Modelling the Dynamic Effects of Oil Shocks on Sovereign Bonds: Empirical Evidence from a VAR Model

Heiko Julian Lehrer

Dissertation written under the supervision of professor

Dr. Eva Schliephake

Dissertation submitted in partial fulfilment of requirements for the MSc in Finance, at the Universidade Católica Portuguesa, 11th of September 2025.

Abstract

There is a significant research gap regarding how oil shocks affect the yield curve of oil-importing countries. Although sovereign bonds and their term structure are the backbone of modern finance, asset management and asset pricing, their resilience to oil shocks is largely underexplored. There is evidence that oil shocks might strongly impact the yield curve. Complementary, this thesis expands the analysis to several oil-importing countries as well as to different kinds of oil shocks by investigating the term-structure dynamics as a response to oil shocks and by comparing the modelled impulse responses to real-world shock responses to gain further understanding whether the appearance of shocks can be utilised to execute macro trades or hedges. Accordingly, country-specific VAR models are defined to capture the impulse response dynamics on the Nelson-Siegel modelled sovereign bond yields for oil shocks. This study underlines that oil shocks are a driver of the yield curve, although their effects are promptly superseded by other external factors. The ability to capture such dynamics varies by country, type of shock, and timing, which precludes the robust approximation of the real yield curve movement. As a result, the usage of these specific models within a practical trading or hedging environment proves to be limited.

Keywords: *oil shocks, sovereign bonds, VAR, vector autoregressive model, impulse response analysis*

Title: *Modelling the Dynamic Effects of Oil Shocks on Sovereign Bonds: Empirical Evidence from a VAR Model*

Author: *Heiko Julian Lehrer*

Resumo

Existe uma lacuna relevante de pesquisa sobre os efeitos dos choques do petróleo na curva de rendimento de países importadores. Embora os títulos soberanos e sua estrutura a termo sejam fundamentais para as finanças modernas, gestão e precificação de ativos, sua resiliência a choques do petróleo continua pouco explorada. Evidências indicam que tais choques podem afetar de forma significativa a curva de rendimento. Esta tese amplia a análise para diferentes países importadores e para distintos tipos de choques, examinando a dinâmica da estrutura a termo em resposta a esses eventos e comparando as respostas de impulso modeladas com as observadas no mundo real. O objetivo é avaliar se tais choques podem ser utilizados na execução de operações macro ou estratégias de hedge. Para isso, são empregados modelos VAR específicos por país, aplicados a rendimentos soberanos modelados segundo Nelson-Siegel. Os resultados mostram que choques do petróleo influenciam a curva de rendimento, mas seus efeitos são rapidamente superados por outros fatores externos. A capacidade de capturar essas dinâmicas varia conforme o país, o tipo de choque e o momento, o que limita a aproximação robusta dos movimentos reais da curva. Assim, a utilização prática desses modelos em ambientes de negociação ou de hedge revela-se restrita.

Palavras-chave: *choques do petróleo, títulos soberanos, VAR, modelo vetorial autorregressivo, análise de resposta a impulso*

Título: *Modelando os Efeitos Dinâmicos dos Choques do Petróleo sobre os Títulos Soberanos: Evidência Empírica a partir de um Modelo VAR*

Autor: *Heiko Julian Lehrer*

Contents

Glossary of Abbreviations

ACF	Autocorrelation Function
ADF	Augmented Dickey-Fuller Test
API	American Petroleum Institute Degree
Brent	Brent Crude
CPI	Consumer Price Index
EIA	U.S. Energy Information Administration
FEVD	Forecast Error Variance Decomposition
GLM	Generalised Linear Models
i.i.d.	Independent and Identically Distributed
IRF	Impulse Response Function
OLS	Ordinary Least Squares
PACF	Partial Autocorrelation Function
RAC	Refiner Acquisition Cost of Crude Oil
SSE	Sum of Squared Errors
SVAR	Structural Vector Autoregressive Regression
US	United States
VAR	Vector Autoregressive Regression
VECM	Vector Error Correction Model
WIP	World Industrial Production Index
WTI	West Texas Intermediate

List of Figures

List of Tables

1 Introduction

According to the UN Trade and Development Organisation UNCTAD (2025), global public debt first exceeded the 100 trillion USD benchmark in 2024. This mounting debt pile alongside a long-term trend of increasing Debt-to-GDP ratios in conjunction with the unified voices of developed and developing countries regarding debt- and financing-related concerns addressing the United Nations General Assembly raises questions about the resilience of the public debt market in general and its reaction to global shocks (UNCTAD, 2025; IMF, 2024).

The vast majority of public debt is financed via sovereign bonds. For most of the developed countries, these are considered safe havens in terms of their risk profile (United States Department of the Treasury, 2025; Eurostat, 2024). Therefore, sovereign bonds cannot only be seen as the backbone of sovereign financing, but rather of the financial markets in general. Thus, this only underlines the importance of studying the resilience of government bonds, as their value impacts institutional investors such as central banks, domestic and non-domestic banks, hedge funds, insurances and pension funds as well as household investors, who are either directly affected as active investors, or indirectly by being dependent on yield, insurance or pension plan payments (Fang, Hardy, & Lewis, 2023).

A myriad of shocks may impact the yield curve of a country, but due to recent events and concerns regarding energy security embedded in an evolving and dynamic, multifaceted macroeconomic environment, this thesis aspires to investigate the complex interrelations of the sovereign bond market and the energy market. As stated by the International Energy Agency (2025), the share of oil within the global energy demand accounts for approximately 30%, which positions oil second among global energy sources, exceeded only by coal with 35%. Although being second in the global energy mix, I chose oil as a representative due to its known interconnectedness with the world economy and its proven track record of severely impacting macroeconomics (Kilian, 2008, 2009; Casoli, Manera, & Valenti, 2024).

Even though the world community demonstrates ambitious objectives to increase the portion of renewable energy sources within the energy composition, the U.S. Energy Information Administration (2023) does not predict a decrease of liquid fuels until the year 2050. This can be attributed to the forecasted overall increase in energy demand and consumption, which is mostly driven by GDP growth and continuous population growth. This is further backed by a demographic analysis of the United Nations (2024), predicting a world population of above 9.5 billion in 2050 and above 10 billion by 2100.

This growth in population is particularly seen in regions where liquid fuels dominate the energy generation (United Nations, 2024; Energy Institute, 2025). Therefore and despite the increase of renewable energies, I expect oil to be highly relevant to satisfy the hunger for energy of the global economy in the future. In consequence, I am investigating the dynamic effects of oil shocks on sovereign bonds and their practical implications for trading and hedging scenarios.

Finally, the structure of this thesis is as follows: Section 2 provides a comprehensive overview of the existing literature with respect to oil market shocks in general and their effects on sovereign bonds. The particular focus lies on the existing scientific gap that this thesis attempts to bridge. Section 3 gives an overview of the data and the variables used throughout the analysis and provides a logical reasoning why this data was chosen and how it is enriching the research. The methodology is discussed in Section 4. Similar to Section 3, this segment offers a broad overview of the methodology used and lays the groundwork for the empirical analysis. Section 5 forms the heart of this thesis by presenting, discussing and interpreting the empirical results gained throughout the creation of this paper. In Section 6, the limitations of the empirical analysis are discussed. Finally, Section 7 summarises the main findings of the analysis, highlights the insights gained, discusses its limitations, and suggests directions for future research.

2 Literature Review

This literature review aims to synthesise historical findings related to oil shocks and state-of-the-art methodologies to provide key insights into the interactions between oil and sovereign bond markets. Historical evidence reaching back to 1983 is supplemented with empirical modelling approaches from 2024 and embedded within a 40-year-ongoing framework of oil shock modelling research. The review starts with the creation and the continuous development of oil shock variables that were consolidated to a robust oil shock modelling framework, which then links oil as a commodity to the broader macroeconomic environment. The oil variables themselves and the macroeconomic variables influenced by them are further reflected in country-specific stock market movements, which build the foundation for researching whether similar movements in the sovereign bond markets can be validated.

Kilian (2009) set the groundwork of modern oil shock modelling by critically examining the standard procedure of testing oil shocks as a consequence of an oil price shock. Although the price might be considered as the synthesis of all variables impacting oil, this assumption does not invalidate concerns of reverse causality. Additionally, due to the interconnectedness of the oil market with the overall economy, using oil as a general proxy breaches with the premise of ceteris paribus. To avoid these sources of error, Kilian (2009) extrapolated three different shocks: oil supply shocks that are driven by regional disruptions, aggregate demand shocks on the basis of global economic activity, and oil-specific demand shocks, which result in speculative trading. These results and extracted shocks are further refined by Baumeister and Hamilton (2019), who observed the differences between supply and demand shocks with a focus on their impact on global economic activity.

