HSBC**	**HSBC**	**HSBC**	**HSBC**	**HSBC**	**HSBC**
2	**BNP**	**SAN**	**SAN**	**SAN**	**BNP**	**SAN**
3	**SAN**	LLOY	**BNP**	**BNP**	**SAN**	**BNP**
4	**BBVA**	**UBSG**	**UBSG**	LLOY	LLOY	LLOY
5	**INGA**	**BNP**	**BBVA**	**BBVA**	**UBSG**	**BBVA**
6	ISP	**BBVA**	LLOY	**UBSG**	**BBVA**	**INGA**
7	**UBSG**	NWG	**INGA**	**STAN**	**INGA**	**CSGN**
8	LLOY	**CSGN**	**BARC**	**BARC**	**CSGN**	**UBSG**
9	**CSGN**	**INGA**	**STAN**	**INGA**	**BARC**	NWG
10	**GLE**	**BARC**	**CSGN**	NWG	**STAN**	**BARC**
11	**BARC**	**UCG**	**NDA**	**CSGN**	AIBG	ISP
12	**STAN**	AIBG	ISP	ISP	ISP	**STAN**
13	NWG	ISP	**GLE**	AIBG	NWG	AIBG
14	**UCG**	**STAN**	**DBK**	**GLE**	**UCG**	**GLE**
15	AIBG	**NDA**	DNB	**UCG**	**GLE**	**DBK**
16	SWEDA	**GLE**	SEB	**NDA**	**DBK**	**UCG**
17	**NDA**	**ACA**	**ACA**	**DBK**	**NDA**	**NDA**
18	**ACA**	**DBK**	SHBB	**ACA**	**ACA**	**ACA**
19	SEB	SWEDA	SWEDA	SEB	CABK	DNB
20	**DBK**	BKIA	AIBG	SWEDA	DANSKE	SEB

Table 6.6: The table shows the top 20 rankings based on estimated mean $\Delta^{\$}CoVaR$ for each method. In the estimation, time period 2013-01-01 to 2014-12-31 was used. The bold banks are the ones included in the G-SIB list as of 2014.

Tickers	A&B	Lasso	PCA	5 random	7 random	9 random
HSBC	1	1	1	1	1	1
UBSG	7	4	4	6	5	8
BNP	2	5	3	3	2	3
SAN	3	2	2	2	3	2
ISP	6	13	12	12	12	11
INGA	5	9	7	9	7	6
BBVA	4	6	5	5	6	5
LLOY	8	3	6	4	4	4
NDA	17	15	11	16	17	17
UCG	14	11	21	15	14	16
ACA	18	17	17	18	18	18
NWG	13	7	24	10	13	9
CABK	23	22	22	21	19	22
KBC	28	25	28	26	29	27
DNB	22	23	15	22	21	19
BARC	11	10	8	8	9	10
SEB	19	21	16	19	23	20
STAN	12	14	9	7	10	12
DBK	20	18	14	17	16	15
SWEDA	16	19	19	20	22	21

Table 6.7: The table shows the ranking overview for the 20 largest banks (as of 2023-03-28) where time period 2013-01-01 to 2014-12-31 was used in calculations. The banks in bold are included in the G-SIB list as of 2014.

Using the years 2013 and 2014 for calculations, tables 6.6 and 6.7 shows the top 20 rankings for each method as well as ranking for the 20 largest banks. Again, most banks included in the G-SIB list are ranked high in all methods. However, Lloyds Banking Group (LLOY) is ranked fairly high (3^{rd} on lasso and 4^{th} on all random selection) in this time period, which is not the case in any of the other time periods. There can be several reasons for this, one being that the bank had a relatively high market capitalization during this specific time period, or if risk in the institution increased more than its peers. Furthermore, table 6.7 shows that HSBC was ranked number one for all methods.

However, Banco Santander (SAN) also showed high consistency, only switching between number two and number three in the rankings. On the contrary, Natwest Group (NAT) showed high dispersion between methods where it was ranked seven using lasso and 24 using PCA.

Rank	**A&B**	**Lasso**	**PCA**	**5 random**	**7 random**	**9 random**
1	**BNP**	**SAN**	**HSBC**	**HSBC**	**HSBC**	**HSBC**
2	**HSBC**	**BNP**	**BNP**	**BNP**	**BNP**	**BNP**
3	ISP	**HSBC**	**SAN**	**SAN**	**INGA**	**SAN**
4	**INGA**	**UBSG**	**UBSG**	**INGA**	**SAN**	**INGA**
5	**SAN**	**INGA**	**INGA**	ISP	**UBSG**	**UBSG**
6	**UBSG**	LLOY	ISP	**UBSG**	ISP	ISP
7	**ACA**	ISP	DNB	LLOY	LLOY	LLOY
8	BBVA	BBVA	BBVA	BBVA	BBVA	BBVA
9	LLOY	**CSGN**	**ACA**	**BARC**	**CSGN**	**CSGN**
10	**BARC**	**ACA**	NDA	**ACA**	**ACA**	**BARC**
11	**CSGN**	NDA	**LLOY**	**CSGN**	**BARC**	**ACA**
12	SEB	**BARC**	SEB	NDA	NDA	NDA
13	DNB	**UCG**	**BARC**	SEB	DNB	DNB
14	GLE	NWG	**CSGN**	GLE	**UCG**	SEB
15	UCG	DNB	SHBB	**UCG**	SEB	GLE
16	NWG	SEB	SWEDA	NWG	**STAN**	**UCG**
17	SWEDA	KBC	GLE	DNB	GLE	NWG
18	NDA	SWEDA	**STAN**	**STAN**	**DBK**	**DBK**
19	**STAN**	GLE	CABK	KBC	NWG	KBC
20	KBC	CABK	**DBK**	**DBK**	KBC	SWEDA

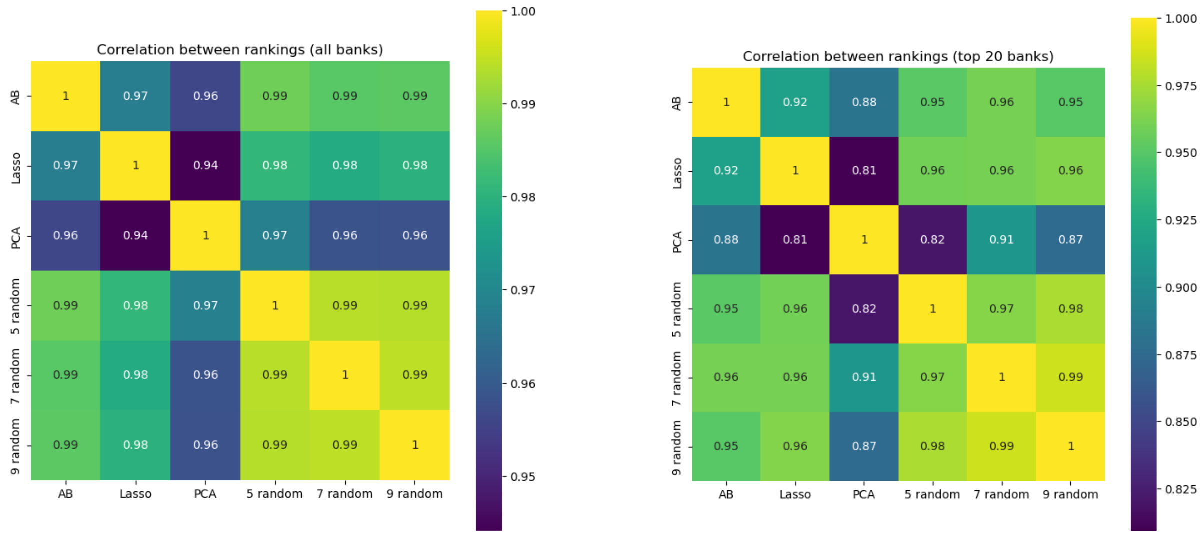
Table 6.8: The table shows the top 20 rankings based on estimated mean $\Delta^{\$}CoVaR$ for each method. In the estimation, time period 2020-01-01 to 2021-12-31 was used. The bold banks are the ones included in the G-SIB list as of 2021.

Rank	A&B	Lasso	PCA	5 random	7 random	9 random
HSBC	2	3	1	1	1	1
UBSG	6	4	4	6	5	5
BNP	1	2	2	2	2	2
SAN	5	1	3	3	4	3
ISP	3	7	6	5	6	6
INGA	4	5	5	4	3	4
BBVA	8	8	8	8	8	8
LLOY	9	6	11	7	7	7
NDA	18	11	10	12	12	12
UCG	15	13	23	15	14	16
ACA	7	10	9	10	10	11
NWG	16	14	28	16	19	17
CABK	22	20	19	22	22	22
KBC	20	17	21	19	20	19
DNB	13	15	7	17	13	13
BARC	10	12	13	9	11	10
SEB	12	16	12	13	15	14
STAN	19	21	18	18	16	21
DBK	24	24	20	20	18	18
SWEDA	17	18	16	21	21	20

Table 6.9: The table shows the ranking overview for the 20 largest banks (as of 2023-03-28) where time period 2020-01-01 to 2021-12-31 was used for calculations. The banks in bold are included in the G-SIB list as of 2021.

Above tables shows the third explicit period for calculation of the $\Delta^{\$}$ CoVaR and the rankings. HSBC is again in the top rankings, ranked 1^{st} for PCA and all random selection methods, however, it is ranked 3^{rd} using lasso and 2^{nd} using Adrian & Brunnermeier. BNP exhibits the most consistent ranking, being ranked either 1^{st} or 2^{nd}, which is also seen in table 6.9. Same as period two, Natwest Group (NWG) fluctuates the most between the different methods, being ranked 28^{th} using PCA and 14^{th} using lasso. Furthermore, Banco Santander (SAN), being one of the more systemically important banks, fluctuates relatively much where it is ranked 1^{st} using lasso and 5^{th} using A&B.

Looking at the rankings across all periods, there are some findings worth mentioning. Firstly, the same banks tend to be ranked high across all periods and across all methods, with HSBC, BNP Paribas (BNP) and Banco Santander (SAN) as examples. However, some banks tend to be ranked high during a certain period but low for a different period. An example of this is Intesa Sanpaolo (ISP), which is ranked between 12 and 14 for period one, but for period three, it is ranked between three and seven. As previously mentioned, this could be linked to a change in market capitalization, but it could also be linked to a change in risk.



(a) All banks (b) 20 largest banks

Figure 6.7: The two figures shows the rank correlation between the six different methods, measured using Spearman's rank correlation. The leftmost figure shows the rank correlation when using all banks in the dataset, while the rightmost figure shows the rank correlation using the 20 largest banks (as of 2023-03-28).

The two figures, 6.7a and 6.7b, shows the Spearman's rank correlation coefficient between the rankings both using all banks, and using the 20 largest banks (as of latest market capitalization). A common denominator between the two figures is that PCA exhibits the lowest correlation with the other methods. This could be linked to PCA not being a variable selection method but rather a method of shrinking the dimension, as mentioned in section 4.2.3. When all banks are included, the lowest correlation is between PCA and lasso with a correlation coefficient of 0.94. For the top 20 banks, the lowest correlation coefficient is again between PCA and lasso and amounts to 0.81. The highest correlation for all banks is between A&B and all random selection as well as the mutual correlation between the random selection methods. They all show a correlation of 0.99. As for the top

20 banks, the highest correlation is between 7 random and 9 random, with a coefficient of 0.99. Overall, the correlation between the rankings for the different methods are lower when only including the top 20 banks, than when all banks are included. Furthermore, a significance test was conducted with the conclusion that all correlation coefficients were significantly different from zero. However, in this setting, it is more appealing to test whether the rank correlation is significantly different from one. For the top twenty banks, it was not possible to reject the null hypothesis of a rank correlation equal to one. Additional testing was conducted on the larger sample of the rank correlations of all banks. The results implied that the null hypothesis was rejected for 8 out of the 36 rank correlations.

7. Discussion

The estimates presented in section 6.1 clearly indicates that the choice of state variable selection method matters. Different methods generate different estimates, and for some banks and across methods, these differences are highly noticeable in euro terms, especially when using the PCA to obtain $\Delta^{\$}$CoVaR estimates. This is important since different estimates introduce uncertainty regarding which estimates one should use when determining a bank's systemic risk contribution. The systemic risk contribution represents the bank's connection to the risk of the financial system as a whole, and from a regulator's point of view, a higher systemic risk contribution could imply regulatory consequences such as being placed in a higher bucket. Hence, if the CoVaR methodology is to be used for determining systemic risk contributions, the fact that different state variable selection methods yield different estimates should be taken into consideration. Which state variables to use, which estimates one should choose and hence consider the most accurate ones, or if one should even use state variables to capture time variation in systemic risk estimates is beyond the scope of this essay. However, the results presented in this essay should be seen as something that questions the expediency of using state variables within Adrian & Brunnermeier's CoVaR methodology to capture the time variation in systemic risk. One should note that these results also have implications for academic research within the area of systemic risk. Future research should put emphasis on the selection of state variables and this essay proposes that if one wishes to use state variables, some motivation may be needed.

Some similarities between the $\Delta^{\$}$CoVaR estimates were brought up in the empirical results section, at least in the mean estimates. The inconsistency between the mean estimate and the time series correlation however points to additional estimation uncertainty, which makes it difficult to draw conclusions. This is the case since similar mean estimates and a low time series correlation suggests that the method showcases a low consistency in the time series dimension while large differences in the mean estimates but a high time series correlation could propose that the methods generate estimates that tend to follow the same trend but differ in the relative estimate levels.

The estimation uncertainty is further highlighted if one considers the development of systemic risk contributions over time presented in section 6.1. During some time periods, different methods estimate different signs of the changes in systemic risk contributions, which implies that one method can mediate the picture that a bank had an increasing systemic risk contribution from one period to another while a different method may say the opposite. This is problematic since it makes it difficult to distinguish how different events and certain periods connect to a bank's risk contribution. As shown by the results presented, it is even possible that the sign of the estimate values differ where one method may produce a negative $\Delta^{\$}$CoVaR estimate. This further showcases the fluctuations in the estimates between methods.