The pioneering research by Hamilton (1983, 2003) about the linkage between oil price shocks and macroeconomic measurements is the interaction node across oil and other markets besides the commodity market. In his research, he examined the relationship between oil price shocks and historical US recessions with the outcome that a rise in oil prices may be associated with a downturn of the overall economy. However, oil price movements are not exclusively reflected within the probability of US recessions, but also in inflation expectations as demonstrated by Aastveit, Bjørnland, and Cross (2023). Aastveit et al. (2023) find that oil demand and supply shocks have a good explanatory power of expected and realised inflation, with the constraint of a different degree of impact based on the duration of the shock. This research by Aastveit et al. (2023) is the basis for Casoli et al. (2024), who studied the relationship of inflation dynamics and energy shocks in Europe after the COVID-19 pandemic, where energy shocks are attributed to either oil shocks or natural gas shocks. However, they show that in the post-pandemic period inflation is mostly driven by natural gas supply shocks rather than oil-specific shocks. The shock framework presented by Casoli et al. (2024) for modelling oil shocks is used throughout this thesis.

These findings lead to the cross-market dependencies of commodities – specifically oil and the equity markets. Jones and Kaul (1996) researched if and how stock returns are impacted by oil price shocks and how those are reflected in the company fundamentals. They show that the impact on the company-specific real cashflow fully justifies the reaction of the stock price. Nevertheless, this finding is a country-specific phenomenon, as this statement has been proven for Canada and the United States, but it is not tenable for Japan and the United Kingdom.

This research is expanded by Sadorsky (1999), who highlights that oil prices have a higher explanatory power of the error variance of equities than interest rates, particularly after accounting for asymmetric effects caused by the oil price volatility. Moreover, Kilian and Park (2009) use a Structural Vector Autoregressive Regression (SVAR) model to gain knowledge about the equity markets' reaction to other types of shocks rather than the price of oil. Thus, they find differences in the markets' reaction based on the reason that caused the shock and in conjunction with industry specifics.

In contrast to the extensive research about the interdependence of oil prices, oil shocks, and equity markets, the research conducted on these linkages to the sovereign bond markets is rather sparse. Kang, Ratti, and Yoon (2014) researched how oil price shocks influence the bond market returns of the United States using a SVAR. They use the shocks promoted by Kilian (2009) to model the return of each sovereign bond tenor separately and analyse the explanatory power across these shocks using a Forecast Error Variance Decomposition (FEVD). Kang et al. (2014) find that a positive demand shock is reflected in a decrease of real returns over the medium term. The same behaviour is observed for an aggregate demand shock, but with a longer shock period of 24 months. They argue that oil shocks explain 27.1% of the variation of the forecasting error variance. Complementary, Ioannidisa and Ka (2018) have a slightly different study design, as they do not test the shocks on separate bond tenors but rather on the term structure. Thus, they extract the short-term and long-term interest rates and the curvature of the yield curve from the plain government bond yields. These are then fed into a SVAR and tested for different shock scenarios as discussed by Kilian (2009).

Ioannidisa and Ka (2018) observe that aggregate demand shocks increase the long-term interest rates for all countries included in the study, whereas market-specific demand shocks have a positive impact on the long-term interest rates solely for oil-importing countries.

Additionally, they observe that oil shocks exhibit a significant degree of explanatory power over movements in the yield curve.

Considering the comprehensive outline of the literature discussing and observing the impact of oil shocks on macroeconomic variables, equity markets and sovereign bond markets, this thesis intends to provide a broad and in-depth overview of how the dynamic effects of oil shocks impact sovereign bonds and their respective term structure. The shock framework used in this thesis is based on the paper by Casoli et al. (2024), which yields more granular insights into the impulse responses compared to the one by Kilian (2009), because the oil-specific demand shocks are further split and hence refined into a consumption demand shock and an inventory demand shock. As initially conceived in Ioannidisa and Ka (2018), I subsequently apply these shocks on the constructed yield curve of several selected countries. However, in contrast to Ioannidisa and Ka (2018), the emphasis of this thesis is not only on theoretically assessing the response of the term structure to an artificial impulse of an oil variable and to further examine the existence of explanatory power within those shocks, but rather on determining whether the chosen variables capture the behaviour of the underlying system to reflect the real-world dynamics. Hence, similar to the paper of Ioannidisa and Ka (2018), the first hypothesis I tested raises the question, if we can observe general or solely country-specific reactions to shocks. The second hypothesis, which expands on the analysis of Ioannidisa and Ka (2018), should examine if there is sufficient evidence to use these patterns in shock responses within a trading or hedging environment. Therefore, I am testing the impulse responses against real-world shock data rather than decomposing the responses using FEVD, which would be the standard approach to test Vector Autoregressive Regression (VAR) models.

3 Data

The aim of this section is to discuss why I chose these variables, how I calculated them and which data transformations I applied, and to present the statistical characteristics of the data used.

3.1 Variable Selection & Description

The data used in this analysis is split into two groups: Firstly, data intended for the simulation of the oil shocks and secondly, the sovereign bonds data. To guarantee comparability and consistency of the analysis, the oil data describes different effects of oil shocks on a global level, which means that the data itself is equal for all countries and sovereign bonds although the effect of the specific oil data or oil shock might have a different effect on or might be reflected differently in the bond yields.

The selection of the oil data to simulate different oil shocks is based on the paper by Casoli et al. (2024), which is defining these based on the fundamental research on oil shock modelling in Kilian (2009), Baumeister and Hamilton (2019) and Aastveit et al. (2023). I aim to study the effects of four different types of shocks on sovereign bonds, three of which are directly linked to the oil market and one that can be seen as an additional parameter to reflect the global economy. Each of these shocks is represented as a dynamic interaction between these variables: *Global Oil Real Price* ($GORP$) as a proxy for consumption demand, *Global Oil Production* (GOP) as a proxy for oil supply, *Global Oil Inventories* (GOI) as a proxy for inventory demand and *Global Economic Activity* (GEA) as a substitute for aggregate demand.

Although the West Texas Intermediate (WTI) and Brent Crude (Brent) indices are often used to account for the price of oil, the fact that both are special grades of Crude oil makes them insufficient and unreliable for my model. WTI and Brent are both oil grades with a high American Petroleum Institute Degree (API) and a low sulphur content and are therefore considered to be of higher quality, which is reflected in the price, compared to Urals or Dubai oil. Hence, using them as a proxy for the *Global Oil Real Price* would introduce a bias of oil-quality and -price in conjunction with local geographical uncertainties, as WTI is the benchmark for US oil and Brent for UK oil (Roncoroni, Fusai, & Cummins, 2015). Therefore, I chose a different methodology based on the Refiner Acquisition Cost of Crude Oil (RAC), which is published on a monthly basis by the U.S. Energy Information Administration (EIA). The RAC is defined as the imported cost of Crude Oil and hence incorporates not only US-market-specific shocks but rather local, foreign and global shocks to the price of oil. Nevertheless, this variable captures the United States' perspective and therefore, the variable must be adapted to the US Consumer Price Index (CPI) to avoid any unnecessary bias, resulting in the definition $GORP = \frac{RAC}{CPI}$ (Casoli et al., 2024).

The *Global Oil Production* published by the EIA is the most straightforward variable, as it already reflects all necessary attributes on a global scale and does not need any correction terms. Furthermore, it is an integral part of the *Global Oil Inventories*. Unfortunately, there is no direct measure of the oil inventories globally, and therefore, a proxy variable must be introduced based on the methodology presented in Kilian and Murphy (2013). They developed a proxy for the *Global Oil Inventories* by applying a scaling ratio of the OECD Petroleum inventories and the United States (US) Petroleum inventories on the US Oil inventories, resulting in the following formula (Kilian & Murphy, 2013):

$$GOI = US\ crude\ oil\ inventories \cdot \frac{OECD\ Petroleum\ inventories}{US\ Petroleum\ inventories} \tag{1}$$

The last oil shock specific variable used in this thesis is the *Global Economic Activity*, which is represented by the World Industrial Production Index (WIP) as proposed by Baumeister and Hamilton (2019). This indicator was specifically developed to depict the world economy within an oil modelling context (Baumeister & Hamilton, 2019). The summary statistics can be found in Table 5 in the Appendix.