As seen in section 6.2, the rankings between the banks tend to differ both over time, but also across methods. Since the rankings are based on $\Delta^{\$}$CoVaR, the size of the institution will matter, which is why e.g., HSBC tends to have a fairly stable ranking, while a smaller bank tends to fluctuate more. An example is Natwest Group in period two, which was ranked seven using lasso and 24 using PCA, which is seen as a large dispersion. Furthermore, HSBC is almost twice as large as the second-largest bank in terms of market capitalization. Hence, one could argue that even though HSBC only fluctuates between a ranking of first and third, this dispersion can still be considered significant because of the size of the bank. If the different selection methods would produce similar results, one could hence expect that the ranking of HSBC would not vary at all. As briefly mentioned above, the capital requirement of HSBC could be different depending on if it is ranked number one or number three looking from a regulator's point of view. One could even argue that the results of this essay may propose that the choice of state variable selection method could influence if a bank is defined as a G-SIB or not because of the variation in systemic risk contribution rankings and estimate values. Hence, the selection of state variables seems to be important for the results. Furthermore, as mentioned in section 6.2, some banks also tend to have a high fluctuation of ranking over time, which could be linked to either change in market capitalization or change in risk (measured as ΔCoVaR). Therefore it is important to also take the time period into account when estimating the systemic risk contributions of the financial institutions, at least in terms of rankings.

Looking at the correlation coefficients of the rankings, it is obvious that they tend to have a high correlation when including all banks. However, the hypothesis test of a rank correlation equal to one showed that 8 out of 36 rank correlations were significantly different from one. When looking at the top 20 banks, some correlations are just over 0.8 (e.g., PCA and lasso). In this setting, one could argue that this is a low coefficient of correlation since if the variable selection method would not matter, then the rankings would be very stable and the correlation should be close to one. This is the case for some of the methods (e.g., 7 random and 9 random), but not for most of the selection methods. Since it was not possible to reject a null hypothesis of correlation equal to one for any of the rank correlations using the top 20 banks, this would imply that rankings indeed is stable across methods. However, the results of the statistical test should be read with some caution due to the very small sample size (20). Hence, the validity of the result can be questioned because of the lower degrees of freedom resulting in fatter tails. Observing the deviation of rankings across all banks, the largest difference between highest and lowest ranking across methods for one bank is 52, and the average difference in ranking across all methods and all banks is approximately 14. For the 20 largest banks, the largest difference in ranking is 15 and average difference across all methods is approximately 4. Hence, even though most rank correlations were not statistically different from one, the changes in ranking can still be argued to be fairly large, which has negative implications for regulators and academic research wishing to rank banks by their systemic risk contributions.

8. Conclusions and future research

The main objective of this essay has been to assess the sensitivity of ΔCoVaR, expressed as $\Delta^{\$}$CoVaR, to the choice of state variable selection method when obtaining time-varying systemic risk estimates. More specifically, the essay has tried to answer the question if different state variables selection methods yield different estimates and systemic risk contribution rankings. The results presented suggests that the risk measure is indeed sensitive to the choice of state variable selection method, and that the risk measure generates different estimates of systemic risk contributions as well as variation in rankings depending on what state variable selection method is used in the estimation. This has obvious implications for regulators and academic research that wish to estimate systemic risk contributions over time, since the estimation uncertainty questions which estimates are the most accurate ones, which banks are most systemically important (ranking), and which state variables should be used if one wants to use them to obtain time-varying systemic risk estimates.

Since, as far as the authors are aware, this essay is the first one to study how the choice of state variables selection method affects systemic risk estimates using CoVaR, the need for future research is obvious. This essay does not contribute to the question regarding which state variables to use when conditionally estimating CoVaR. Hence, future research could investigate what state variables should be included in the estimation. This is however conditioned on that one wants to use state variables at all to capture time variation, which the results of this essay in some sense questions. Therefore, it would make sense for future research to study the expediency of additional estimation methods when estimating systemic risk contributions over time.

9. References

Abdi, Hervé and Lynne J Williams (2010). "Principal component analysis". In: *Wiley interdisciplinary reviews: computational statistics* 2.4, pp. 433–459.

Acharya, Viral and Matthew P. Richardson (2009). "CAUSES OF THE FINANCIAL CRISIS". In: *Critical Review* 21, pp. 195–210.

Adrian, Tobias and Markus K Brunnermeier (2011). *CoVaR*. Tech. rep. National Bureau of Economic Research.

Adrian, Tobias and Markus K. Brunnermeier (July 2016). "CoVaR". In: *American Economic Review* 106.7, pp. 1705–41. DOI: 10.1257/aer.20120555. URL: https://www.aeaweb.org/articles?id=10.1257/aer.20120555.

Al-Ameer, Maryam, Walaa Hammad, Areej Ismail, and Allam Hamdan (Sept. 2018). "The Relationship of Gold Price with the Stock Market: The Case of Frankfurt Stock Exchange". In: *International Journal of Energy Economics and Policy* 8, pp. 357–371.

Asgharian, Hossein, Dominika Krygier, and Anders Vilhelmsson (2022). "Systemic risk and centrality: The role of interactions". In: *European Financial Management* 28.5, pp. 1199–1226.

Basel Committee on Banking Supervision, BCBS (2018). *Global systemically important banks: revised assessment methodology and the higher loss absorbency requirement*. Tech. rep. Bank for International Settlements.

Bernal, Oscar, Jean-Yves Gnabo, and Grégory Guilmin (2014). "Assessing the contribution of banks, insurance and other financial services to systemic risk". In: *Journal of Banking & finance* 47, pp. 270–287.

Boeckx, Jef, Maarten Dossche, and Gert Peersman (2014). "Effectiveness and transmission of the ECB's balance sheet policies". In.

Bonferroni, Carlo E (1935). "Il calcolo delle assicurazioni su gruppi di teste". In: *Studi in onore del professore salvatore ortu carboni*, pp. 13–60.

Borri, Nicola, Marianna Caccavaio, Giorgio Di Giorgio, and Alberto Maria Sorrentino (2012). "Systemic risk in the European banking sector". In: *Available at SSRN 2112919*.

Boyd, John H, Jian Hu, and Ravi Jagannathan (2005). "The stock market's reaction to unemployment news: Why bad news is usually good for stocks". In: *The Journal of Finance* 60.2, pp. 649–672.

Brooks, Chris (2019). *Introductory econometrics for finance*. Cambridge university press.

Cameron, A Colin and Pravin K Trivedi (2005). *Microeconometrics: methods and applications*. Cambridge university press.

Castro, Carlos and Stijn Ferrari (2014). "Measuring and testing for the systemically important financial institutions". In: *Journal of Empirical Finance* 25, pp. 1–14.

ECB (2009). *FINANCIAL STABILITY REVIEW december 2009*. Tech. rep. European Central Bank.

Financial Stability Board, FSB (2022). *FSB publishes 2022 G-SIB list*. URL: `https://www.fsb.org/2022/11/fsb-publishes-2022-g-sib-list/` (visited on 05/02/2023).

Girardi, Giulio and A Tolga Ergün (2013). "Systemic risk measurement: Multivariate GARCH estimation of CoVaR". In: *Journal of Banking & Finance* 37.8, pp. 3169–3180.

Hanif, Hasan, Muhammad Naveed, and Mobeen Ur Rehman (2020). "Extending the forward systemic risk measure: Do sector level variables matter?" In: *Cogent Business & Management* 7.1, p. 1809944.

Hull, John C. (2018). *Risk managament and Financial Institutions*. 5th ed. Wiley.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani (2021). *An introduction to statistical learning*. 2nd ed. Springer.