The second portion of the data consists of sovereign bonds data, more specifically, the yields of the bonds. The bonds data is from Refinitiv Workspace and represented by the Refinitiv Benchmark Indices for different countries and tenors/maturities. Since the aim of this thesis is to study the effects of oil market shocks on sovereign bonds and their yield curve, I collected the most comprehensive daily sovereign bonds data across as many tenors as possible to gain highly granular insights into the yield curve. However, data availability is highly dependent on the country, as Refinitiv Workspace only offers limited Benchmarking Indices for some of them. The maximum of available benchmarks are 3 months, 2 years, 3 years, 5 years, 7 years, 10 years and 30 years benchmarks. The availability of these time series in combination with sufficient historical data coverage and the relevancy for the analysis form the inclusion criteria for each country.

The data points queried from Refinitiv Workspace are the Bid Yield (BID), the Ask Yield (ASK) and the Mid Yield (YLD). Given that the analysis is to be buyer-seller neutral, the Mid Yield serves as an input into the VAR model and the Nelson-Siegel model. However, this data point is often not available and thus must be generated artificially. Firstly, the long-term Bid-Ask Spread ($SPRD$) is calculated based on $SP\bar{R}D = \frac{1}{n}\sum_{i=1}^{N} BID - ASK$, where n is the total number of observations in the time series. Secondly, the following methodology is used to construct the Mid Yield when it is not available:

$$YLD = \begin{cases} YLD, & \text{if Mid is available} \\ \frac{ASK+BID}{2}, & \text{if both Ask and Bid are available} \\ ASK + \frac{SPRD}{2}, & \text{if only Ask is available} \\ BID - \frac{SPRD}{2}, & \text{if only Bid is available} \\ undefined, & \text{if no valid data is available} \end{cases} \tag{2}$$

It is important to highlight two limitations of this construction methodology. The first limitation is that a forward-looking Bid-Ask Spread is being incorporated into the Mid Yield. Hence, the assumption for the Bid-Ask Spread is an averaging to the mean over a long period. I could bypass this limitation by applying an expanding Bid-Ask Spread, however, due to the already limited coverage of historical data and the otherwise time-variable averaging out of shocks, I decided to use the static Bid-Ask Spread. The second limitation is the averaging of shocks in the Bid-Ask Spread, which limits the interpretability, thus I decided to focus on the static version to keep the analysis buyer/seller neutral.

After discussing oil and sovereign bond variable specific methods and reasonings, there are a few remaining issues worth mentioning. Since this thesis studies the effects of shocks, outliers in the data were intentionally retained and not excluded. This, alongside a linear approximation for missing values, prepares the data optimally for further analysing shock scenarios.

3.2 Shock Events

This subsection presents the shock events I chose for the analysis and verification of the quality of the VAR model. To avoid any bias within the analysis, I selected events solely based on quantitative rather than qualitative factors. It is rather the intensity of the shock that is decisive than the rationale behind the shock's occurrence. To give a more holistic view without unnecessarily prolonging the analysis, I evaluated two scenarios per variable throughout the empirical analysis of this thesis.

The methodology used to select the event dates is ranking the z-scores in a descending order and picks the first two dates including some exceptions. The first exception being for the variable $GORP$ for August 2024, due to a lack of data, and the second exception referring to GOP for May 2020, because this date is already analysed for $GORP$. Additionally, the constraint is set that the event date must occur within the timeframe of January 2020 and January 2024, so I have sufficient historical data to train the VAR model and I have ample data after the shock to assess the model's output against the real data. Table 1 outlines the complete timeline of events.

	Rank	Shock Date	Z-Score*
GORP	1	May 2020	1.154
	2	March 2023	1.154
GOI	1	November 2020	1.925
	2	June 2022	1.902
GOP	1	February 2021	1.769
	2	January 2024	1.616
GEA	1	January 2020	4.477
	2	April 2022	1.604

***** *z-score is calculated on the transformed variables*

Table 1: *Overview of the shock events*

The z-score of each oil variable is computed on a rolling basis:

$$Z_t = \frac{x_t - \mu_{t,n}}{\sigma_{t,n}} \tag{3}$$

, where x_t is the data point at time t, $\mu_{t,n}$ describes the average of the data over the historical timeframe of n at time t, and $\sigma_{t,n}$ represents the standard deviation at time t with respect to the lookback period of n. Accordingly, I selected various lookback windows to capture the cyclical behaviour of the variable, while mitigating the influence of noise. Hence, $GORP$ has a relatively short cycle being captured with a lookback window of 3 months, whereas GOI has a lookback period of 6 months. Additionally, GOP uses an n of 9 months and GEA is depicted in a window of 24 months. These lookback periods are backed by academic literature from Hamilton (2003), Kilian and Park (2009), Baxter and King (1999) and Christiano and Fitzgerald (2003). As a result of limited historical coverage concerning oil data and sovereign bonds data, the shorter end of the spectrum for each variable is used in this analysis.

3.3 Country Selection

After I decided which events to analyse, the next integral part was the selection of the target countries. Sovereign bond yields are influenced by a range of macroeconomic factors, including inflation, unemployment, and interest rates. These variables are, in turn, shaped – directly or indirectly – by trade surpluses, trade imbalances and capital flows, both of which are closely linked to patterns of imports and exports (Cuñat & Zymek, 2022). Therefore, I linked the selection of the countries to the import and export volume of oil, as it is reasonable to expect that the stronger a commodity dominates a country's economy, the more sensitively that country's bond yield curve will react to changes in the commodity's price.

I tested different approaches to formulate each country's dependency on oil by analysing the ratio between the oil trade balance to the overall balance of trade and by observing the impact of oil on GDP. However, as this associates the oil trade volume with other macroeconomic variables, which themselves are influenced by the underlying commodity, I opted to only look at the top ten importers and exporters, based purely on the absolute values to avoid a selection bias regarding the reference value.

A final criterion to be met by the countries is the availability of the sovereign bonds data. I applied the Refinitiv Benchmark Indices as proxies for the behaviour of the sovereign bonds, yet that imposes the limitation that those indices do not exist for all countries. Hence, if there is no availability of the benchmark indices for different maturities, I excluded the country from the selection.