Katircioglu, Salih Turan, Kamil Sertoglu, Mehmet Candemir, and Mehmet Mercan (2015). "Oil price movements and macroeconomic performance: Evidence from twenty-six OECD countries". In: *Renewable and Sustainable Energy Reviews* 44, pp. 257–270. ISSN: 1364-0321. DOI: `https://doi.org/10.1016/j.rser.2014.12.016`. URL: `https://www.sciencedirect.com/science/article/pii/S1364032114010685`.

Kaufman, George G. and Kenneth E. Scott (2003). "What Is Systemic Risk, and Do Bank Regulators Retard or Contribute to It?" In: *The Independent Review* 7, p. 371.

Koenker, Roger and Gilbert Bassett Jr (1978). "Regression quantiles". In: *Econometrica: journal of the Econometric Society*, pp. 33–50.

Krygier, Dominika (2014). "Measuring systemic risk in the Nordic countries An application of CoVaR". Master's Thesis, Supervisor: Professor Hans Byström, Lund University.

Le Cook, Benjamin and Willard G Manning (2013). "Thinking beyond the mean: a practical guide for using quantile regression methods for health services research". In: *Shanghai archives of psychiatry* 25.1, p. 55.

Liu, Shuming (2015). "Investor Sentiment and Stock Market Liquidity". In: *Journal of Behavioral Finance* 16.1, pp. 51–67. DOI: 10.1080/15427560.2015.1000334. eprint: https://doi.org/10.1080/15427560.2015.1000334. URL: https://doi.org/10.1080/15427560.2015.1000334.

López-Espinosa, Germán, Antonio Moreno, Antonio Rubia, and Laura Valderrama (2012). "Short-term wholesale funding and systemic risk: A global CoVaR approach". In: *Journal of Banking & Finance* 36.12, pp. 3150–3162.

Mendonça, Helder Ferreira de and Rafael Bernardo da Silva (2018). "Effect of banking and macroeconomic variables on systemic risk: An application of ΔCOVAR for an emerging economy". In: *The North american Journal of Economics and finance* 43, pp. 141–157.

Montagna, Mattia, Gabriele Torri, and Giovanni Covi (Dec. 2020). *On the origin of systemic risk*. Working Paper Series 2502. European Central Bank. URL: https://ideas.repec.org/p/ecb/ecbwps/20202502.html.

Morgan, JP (1996). *RiskMetrics Technical Document*. Tech. rep. Morgan Guaranty Trust Company of New York.

OECD (2023). *Consumer Confidence Index (CCI)*. URL: https://data.oecd.org/leadind/consumer-confidence-index-cci.htm.

Petrella, Lea, Alessandro G Laporta, and Luca Merlo (2019). "Cross-country assessment of systemic risk in the European stock market: evidence from a CoVaR analysis". In: *Social Indicators Research* 146.1-2, pp. 169–186.

Porter, Stephen R (2015). "Quantile regression: Analyzing changes in distributions instead of means". In: *Higher Education: Handbook of Theory and Research: Volume 30*, pp. 335–381.

Roengpitya, Rungporn and Phurichai Rungcharoenkitkul (2011). "Measuring systemic risk and financial linkages in the Thai banking system". In: *Systemic Risk, Basel III, Financial Stability and Regulation*.

Silva, Walmir, Walmir Silva, Herbert Kimura, and Vinicius Amorim Sobreiro (2017). "An analysis of the literature on systemic financial risk: A survey". In: *Journal of Financial Stability* 28, pp. 91–114.

Smaga, P (2014). *The Concept of Systemic Risk. Systemic Risk Centre Special Paper No 5, The London School of Economics and Political Science, August 2014*.

Thabet, Errefat Hamoud (2014). "Examining the long run relationship between the US money supply (M2) and The Canadian Stock Market". In: *International Journal of Economics and Finance* 6.10, pp. 1916–9728.

Tibshirani, Robert (1996). "Regression shrinkage and selection via the lasso". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 58.1, pp. 267–288.

Wen, Fenghua, Kaiyan Weng, and Wei-Xing Zhou (2020). "Measuring the contribution of Chinese financial institutions to systemic risk: an extended asymmetric CoVaR approach". In: *Risk Management* 22, pp. 310–337.

Xu, Qifa, Mengting Li, Cuixia Jiang, and Yaoyao He (2019). "Interconnectedness and systemic risk network of Chinese financial institutions: A LASSO-CoVaR approach". In: *Physica A: Statistical Mechanics and its Applications* 534, p. 122173.

Appendix A-I

A List of banks

The table shows all banks included in the data set with their respective ticker, weight (as percent of system) and market capitalization, gathered March 28^{th} 2023.

Bank	**Ticker**	**Weight**	**MC (€)**
HSBC	HSBC	11.1500%	122,255,818,752
UBS Group	UBSG	5.8385%	64,017,055,744
BNP Paribas	BNP	5.7604%	63,160,750,080
Banco Santander	SAN	4.9716%	54,512,082,944
Intesa Sanpaolo SpA	ISP	3.9572%	43,389,415,424
ING Groep	INGA	3.5904%	39,367,217,152
Banco Bilbao Vizcaya Argentaria SA	BBVA	3.4593%	37,929,431,040
Lloyds Banking Group PLC	LLOY	3.1969%	35,053,301,760
Nordea Bank Abp	NDA	3.0891%	33,871,216,640
UniCredit SpA	UCG	2.9294%	32,119,873,536
Credit Agricole	ACA	2.7567%	30,225,739,776
Natwest Group	NWG	2.5816%	28,306,833,408
CaixaBank	CABK	2.4071%	26,392,498,176
KBC Groep	KBC	2.3171%	25,405,618,176
DNB Bank	DNB	2.2717%	24,908,630,016
Barclays PLC	BARC	2.2239%	24,384,319,488
Skandinaviska Enskilda Banken	SEB	1.9248%	21,104,467,968
Standard Chartered PLC	STAN	1.7480%	19,166,105,600
Deutsche Bank AG	DBK	1.6507%	18,098,995,200
Swedbank	SWEDA	1.5803%	17,327,513,600
Societe Generale	GLE	1.4447%	15,840,896,000
Danske Bank	DANSKE	1.4436%	15,828,756,480
Svenska Handelsbanken	SHBB	1.3821%	15,153,892,352
Julius Baer Group	BAER	1.1964%	13,118,524,416
ABN AMRO Bank	ABN	1.1808%	12,946,753,536