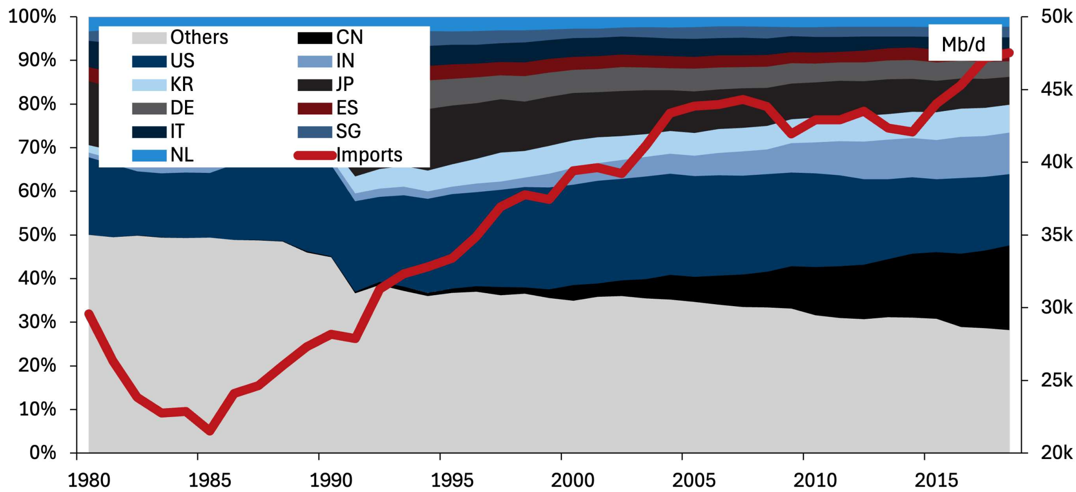

Figure 1: *Crude imports per country (inc. lease condensate)*

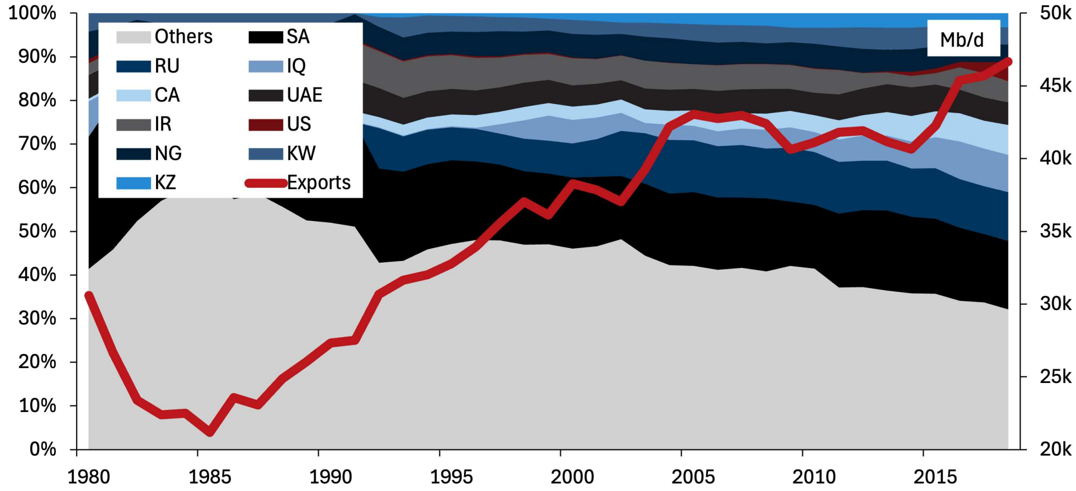

Figure 2: *Crude exports per country (inc. lease condensate)*

Figure 1 visualises the top ten importers, whereas Figure 2 shows the top ten exporters of crude oil including lease condensate. Apart from the United States and Canada, the oil-exporting countries do not provide sufficient sovereign bonds data for the analysis. However, the United States is already analysed as part of the oil-importing sample (Figure 1) and hence, does not need any further analysis from an oil-exporting perspective, as the empirical analysis applied in this thesis is identical for oil-importing and oil-exporting countries. This yields Canada as the only exporting nation.

The original research design included a comparative analysis of oil shocks on sovereign bonds of oil-importing and oil-exporting countries. From Figure 2, I gained evidence that there is suboptimal data to implement a comparative analysis. Taking Canada's unique national profile and its concentrated oil-export relationship with the United States – 97% of oil exports go to the United States ("Crude Oil Industry Overview", 2023) – it cannot be considered a representative proxy for oil-exporting countries on a broader scale. As this thesis strives to gain generalised insights across countries without focusing on unique specificities, I decided to modify the study design to only incorporate oil-importing countries.

4 Methodology

After specifying the variables used throughout this analysis and their rational, this section outlines the methodologies, transformations, tests and models, their theory and function, used in the empirical analysis to build a resilient dynamic system.

4.1 Data Transformations

Before conducting further research to make the final assembly of the VAR model possible, another final step is needed. Although the data is sufficiently prepared for a shock modelling scenario in general, the VAR model used in this thesis requires a specific transformation to build a stable and yet dynamic system, which I will discuss throughout this section.

To ensure stability and subsequently model reliable shock scenarios, a VAR model assumes its data to be stationary and its residuals not to be serially correlated. A proven strategy to prevent or at least decrease serial correlation is to transform the data by taking the differences of the logarithms (Casoli et al., 2024; Wooldridge, 2010). Nevertheless, transforming the data this way makes it more difficult to interpret the results and faces the issue that the natural logarithm is not defined for negative values. To evade this problem, a constant term is added within the logarithm. Hence, the oil variables discussed above are transformed using this formula $\Delta \log(1 + x_t) = \log(1 + x_t) - \log(1 + x_{t-1})$.

The sovereign bonds data itself does not need any further transformations. The raw sovereign bonds data is not fed into the VAR model itself, but rather into the Nelson-Siegel model to extract the *level* (β_0), *slope* (β_1) & *curvature* (β_2) of the yield curve. The Nelson-Siegel model does not necessitate any kind of transformation for the input data (Nelson & Siegel, 1987). Nonetheless, these modelled variables must be differenced to enhance stationarity and address serial correlation (Wooldridge, 2010). Given the characteristics of the data — negative values bigger than -1 and small numbers — it is more suitable to use the standard difference rather than the logarithmic difference $\Delta x = x_t - x_{t-1}$.

The usage of different transformations for oil and sovereign bonds data is not necessarily a problem, but going forward it does introduce further challenges, as the interaction of the impulse and response variables becomes more complex. The different scale of the variables must be noted. Consequently, if a log-transformed variable is shocked on a non-log-transformed variable, it results in a change of absolute units rather than percentages.

4.2 Nelson-Siegel Model

The Nelson-Siegel model is a parametric framework designed to dynamically capture the term structure of interest rates and government bond yields. This methodology is specifically designed to model the forward rate curve based on four parameters: β_0, which proxies the long-term interest rate level, β_1, which represents the short-term level, and β_2, which models the mid-term curvature of the yield curve. The fourth parameter λ can be seen as a control parameter in conjunction with β_2 (Nelson & Siegel, 1987; Diebold & Li, 2006).

$$f(\tau) = \beta_0 + \beta_1 e^{-\lambda\tau} + \beta_2 \tau\lambda e^{-\lambda\tau} \tag{4}$$

When analysing the equation of the Nelson-Siegel model by assigning $\tau = 0$ and converging $\tau \longrightarrow \infty$, it is shown that each parameter corresponds to the explanation above. If $\tau = 0$, the term simplifies to $\beta_0 + \beta_1$, which proves that β_1 models the short-term level, as β_0 is the result of the model under the convergence constraint of $\tau \longrightarrow \infty$ representing the long-term level. β_2 serves as a bridge between β_0 and β_1 and therefore defines the shape on the back of λ (Diebold & Li, 2006). Moreover, Equation (4) is used to derive the forward rate curve (Nelson & Siegel, 1987).

$$y_t(\tau) = \beta_0 + \beta_1 \frac{1 - e^{-\lambda\tau}}{\lambda\tau} + \beta_2 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) \tag{5}$$

This equation is evaluated on the panel data for each day throughout the different tenors of sovereign bonds. To fit the model and estimate the parameters, an Ordinary Least Squares (OLS) algorithm is used that minimises the Sum of Squared Errors (SSE), which is calculated by Equation (6) (Wooldridge, 2010).

$$[\beta_0, \widehat{\beta_1, \beta_2}, \lambda]^\top = \arg\min \; SSE = \sum_{i=1}^{n} \epsilon^2 = \sum_{i=1}^{n} (y_{i,\text{obs}} - y_{i,\text{pred}})^2 \tag{6}$$

Finally, the estimated parameters are transformed using the methodology discussed in Section 4.1. These transformed variables serve as input to the VAR model, alongside the oil data.

4.3 Augmented Dickey-Fuller Test

A strict requirement of VAR models is the stationarity of the data (Wooldridge, 2010). The Augmented Dickey-Fuller Test (ADF) is the standard method to evaluate if the data complies with this restriction, or if further steps must be taken to transform the data into stationary data. The concept of a stationary process is best described by the example of a random walk W_t, which by definition is defined as per below:

$$\begin{aligned} W_0 &= 0 \\ W_t &= W_{t-1} + Z_t, \quad \text{for} \quad t \geq 1 \end{aligned} \tag{7}$$

with Z_t being a sequence of randomly distributed variables that fulfil the Independent and Identically Distributed (i.i.d.) criterion. The distribution of the variables is not necessarily normal, it can be any type of probability distribution. However, to avoid unnecessary complexity, it is assumed to be normal for Z_t and it can therefore be represented as below (Wooldridge, 2010):

$$Z_t \sim N(0, \sigma^2) \tag{8}$$

Strict stationarity is present if the joint probability distribution of random variables is equal to the joint probability distribution of the random walk process including a Δ in time t. If this is given, the process itself is time neutral. The more common type of stationarity is weak stationarity, which is defined in Equation (9) (Wooldridge, 2010; Fuller, 1995).

$$\begin{aligned} \forall t, \quad \mu &= E(W_t) \\ \forall t, \quad \sigma^2 &= Var(W_t) \\ \forall t, \quad Cov(X_t, X_{t-\Delta t}) &= \gamma(\Delta t) \end{aligned} \tag{9}$$

Using the ADF Test, the data can be tested for weak stationarity. The ADF Test is a generalisation of the standard Dickey-Fuller Test, which simply tests for the presence of a unit root by running a regression with one lag variable (Fuller, 1995; Dickey & Fuller, 1979).

$$y_t - y_{t-1} = \Delta y_t = \alpha + \delta y_{t-1} + \epsilon_t \tag{10}$$

In comparison, the ADF Test offers the possibility to include multiple lags, which makes it applicable for processes with $AR(n)$ (Fuller, 1995; Dickey & Fuller, 1979).

$$y_t - y_{t-1} = \Delta y_t = \alpha + \delta y_{t-1} + \sum_{i=1}^{n} \gamma_i \Delta y_{t-i} + \epsilon_t \tag{11}$$

In this regression model, the key variable is δ, as it determines the presence of a unit root if one exists. Based on this knowledge the testing hypothesis is defined as (Fuller, 1995; Dickey & Fuller, 1979):

$$\begin{aligned} H_0 &= \delta = 0, \quad \text{non-stationary} \\ H_1 &= \delta < 0, \quad \text{stationary} \end{aligned} \tag{12}$$

The ADF Test is testing for stationarity of the oil data and the Nelson-Siegel modelled sovereign bond yields, and therefore whether they are valid to use in a VAR model. In the following subsection, I will discuss the dynamic oil market VAR model.

4.4 Vector Autoregressive Model

Before discussing the functionality and methodology of the VAR model, I will clarify why I chose this type of modelling technique for this specific use-case. A VAR model treats all variables as endogenous. Hence, the VAR model evaluates the variables' states based on the internal interactions of all other variables. This is favourable, as the relationship between the oil variables and the term structure is bidirectional due to the intrinsic feedback loop within the model. Therefore, the issue of reverse causality is annulled – one variable causes an effect on the other variable and vice versa. Additionally, a VAR model offers a unique approach to test the behaviour of the system under the condition of an unexpected shock of one of the variables and how the affected variables respond to this shock. Hereby, the model accounts for time dynamics and lagged effects. Yet, it must be stated that the plain VAR model is solely based on the mathematical interconnectedness of the variables and is not restricted by theoretical assumptions. This reduces an assumption bias in the model and, consequently, the room for misspecification, but particularly during structural analysis, this can lead to unexpected results, as the mathematically driven variable interactions are not backed by economic theory and thus might produce unrealistic results. To introduce theory-backed assumptions to the variable interactions of the VAR model, a SVAR model must be used (Lütkepohl, 2005).

4.4.1 Model

A VAR model is a system of predictive Generalised Linear Models (GLM). Hence, a VAR model consists of multiple dependent variables, each defined by a regression model on the predecessors of its peer variables. The lagged variables are acting as independent predictors, thus benefitting from the introduction of autoregressive effects. They are not only independently influencing the models' variables but impacting all variable-containing equations. Consequently, a model of k variables and a lag of p depends on the lagged estimates of itself and is described by $VAR(p)$. For example, within a system of three variables $k = 3$ (x, y, z), x_t is dependent on x_{t-q}, y_{t-q} & z_{t-q}, for $q \in \{1, 2, \ldots, t-q\}$. Following this logic, the equations below are constructed (Lütkepohl, 2005; Stock, n.d.):

$$
\begin{aligned}
x_t &= \gamma_x + \sum_{q=1}^{p} (a_{x,q} x_{t-q} + b_{x,q} y_{t-q} + c_{x,q} z_{t-q}) + \epsilon_{t,x} \\
y_t &= \gamma_y + \sum_{q=1}^{p} (a_{y,q} x_{t-q} + b_{y,q} y_{t-q} + c_{y,q} z_{t-q}) + \epsilon_{t,y} \\
z_t &= \gamma_z + \sum_{q=1}^{p} (a_{z,q} x_{t-q} + b_{z,q} y_{t-q} + c_{z,q} z_{t-q}) + \epsilon_{t,z}
\end{aligned}
\tag{13}
$$

These equations can be generalised into:

$$
y_t^i = \gamma^i + \sum_{q=1}^{p} \sum_{j=1}^{k} \beta_{j,q} y_{t-q}^j + \epsilon^i \tag{14}
$$

, where γ^i is the intercept for the i-th equation, ϵ^i is the error term and $\beta_{j,q}$ is the OLS estimated coefficient for the j-th variable (Lütkepohl, 2005; Stock, n.d.).

Additionally, this condensed form can then be vectorised into the following matrix notation:

$$
Y_t = \Gamma + \sum_{q=1}^{p} K_q Y_{t-q} + \varepsilon_t \tag{15}
$$

Hereby, Γ is a vector of all intercepts with the dimensions of $k \times 1$, K_q $(k \times k)$ represents the coefficients and captures the influence on the preceded variables Y_{t-q} $(k \times 1)$, and ε_t is the error term at time t $(k \times 1)$ (Lütkepohl, 2005; Stock, n.d.).

4.4.2 Determining the Optimal Lag

An integral part of the VAR model is the specification of the lag p. More precisely, how many predecessors are used to describe the variables at time t. Hereby, a balance between achieving a good fit, avoiding overfitting, and minimising excessive data usage must be found. The information criteria evaluated in this thesis are: AIC, HQ, SC & FPE. As a result of the empirical analysis – testing and research – in combination with the existing literature, I chose AIC as the information criterion for all VAR models throughout this thesis (Lütkepohl, 2005; Ivanov & Kilian, 2005).

Based on Lütkepohl (2005) and the research by Kilian (2001), the AIC is often the preferred information criterion in the context of forecasting and impulse response analysis. In essence, Lütkepohl (2005) and Kilian (2001) argue that although the AIC tends to overfit the model by asking for a larger amount of lags, it reflects the long-term variable interactions the best. However, this can introduce side effects into the model with severe impact on the precision and accuracy of the model's forecasts and impulse responses. Nevertheless, by including more lags, the critical case of misspecification can be minimised. Equation (16) is the formula for AIC.

$$AIC(p) = ln(\Sigma(p)) + \frac{2K^2p}{T} \tag{16}$$

The first term of the equation $ln(\Sigma(p))$, which is the logarithm of the covariance matrix of the residuals ($Cov(\varepsilon_t)$), accounts for the fit of the model, whereas the second part ($\frac{2K^2p}{T}$) penalises complexity. That means, the higher the lag, the more parameters must be estimated, and the severer the penalisation (Lütkepohl, 2005; Ivanov & Kilian, 2005).

4.4.3 Robustness Test – Serial Correlation

For accurate forecasts and shock results, the VAR model demands the residuals not to be serially correlated. This means that the error terms need to be independent and truly random. Hence, a previous error term does not influence and has no predictive power regarding the current error term (Lütkepohl, 2005):

$$P(\epsilon_t \mid \epsilon_{t-q}) = P(\epsilon_t), \text{ for } q \in \{1, 2, \ldots, t-q\} \tag{17}$$

A common and visual approach to detect autocorrelation effects is the usage of autocorrelation and partial autocorrelation charts. Based on those, it can be seen at which lag which intensity of autocorrelation occurs, and thus the number of lags can be amended accordingly. However, the usage of autocorrelation charts does not cover all possibilities, as the autocorrelations might depend on each other. Consequently, these charts might severely underestimate the occurrence of serial correlation (Lütkepohl, 2005).

To avoid this pitfall, I used a more accurate test: the Portmanteau Test. This test evaluates the autocorrelation effects for different lag lengths. The test statistic Q_h, where h is the number of lags included, is given below (Lütkepohl, 2005).

$$\begin{aligned} Q_h = T \sum_{j=1}^{h} \operatorname{tr}(\hat{\Sigma}_j' \hat{\Sigma}_0^{-1} \hat{\Sigma}_j \hat{\Sigma}_0^{-1}) \\ \hat{\Sigma}_i = \frac{1}{T} \sum_{t=i+1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_{t-i}^T \end{aligned} \tag{18}$$

The two hypotheses are:

$$\begin{aligned} H_0: \quad \Sigma_0 = \Sigma_1 = \Sigma_h = 0 \\ H_1: \quad \exists\, \Sigma_i \neq 0, \text{ for } i \in \{1, 2, \ldots, h\} \end{aligned} \tag{19}$$

, where Σ is defined as the covariance matrix of the residuals. Accordingly, the following emerges: If the residuals exhibit significant autocorrelation at any lag up to h, H_0 must be rejected (Lütkepohl, 2005).

4.4.4 Robustness Test – Heteroskedasticity

After testing for serial correlation, I am investigating the residuals for heteroskedastic effects. Heteroskedasticity describes the existence of a time-varying variance, which should not occur for a fully robust regression model, as this would require the residuals to be homoskedastic. In comparison to serial correlation effects, heteroskedasticity does not influence the result or the reliability of the tested VAR model, it rather affects the standard errors and t/F-statistics. Hence, if the existence of heteroskedasticity is proven, the standard errors need to be adjusted to paint a credible picture (Wooldridge, 2010).