Erste Groupe Bank	EBS	1.1132%	12,206,319,616
Commerzbank	CBK	1.0554%	11,571,784,704
Bank of Ireland	BIRG	0.9037%	9,909,084,160
AIB Group	AIBG	0.8807%	9,656,423,424
FinecoBank	FBK	0.7687%	8,428,805,632
Mediobanca	MB	0.7068%	7,750,324,224
Banque Cantonale Vaudoise	BCVN	0.6730%	7,378,747,392
OTP Bank Nyrt	OTP	0.6646%	7,286,659,072
PKO Bank Polski	PKO	0.6613%	7,250,469,888
Santander Bank Polska	SPL	0.5679%	6,227,135,488
Banca Mediolanum	BMED	0.5486%	6,015,197,184
Banco de Sabadell	SAB	0.4998%	5,480,663,552
Banco BPM	BAMI	0.4756%	5,215,256,576
Investec Group	INVP	0.4292%	4,705,804,288
Bank Polska Kasa Opieki	PEO	0.4154%	4,554,596,352
Bankinter SA	BKT	0.4137%	4,535,678,464
Eurobank Ergasias Services and Holdings	EUROB	0.4088%	4,482,498,048
ING Groep	ING	0.4081%	4,474,834,944
Graubundner Kantonalbank	GRKP	0.3956%	4,337,618,944
Raiffeisen Bank International	RBI	0.3930%	4,309,109,248
Jyske Bank	JYSK	0.3703%	4,059,709,184
National Bank of Greece	ETE	0.3646%	3,997,305,088
Oberbank	OBS	0.3478%	3,813,188,352
Zagrebacka Banka	ZABA	0.3476%	3,810,879,232
Ringkjoebing Landbobank	RILBA	0.3315%	3,634,752,512
Luzerner Kantonalbank	LUKN	0.3153%	3,456,806,656
Avanza Bank	AZA	0.3069%	3,365,056,768
BAWAG Group	BG	0.3059%	3,354,449,920
BPER Banca	BPER	0.2931%	3,213,980,672
Bper Banca	BPE	0.2918%	3,199,822,080
Vontobel Holding	VONN	0.2913%	3,194,292,992
Credit Suisse	CSGN	0.2862%	3,138,318,336
Banco Comercial Portugues	BCP	0.2711%	2,972,921,856
St Galler Kantonalbank	SGKN	0.2687%	2,945,785,344

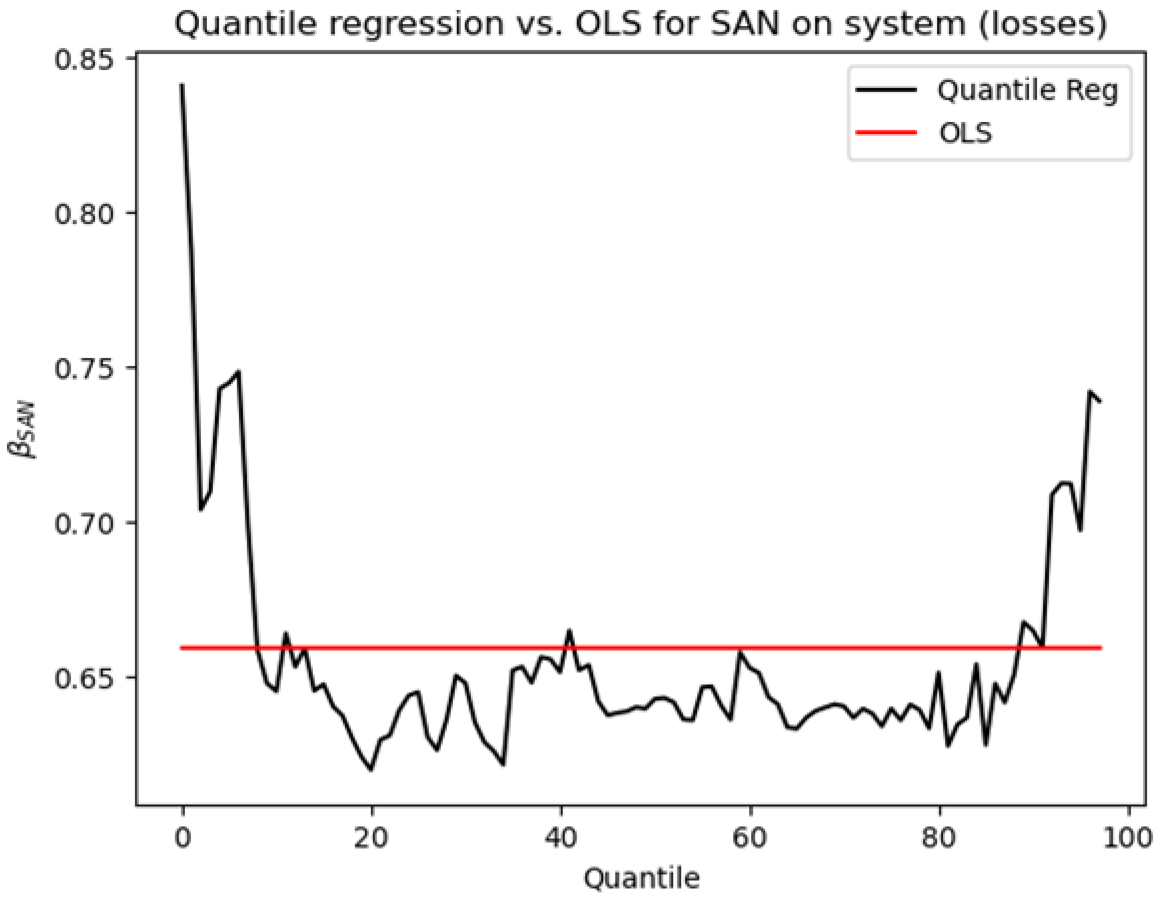
Basler Kantonalbank	BSKP	0.2584%	2,833,109,248
Banca Transilvania	TLV	0.2570%	2,817,675,776
Unicaja Banco	UNI	0.2353%	2,580,498,176
Alpha Services and Holdings	ALPHA	0.2307%	2,529,774,592
Banca Monte dei Paschi di Siena	BMPS	0.2217%	2,431,201,024
Thurgauer Kantonalbank	TKBP	0.2177%	2,387,203,328
Piraeus Financial Holdings	TPEIR	0.2154%	2,361,943,808
Sydbank	SYDB	0.2138%	2,344,060,928
Bank Handlowy w Warszawie	BHW	0.2085%	2,286,071,808
Zuger Kantonalbank	ZUGER	0.2036%	2,232,319,488
Berner Kantonalbank	BEKN	0.1952%	2,140,507,264
Virgin Money UK	VMUK	0.1943%	2,129,988,480
Caisse Regionale de Credit Agricole Mutuel de Paris et d'Ile-de-France	CAF	0.1853%	2,031,612,672
Basellandschaftliche Kantonalbank	BLKB	0.1794%	1,967,403,648
Liechtensteinische Landesbank	LLBN	0.1691%	1,854,237,696
Moneta Money Bank	MONET	0.1624%	1,780,529,280
Spar Nord Bank	SPNO	0.1594%	1,747,320,576
BNP Paribas Bank Polska	BNPPPL	0.1453%	1,593,497,472
Bank fuer Tirol und Vorarlberg	BTUV	0.1393%	1,527,006,208
Close Brothers Group	CBG	0.1349%	1,479,396,992
Banque Cantonale de Geneve	BCGE	0.1294%	1,418,928,000
Nova Ljubljanska Banka	NLBR	0.1284%	1,408,000,000
Arion Bank	ARION	0.1240%	1,359,271,040
Sparebank 1 SMN	MING	0.1234%	1,352,848,768
Permanent tsb	IL0A	0.1189%	1,303,958,016
Grenke	GLJ	0.1001%	1,097,295,488
Alior Bank	ALR	0.0939%	1,029,293,632
Caisse Regionale de Credit Agricole Mutuel du Languedoc	CRLA	0.0858%	940,226,560
Caisse Regionale de Credit Agricole Mutuel Nord de France	CNDF	0.0832%	912,420,800
FlatexDEGIRO	FTK	0.0730%	800,237,568
Aktia Bank Oyj	AKTIA	0.0616%	675,374,464

Collector Bank	COLL	0.0606%	664,709,312
Oma Saastopankki Oyj	OMASP	0.0575%	630,898,432
Alandsbanken Abp	ALBBV	0.0572%	627,012,544
Attica Bank	TATT	0.0538%	589,944,384
Caisse Regionale de Credit Agricole Atlantique Vendee	CRAV	0.0513%	562,964,352
Vestjysk Bank	VJBA	0.0492%	539,960,256
Bank of Valletta	BOV	0.0479%	525,464,352
Caisse Regionale de Credit Agricole Mutuel Sud Rhone Alpes	CRSU	0.0460%	503,934,624
Caisse Regionale de Credit Agricole Mutuel Alpes Provence	CRAP	0.0452%	495,509,344
Komercijalna Banka AD Skopje	KMB	0.0418%	458,836,096
Schweizerische Nationalbank	SNBN	0.0410%	449,307,136
Banco di Desio e della Brianza	BDB	0.0399%	438,023,552
Caisse Regionale de Credit Agricole Mutuel de Normandie-Seine	CCN	0.0395%	432,990,848
Credit Agricole Loire Haute-Loire	CRLO	0.0388%	425,001,216
Resurs Holding	RESURS	0.0366%	401,219,968
Caisse Regionale de Credit Agricole Mutuel de la Touraine et du Poitou	CRTO	0.0360%	395,045,216
ProCredit Holding	PCZ	0.0339%	372,238,464
NLB Banka AD Skopje	TNB	0.0329%	360,515,808
HSBC Bank Malta	HSB	0.0322%	353,099,968
Glarner Kantonalbank	GLKBN	0.0306%	335,776,832
Caisse Regionale de Credit Agricole Mutuel d'Ille-et-Vilaine	CIV	0.0292%	319,708,352
Hypothekarbank Lenzburg	HBLN	0.0282%	309,059,136
Lån & Spar Bank	LASP	0.0281%	307,688,576
Caisse Regionale de Credit Agricole du Morbihan	CMA	0.0270%	295,627,488
Addiko Bank	ADKO	0.0263%	288,600,000
Coop Pank	CPA1T	0.0258%	282,902,016
TF Bank	TFBANK	0.0230%	251,963,664
Hrvatska postanska banka	HPB	0.0211%	230,807,248

BankNordik	BNORDIK	0.0204%	223,622,800
Metro Bank PLC	MTRO	0.0179%	196,403,920
Bank Ochrony Srodowiska	BOS	0.0161%	176,705,840
Skjern Bank	SKJE	0.0145%	159,194,304
Secure Trust Bank	STB	0.0130%	143,029,328
Fynske Bank	FYNBK	0.0130%	143,010,240
Djurlands Bank	DJUR	0.0123%	134,856,224
Banca Sistema	BST	0.0098%	107,120,840
Kreditbanken	KRE	0.0086%	93,945,784
Lollands Bank	LOLB	0.0076%	83,236,264
Totalbanken	TOTA	0.0062%	68,417,688
Mons Bank	MNBA	0.0049%	54,241,668
Texim Bank	TXIM	0.0048%	52,150,956
Istarska Kreditna Banka Umag	IKBA	0.0045%	49,136,000
Wiener Privatbank	WPB	0.0030%	32,530,192
Fellow Bank	FELLOW	0.0027%	29,944,608
Hvidbjerg	HVID	0.0020%	22,445,926
Slatinska Banka	SNAB	0.0006%	6,111,164
Natixis SA	KN	0.0000%	
Banco Bilbao Vizcaya Argentaria	BVA	0.0000%	
Bankia	BKIA	0.0000%	
Liberbank	LBK	0.0000%	
NLB Skladi - Globalni uravnotezeni	NLBKOMB	0.0000%	

B OLS vs Quantile Regression

Difference between estimated β using ordinary least squares (OLS) and quantile regression. The regression run is losses of Santander on the system losses.



C All state variables

The table shows all state variables included in the data set with a description and a motivation to why it should be included.

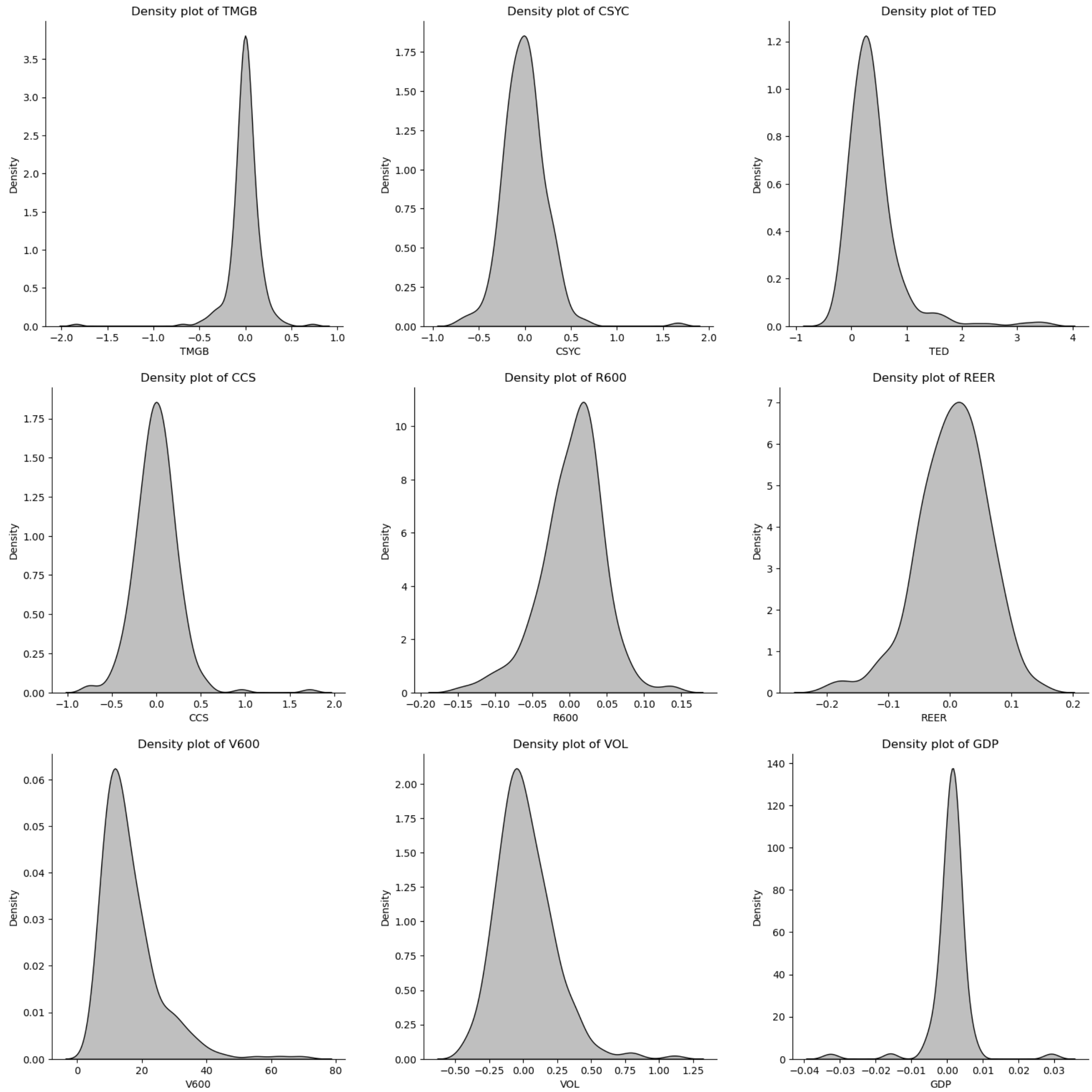
Symbol	**State variable description**	**Motivation**
TMGB	Change in three month german bond yield	Captures future economic activity and inflation as well as short term liquidity risk (Asgharian et al. 2022)
CSYC	Change in slope of german yield curve (calculated as 10y government bond - 3m treasury bill)	Captures time variation in return tails (ibid.)
TED	Spread between three month EURIBOR (Euro Interbank Offered Rate) and three month german govenment bond yield	Captures short-term liquidity risk (Adrian and Markus K. Brunnermeier 2016)
CCS	Change in credit spread (calculated as Moody's Baa-rated bond index - 10y german bond index)	Captures time variation in return tails (Asgharian et al. 2022)
R600	Return of Stoxx600 index	Controls for equity market return (ibid.)
REER	Real estate sector excess return over financial sector	Controls for equity market return (ibid.)
V600	Volatility of Stoxx600 (VIX-index of Stoxx600)	Captures uncertainty and investor sentiment (ibid.)
VOL	Percentage change in volume of Stoxx600	Captures investor sentiment (Liu 2015)
GDP	Percentage change in German GDP	Captures future economic activity (Roengpitya and Rungcharoenkitkul 2011)