The test I applied is Engle's ARCH Test, who formulated the two hypotheses (Engle, 1982):

$$\begin{aligned} H_0: \quad \alpha_1 = \alpha_2 = \alpha_p = 0 \\ H_1: \quad \exists\, \alpha_i \neq 0, \text{ for } i \in \{1, 2, \ldots, p\} \end{aligned} \tag{20}$$

In this case, α_X are the constants used in the $ARCH(p)$ model, which is specified by the following equation (Engle, 1982):

$$\begin{aligned} \sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + ... + \alpha_p \epsilon_{t-p}^2 \\ \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 \end{aligned} \tag{21}$$

σ_t^2 is the modelled volatility of the residuals (ϵ_t) of each of the VAR models' equations, assuming that $\epsilon_t \sim N(0, \sigma_t^2)$ (Engle, 1982).

The next step of the testing procedure is a predictive regression on the squared residuals to test for autocorrelation within the error terms. Engle (1982) is using the effect that if the squared errors are autocorrelated, the plain error itself must have heteroskedasticity effects (Engle, 1982). Furthermore, the Lagrange Multiplier (LM) test statistics are evaluated (Heij, 2004):

$$LM = nR^2 \tag{22}$$

n is the sample size of the data, R^2 is defined as the goodness-of-fit and $LM \sim \chi^2(p)$. Thus, the p-value or critical value computation based on the χ^2 distribution must be conducted (Heij, 2004).

4.4.5 Robustness Test – Variable Stability

The VAR model assumes its parameters to be constant over time. One approach to test for constancy within a sample period is to run the CUSUM Test (Heij, 2004). The CUSUM Test searches the residuals within a regression model for deviations in its average value, which would be a strong indicator for the existence of a structural break that occurred during the sample time of the tested variable (Heij, 2004; Brown, Turbin, & Evans, 1975).

The test is based on the assumption of normality and the addition of the periods' average in a cumulative procedure, which is continuously compared to a previously defined critical value. Hence, if the critical boundaries are breached, the null hypothesis of stable regression coefficients is rejected (Heij, 2004; Brown et al., 1975).

$$W_r = \sum_{t=k+1}^{r} \frac{w_t}{s} \tag{23}$$

w_t is the average of the sample and s is the standard deviation of the residuals for all observations (Heij, 2004). These values are afterwards compared to the critical boundaries as discussed by Heij (2004):

$$|W_r| > 2\sqrt{r-k} \tag{24}$$

4.4.6 Robustness Test – System Stability

System stability is a decisive factor for the VAR model to be considered robust. A VAR model is a system of equations, and hence it must be ensured that if a variable gets shocked, the system behaviour remains predictable and does not diverge but rather dissipates over time. Therefore, in an unstable VAR model the shock might persist indefinitely rather than converge to the initial state. In this case, the VAR model would produce disturbed forecasts and shock results. Thus, the VAR model's specification is insufficient for the provided use-case and a different model, for example a Vector Error Correction Model (VECM) should be taken into consideration (Lütkepohl, 2005).

Lütkepohl (2005) establishes the following condition for system stability:

$$det(I - Az) \neq 0, \quad \text{for } |z| \leq 1 \tag{25}$$

In this case, I is the identity matrix, A describes the companion matrix and z is the variable or variables to be solved. Therefore, we can consider the VAR model as stable if no roots are in or on the unit circle with $z \in \mathbb{C}$ (Lütkepohl, 2005).

Furthermore, Pfaff and Stigler (2024) argue that the definition of system stability by Lütkepohl (2005) is identical to the following:

$$\forall i, \quad |\lambda_i| < 1 \tag{26}$$

λ_i are the eigenvalues of the VAR model. Hence, all eigenvalues of the system must be embedded in the complex unit circle ($\lambda_i \in \mathbb{U}$) to be able to consider the system stable (Pfaff & Stigler, 2024).

4.4.7 Impulse Response Function

The Impulse Response Function (IRF) is the key component of the impulse response analysis and further forms the core component of this thesis. Hereby, it is of high interest to study the response of a variable based on an impulse on an alternative variable. The strength of the VAR model is the simulation of the impulse response behaviour not as a stand-alone relation but embedded within a multidimensional system. Additionally, an exogenous shock can be propagated through time, which then results in a forecast of shock scenario responses (Lütkepohl, 2005).

Assuming a generic $VAR(1)$ model as defined in Equation (15), an exogenous shock is simulated by modifying the error term u_t. Thus, the variable of interest is activated by applying the impulse value, whereas all other variables are deactivated by setting their values to zero (Lütkepohl, 2005).

$$\begin{gathered} u_0 = e_i \cdot \alpha \\ \begin{bmatrix} u_{1,0} \\ u_{2,0} \\ \vdots \\ u_{n,0} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \cdot \alpha \end{gathered} \tag{27}$$

This term scales the unit vector e_i by α and adjusts u_t accordingly. Additionally, this impulse is propagated forward in time to analyse the time-varying impact following the shock (Lütkepohl, 2005).

5 Empirical Analysis

5.1 Structure

This section builds the bridge between the data selection, the transformation process and the methodology used to implement and test the model on a practical and theoretical layer. Although the methodology section is already structured in the order of application, Figure 3 visually showcases all steps throughout the empirical analysis.

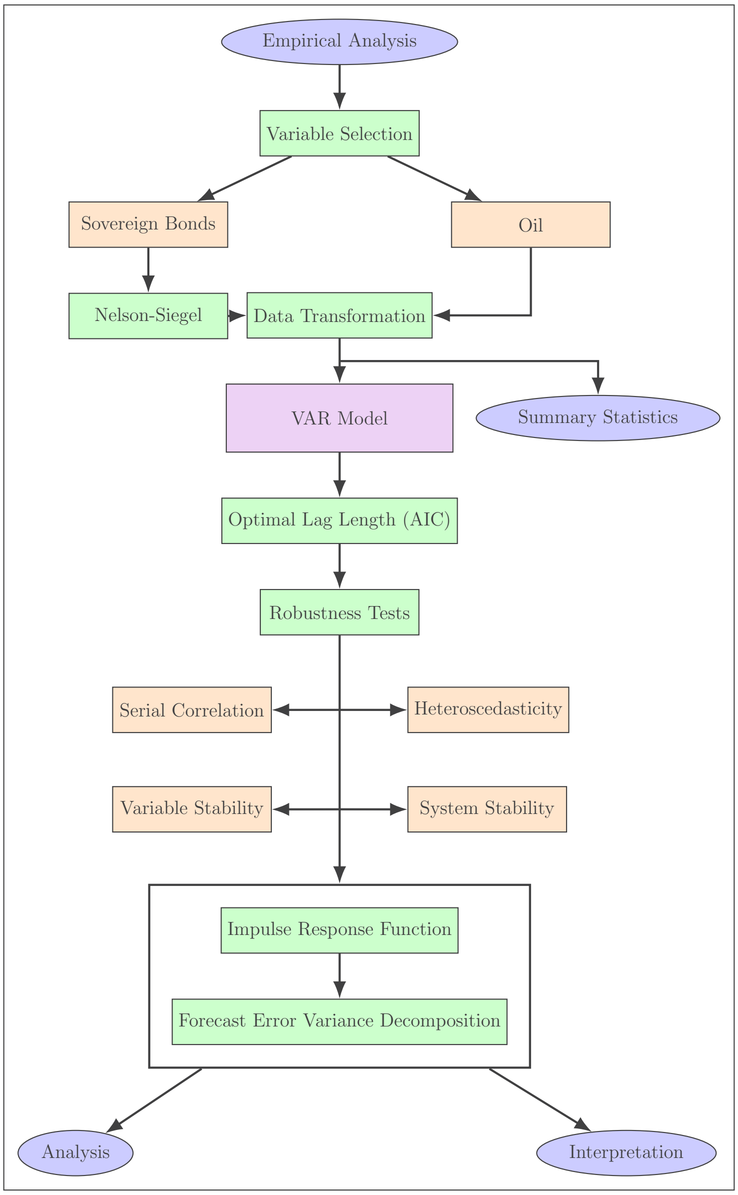

Figure 3: *Overview of the Empirical Analysis*

5.2 Data

On the basis of the research conducted by Casoli et al. (2024) and Wooldridge (2010), the input variables for the VAR model are required to be stationary. To test the data for stationarity, I applied the ADF Test on each series. Evaluating the ADF on the raw data series results in not rejecting the hypothesis H_0, which means that the data is non-stationary and hence not applicable within a VAR model. Consequently, the data transformations discussed in Section 4.1 are imposed on the raw data series. Accordingly, the hypothesis H_0 can be rejected with a p-value of 0.01 for all variables, ensuring stationary data. The detailed ADF results can be found in Table 6 and Table 7 in the Appendix.