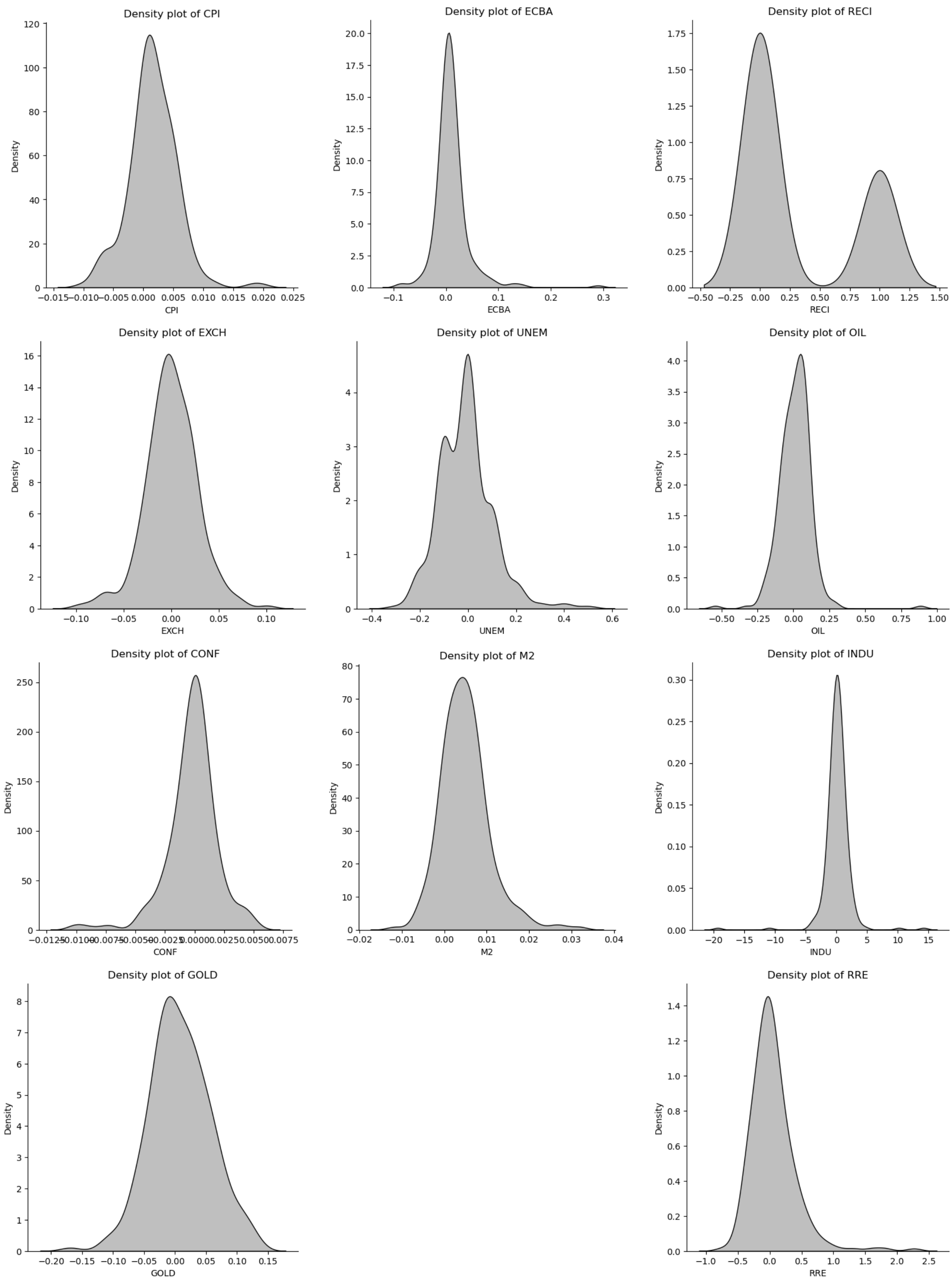
CPI	Percentage change in German CPI	Captures future economic actvity and controls for inflation (Roengpitya and Rungcharoenkitkul 2011)
ECBA	Percentage change in ECB assets	Captures future economic activity and return tails (Boeckx et al. 2014)
RECI	Recession indicator (dummy variable)	Captures the state of the economy
EXCH	Return of EUR/USD exchange rate	Captures time variation in return tails (Petrella et al. 2019)
UNEM	Change in European unemployment rate	Captures future economic activity and time variation in return tails (Boyd et al. 2005)
OIL	Percentage change in oil price	Captures future economic activity (Katircioglu et al. 2015)
CONF	Percentage change in OECD Europe consumer confidence index	Captures uncertainty, investor sentiment and future economic activity (OECD 2023)
M2	Percentage change in ECB money supply, M2	Captures time variation in return tails (Thabet 2014)
INDU	Change in Eurozone indutrial production	Captures future economic activity (Roengpitya and Rungcharoenkitkul 2011)
GOLD	Percentage change in gold price	Captures time variation in return tails (Al-Ameer et al. 2018)
RRE	Return of Stoxx600 Real Estate index	Controls for equity market return (same reasoning as for R600, (Asgharian et al. 2022))

RBANK	Return of Stoxx600 Bank index	Controls for equity market return (same reasoning as for R600, (Asgharian et al. 2022))
VBANK	Volatility of Stoxx600 Bank index (VIX-index)	Captures uncertainty and investor sentiment (same reasoning as for V600, (ibid.))
VOLB	Percentage change in volume of Stoxx600 Bank index	Captures investor sentiment (same reasoning as for VOL, (Liu 2015)
VRE	Volatility of Stoxx600 Real Estate index (VIX-index)	Captures uncertainty and investor sentiment (same reasoning as for V600, (Asgharian et al. 2022))