5.3 VAR Model Configuration

An integral part for the success and validity of a VAR model is the correct specification of the lags used. On the one hand, I had to model the data complexities without overfitting the historical training data, and on the other hand I had to balance the data constraints regarding the number of observations. Given that the data available is not on a high frequency scale but on a monthly one, the restriction regarding the sample size for each country and shock event is eminent.

However, as a preliminary step in determining the appropriate lag length, I tested multiple information criteria and analysed each country for each shock. The information criteria observed are: AIC, HQ, SC & FPE. As already implied within Section 4.4.2, AIC is the main information criterion used as a baseline for the optimal lag length. Yet, this is a finding based on a combination of literature review and practical research and hence, I will discuss the reasoning for using the AIC in this section rather than in the Methodology.

Furthermore, to not solely rely on the literature and to validate that the literature is in line with the data used in this research, I observed the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) charts on a randomly selected sample of a combination of countries and shocks. The result does not invalidate the usage of AIC, however, a seasonal dependency in the data is evident. Thus, to address the seasonal component of the data, I added eleven dummy variables to the input data. Nevertheless, the AIC is validated as a good indicator for the lags of the VAR model and is further used in the model parametrisation.

To get an unbiased overview and to spot large discrepancies among the information criteria, I evaluated all four criteria for each shock event per country. This yields the results shown in Table 2.

		Global Oil Real Price				Global Oil Inventories				Global Oil Production				Global Economic Activity			
		AIC	HQ	SC	FPE	AIC	HQ	SC	FPE	AIC	HQ	SC	FPE	AIC	HQ	SC	FPE
Panel A:																	
1	United States	4	1	1	4	4	1	1	4	4	1	1	4	4	1	1	4
1	China	16	1	1	1	1	1	1	1	1	1	1	1	16	1	1	2
1	Germany	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	2
1	Spain	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	India	2	1	1	2	2	1	1	2	3	1	1	3	2	1	1	2
1	Italy	3	1	1	3	3	1	1	3	3	1	1	3	2	1	1	2
1	Japan	3	1	1	3	2	1	1	2	2	1	1	2	2	1	1	2
1	South Korea	3	1	1	3	2	1	1	2	3	1	1	2	2	1	1	2
1	Netherlands	2	1	1	1	1	1	1	1	1	1	1	1	2	1	1	2
1	Singapore	2	1	1	2	2	1	1	2	2	1	1	2	2	1	1	2
Panel B:																	
2	United States	4	1	1	4	4	1	1	4	4	1	1	4	4	1	1	4
2	China	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	Germany	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	Spain	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	India	3	1	1	3	3	1	1	3	3	1	1	3	3	1	1	3
2	Italy	3	1	1	3	3	1	1	3	3	1	1	3	3	1	1	3
2	Japan	2	1	1	2	2	1	1	2	2	1	1	2	2	1	1	2
2	South Korea	3	1	1	3	2	1	1	2	3	1	1	3	2	1	1	2
2	Netherlands	2	1	1	2	2	1	1	2	2	1	1	2	1	1	1	1
2	Singapore	2	1	1	2	2	1	1	2	1	1	1	1	1	1	1	1

Table 2: *Country level information criteria by shock event*

When examining Table 2, there are two main findings. For each VAR model, HQ and SC equal factor 1, indicating a lag length of 1, whereas AIC and FPE suggest different lag lengths. Per se, this is neither positive nor negative, as I am expecting HQ and SC to return lower values than AIC and FPE, due to a higher penalisation in the calculation (Lütkepohl, 2005). However, the non-existence of variability in HQ and SC for all of the VAR models in combination with the previously conducted ACF and PACF experiment raises serious concerns whether a lookback window of 1 is really a decisive factor to capture the complex intricacies and interferences of the system. In contrast, AIC and FPE do offer different lag lengths for various VAR models, yet without experiencing large divergences.

There are two values and therefore two shock events that are big outliers regarding their AIC, both of which refer to China. Firstly, the VAR model testing an oil price shock ($GORP$) of 1.154σ in May 2020 (Table 1) and secondly, testing the economic activity shock (GEA) of 4.447σ in January 2020 (Table 1). This is a strong signal that there is a fundamental difference in the underlying sovereign bonds data of China and how China's sovereign bonds interfere with the oil variables. An AIC of 16 for this country's model suggests a poorer fit compared to the other models, as their AIC values are all at 4 or less. This indicates greater difficulty in capturing the system's behavior for China.

According to Lütkepohl (2005) and Hamilton (1994), there is a multitude of reasons why such a response arises. Firstly, differences in the data volatility and noise might cause such a result. Secondly, the dynamics of China might not be fully captured by the variables within the system, hence, an omitted variable error occurs. This might go hand in hand with non-linearities for the behaviour of China. However, as the AIC for China yields appropriate results for the other shock events and timespans, non-linearities do not seem to be the most reasonable explanation, as those non-linearities would be observable throughout the model. Thirdly, the high AIC might be attributable to structural breaks throughout the time period observed. Yet, it must be noted that the explosion in the lag length only unfolds for the AIC, but for none of the other information criteria. Nonetheless, the high AIC does not necessarily mean that the model does not capture the behaviour of the underlying data and will therefore yield wrong results. Going further, the analysis and interpretation of those shocks must be set in relation to these findings.

5.4 Robustness Tests

The robustness tests are the first step to analyse the quality and correctness of the VAR models. Hereby, the minimum requirements of the VAR model are tested to qualify for further analysis regarding the expectation of valid and trustworthy results. The VAR model necessitates its residuals to not exhibit serial correlation while at the same time being homoskedastic. If both of these constraints are given, the VAR model is considered to be robust (Wooldridge, 2010; Lütkepohl, 2005).

I conducted the Portmanteau Test to see if serial correlation effects are present within the residuals of the models. To justify the use of the VAR model and its inclusion in subsequent analysis, the null hypothesis H_0 must not be rejected. Failure to do so would result in inefficient parameter estimates and incorrect standard errors, thereby compromising the reliability of forecasts and impulse response functions (Wooldridge, 2010; Lütkepohl, 2005).

		Global Oil Real Price		Global Oil Inventories		Global Oil Production		Global Economic Activity	
		Portmanteau	ARCH	Portmanteau	ARCH	Portmanteau	ARCH	Portmanteau	ARCH
Panel A:									
1	United States	0.0982	0.0000 (***)	0.2415	0.0000 (***)	0.1479	0.0000 (***)	0.0058 (**)	0.0001 (***)
1	China	0.0000 (***)	1.0000 1	0.0061 (**)	0.0000 (***)	0.0017 (**)	0.0000 (***)	0.0000 (***)	1.0000 1
1	Germany	0.0180 (*)	0.0000 (***)	0.0038 (**)	0.0000 (***)	0.0009 (***)	0.0000 (***)	0.0045 (**)	0.0000 (***)
1	Spain	0.3120	0.0000 (***)	0.3916	0.0000 (***)	0.2349	0.0000 (***)	0.0029 (**)	0.0000 (***)
1	India	0.4015	0.0000 (***)	0.1073	0.0000 (***)	0.1403	0.0012 (**)	0.0808	0.0000 (***)
1	Italy	0.0005 (***)	0.0000 (***)	0.0017 (**)	0.0000 (***)	0.0006 (***)	0.0000 (***)	0.0000 (***)	0.0000 (***)
1	Japan	0.0191 (*)	0.0000 (***)	0.0002 (***)	0.0000 (***)	0.0000 (***)	0.0000 (***)	0.0000 (***)	0.0000 (***)
1	South Korea	0.0154 (*)	0.0000 (***)	0.0001 (***)	0.0000 (***)	0.0031 (**)	0.0000 (***)	0.0028 (**)	0.0000 (***)
1	Netherlands	0.1234	0.0000 (***)	0.0959	0.0000 (***)	0.0474 (*)	0.0000 (***)	0.0064 (**)	0.0000 (***)
1	Singapore	0.7166	0.0000 (***)	0.6491	0.0000 (***)	0.4610	0.0000 (***)	0.2179	0.0000 (***)
Panel B:									
2	United States	0.2680	0.0000 (***)	0.3382	0.0000 (***)	0.1255	0.0000 (***)	0.3174	0.0000 (***)
2	China	0.0000 (***)	0.0000 (***)	0.0000 (***)	0.0000 (***)	0.0000 (***)	0.0000 (***)	0.0000 (***)	0.0000 (***)
2	Germany	0.0003 (***)	0.0000 (***)	0.0002 (***)	0.0000 (***)	0.0001 (***)	0.0000 (***)	0.0001 (***)	0.0000 (***)
2	Spain	0.0920	0.0000 (***)	0.1036	0.0000 (***)	0.0415 (*)	0.0000 (***)	0.0664	0.0000 (***)
2	India	0.0776	0.0000 (***)	0.0920	0.0000 (***)	0.0526	0.0000 (***)	0.0700	0.0000 (***)
2	Italy	0.0001 (***)	0.0000 (***)	0.0001 (***)	0.0000 (***)	0.0001 (***)	0.0000 (***)	0.0001 (***)	0.0000 (***)
2	Japan	0.0000 (***)	0.0000 (***)	0.0000 (***)	0.0000 (***)	0.0000 (***)	0.0000 (***)	0.0000 (***)	0.0000 (***)
2	South Korea	0.0042 (**)	0.0000 (***)	0.0000 (***)	0.0000 (***)	0.0027 (**)	0.0000 (***)	0.0000 (***)	0.0000 (***)
2	Netherlands	0.0999	0.0000 (***)	0.0368 (*)	0.0000 (***)	0.0616	0.0000 (***)	0.0042 (**)	0.0000 (***)
2	Singapore	0.0562	0.0000 (***)	0.0764	0.0000 (***)	0.0042 (**)	0.0000 (***)	0.0033 (**)	0.0000 (***)