D Distribution of state variables

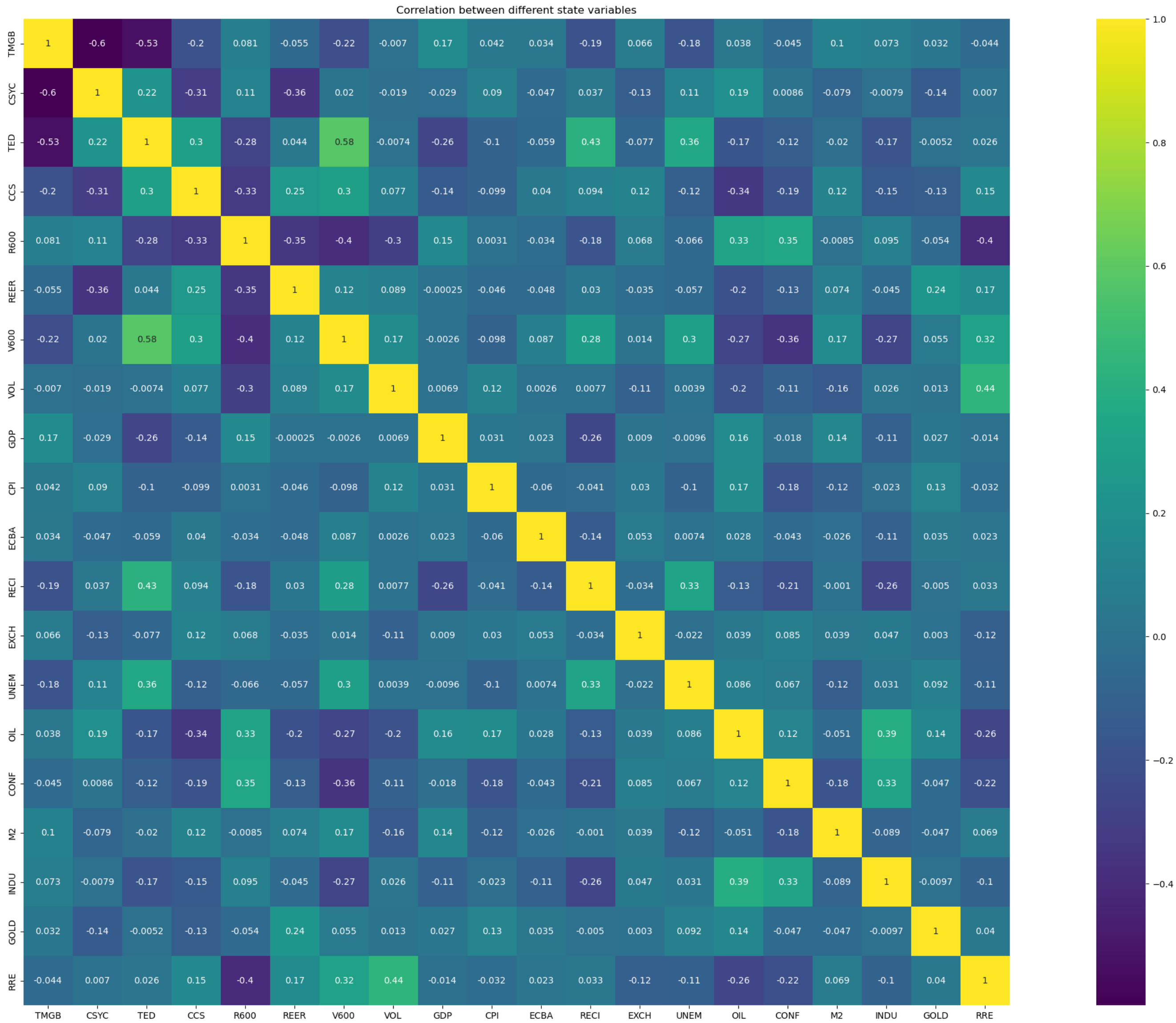
The plots below show the density distribution of all state variables included in the data set.




Density plot of CPI
Density
CPI
Density plot of ECBA
Density
ECBA
Density plot of RECI
Density
RECI
Density plot of EXCH
Density
EXCH
Density plot of UNEM
Density
UNEM
Density plot of OIL
Density
OIL
Density plot of CONF
Density
CONF
Density plot of M2
Density
M2
Density plot of INDU
Density
INDU
Density plot of GOLD
Density
GOLD
Density plot of RRE
Density
RRE

E Correlation heatmap

The figure shows the correlation between each of the state variables in a heatmap (after the four variables with a correlation above 0.70 were removed).



F Selected state variables

The chosen state variables in each of the estimation methods are shown below.

Method	Selected state variables
Adrian & Brunnermeier	TMGB, CYSC, TED, CCS, R600, REER, V600
Lasso	CSYC, EXCH, CONF
PCA	N/A
Random 5 (draw 1)	GOLD, CSYC, RRE , GDP, ECBA
Random 5 (draw 2)	CCS, M2, V600, ECBA, TED
Random 5 (draw 3)	EXCH, R600, GOLD, TMGB, CPI
Random 5 (draw 4)	OIL, TED, CSYC, INDU, R600
Random 5 (draw 5)	RRE, CCS, GOLD, V600, UNEM
Random 7 (draw 1)	TED, REER, INDU, RRE, EXCH, CSYC, RECI
Random 7 (draw 2)	INDU, VOL, REER, V600, GDP, TED, CONF
Random 7 (draw 3)	CSYC, INDU, TED, REER, RECI, TMGB, GOLD
Random 7 (draw 4)	VOL, RECI, V600, CONF, GOLD, R600, EXCH
Random 7 (draw 5)	CCS, V600, VOL, REER, M2, CONF, TED
Random 9 (draw 1)	VOL, ECBA, REER, V600, CCS, GOLD, UNEM, TED, OIL
Random 9 (draw 2)	CCS, REER, ECBA, V600, OIL, RRE, CONF, CPI, GOLD
Random 9 (draw 3)	RRE, UNEM, CPI, VOL, OIL, M2, ECBA, GDP, REER
Random 9 (draw 4)	RECI, EXCH, TMGB, CSYC, GDP, REER, VOL, CONF, OIL
Random 9 (draw 5)	RRE, REER, TED, CSYC, R600, OIL, CCS, M2, INDU

G G-SIBs

The G-SIB list as of November 2022. The banks are sorted in alphabetical order in their respective bucket. The percentage in parenthesis is the additional loss absorbing capital imposed on the bank, as a percentage of risk-weighted assets.

Bank	**Area**	**Bucket**
JP Morgan Chase	United States	4 (2.5%)
Bank of America	United States	3 (2.0%)
Citigroup	United States	
HSBC	Europe	
Bank of China	Asia	2 (1.5%)
Barclays	Europe	
BNP Paribas	Europe	
Deutsche Bank	Europe	
Goldman Sachs	United States	
Industrial and Commercial Bank of China	Asia	
Mitsubishi UFJ FG	Asia	
Agricultural Bank of China	Asia	1 (1.0%)
China Construction Bank	Asia	
Credit Suisse	Europe	
Groupe BPCE	Europe	
Groupe Crédit Agricole	Europe	
ING	Europe	
Mizuho FG	Asia	
Morgan Stanley	United States	
Royal Bank of Canada	North America	
Santander	Europe	
Société Générale	Europe	
Standard Chartered	Europe	
State Street	United States	
Sumitomo Mitsui FG	Asia	

Toronto Dominion	North America
UBS	Europe
UniCredit	Europe
Wells Fargo	United states

H Standard deviation heatmap

A heatmap over the standard deviations for each bank across each method (for the 20 largest banks as of March 2023).