*p <= 0.001 → ***, reject H0 with α = 0.1%*
*p <= 0.01 → **, reject H0 with α = 1%*
*p <= 0.05 → *, reject H0 with α = 5%*

Table 3: *Country level Portmanteau and Engle's ARCH Tests by shock event*

There are two main observations resulting from Table 3. Firstly, there are countries where the underlying variables seem to be sufficient to capture the behaviour of the symbiosis between the sovereign bonds and oil data. On the other hand, there are countries where the null hypothesis (H_0) is rejected, and hence, the VAR model does not fully capture the complex behaviour for these countries. The countries that fail the Portmanteau Test are: China, Germany, Italy, Japan and South Korea. Reasons for having serially correlated residuals are either an incorrectly specified lag length, omitted variables, the usage of non-stationary data or structural breaks during the analysed time periods. Considering the reasoned choice of the lag length in conjunction with the rejected null hypothesis (H_0) of the ADF Tests, the existence of serial correlation is constrained to either omitted variables or structural breaks (Wooldridge, 2010; Lütkepohl, 2005).

Secondly, I observed a time dependence in certain VAR models, although a serial correlation is not exhibited. There is only one country that does not display serial correlation throughout all shocks tested: India. All the other countries have at least one occurrence of serial correlation. For example, the Portmanteau Test for the United States is rejected for the first Global Economic Activity (GEA) shock. Other countries demonstrate the existence of serial correlation for the second Global Oil Inventories (GOI) shock, but not for the first. Therefore, the time difference between the first and second GOI shock accounts for the presence or absence of serial correlation. Thus, there is evidence for the existence of a time-varying component within the models. Although not fully providing proof, it solidifies the presumption that the serial correlation is attributable to structural breaks rather than omitted variables.

In addition to the Portmanteau Test, I applied Engle's ARCH Test, which examines the residuals for heteroskedastic effects. Based on the insights derived from the serial correlation tests and the assumption of a time-varying variance, I expect Engle's ARCH Test to yield strong evidence for significant heteroskedasticity. This assumption is intensified by examining the results in Table 3 more thoroughly.

Table 3 confirms that all VAR models are affected by heteroskedasticity, with the exception of two, which are both related to China in the context of a Global Oil Real Price shock ($GORP$) and a Global Economic Activity shock (GEA). In both cases, the null hypothesis is not rejected, indicating homoskedasticity. On the contrary, both scenarios yield a result of 1 for Engle's ARCH Test, which raises significant concerns, as it appears to be a theoretical value rather than a valid outcome supported by empirical data. It implies that there is no heteroskedasticity in the data. This highly doubtful result alongside the constant serial correlation and the lag selection lead to the conclusion that the specified VAR model does not reflect China's specificities and hence, no insights into the behaviour of Chinese sovereign bonds under oil shocks can be gained (Wooldridge, 2010; Engle, 1982).

Nevertheless, Engle's ARCH Test proves that the conjecture of a time-varying component is correct, which contradicts the assumption of a static variance for a fully robust regression model. Yet, the coefficients estimated by the VAR model for each equation are still unbiased and correct. Hence, the interdependence of the variables does still correctly reflect the real-world behaviour. However, the t-statistics are affected by the existence of heteroskedasticity. Therefore, some of the p-values might be misleading (Wooldridge, 2010). To deal with this, robust standard errors by White (1980) can be used. Nevertheless, due to the relatively small sample size compared to the number of variables, which can damage the validity of the robust standard errors, and because of the objective of studying the impulse response function rather than making an inference on the statistical significance of variables, I did not evaluate the robust standard errors in this particular case (Lütkepohl, 2005; White, 1980).

Although this contradicts to some extent the selection and filtration of the countries during the serial correlation testing stage, these countries are still used in the further analysis, as the point estimates of the impulse response function are not affected by a time-varying variance. On the contrary, it affects the confidence intervals for the impulse response function (Wooldridge, 2010; Lütkepohl, 2005).

To gain further insights into the time-varying component by attributing it to the causing variables, I tested the variable stability through a CUSUM Test, which is run on the residuals of each VAR model. This test demonstrates the cumulative sums of the residuals on a timescale and thereby analyses the behaviour of the models' coefficients. The variables are stable if the coefficients are constant ($\mu = 0$). If the coefficients are not constant, there is evidence for the presence of a time-varying component within that variable. Thus, by applying this test on the VAR models' residuals, the suspicion of a time-varying component is confirmed and observed on a more granular level (Heij, 2004; Brown et al., 1975).

		Global Oil Real Price			Global Oil Inventories			Global Oil Production			Global Economic Activity		
		CUSUM		**Stability**	**CUSUM**		**Stability**	**CUSUM**		**Stability**	**CUSUM**		**Stability**
Panel A:													
1	United States	✓	-	✓	✓	-	✓	✓	-	✓	✓	-	✓
1	China	✓	-	✓	✓	-	✓	✓	-	✓	✓	-	✓
1	Germany	✓	-	✓	✓	-	✓	✓	-	✓	✓	-	✓
1	Spain	✓	-	✓	✓	-	✓	✓	-	✓	✓	-	✓
1	India	✓	-	✓	✓	-	✓	✓	-	✓	X	*(β1)*	✓
1	Italy	✓	-	✓	✓	-	✓	✓	-	✓	✓	-	✓
1	Japan	✓	-	✓	✓	-	✓	✓	-	✓	✓	-	✓
1	South Korea	✓	-	✓	✓	-	✓	✓	-	✓	✓	-	✓
1	Netherlands	✓	-	✓	✓	-	✓	✓	-	✓	✓	-	✓
1	Singapore	✓	-	✓	✓	-	✓	✓	-	✓	✓	-	✓
Panel B:													
2	United States	✓	-	✓	✓	-	✓	X	*(GOI)*	✓	✓	-	✓
2	China	X	*(GOI)*	✓	X	*(GOI)*	✓	X	*(GOI)*	✓	✓	-	✓
2	Germany	✓	-	✓	✓	-	✓	X	*(GOI)*	✓	✓	-	✓
2	Spain	X	*(GOI)*	✓	✓	-	✓	X	*(GOI)*	✓	✓	-	✓
2	India	X	*(GOI)*	✓	X	*(GOI)*	✓	X	*(GOI)*	✓	X	*(β1)*	✓
2	Italy	✓	-	✓	✓	-	✓	X	*(GOI)*	✓	✓	-	✓
2	Japan	X	*(GOI)*	✓	✓	-	✓	X	*(GOI)*	✓	✓	-	✓
2	South Korea	X	*(GOI)*	✓	✓	-	✓	X	*(GOI)*	✓	✓	-	✓
2	Netherlands	X	*(GOI)*	✓	✓	-	✓	X	*(GOI)*	✓	✓	-	✓
2	Singapore	X	*(GOI)*	✓	X	*(GOI)*	✓	X	*(GOI)*	✓	✓	-	✓

Table 4: *Country level CUSUM and stability tests by shock event*

Table 4 presents the results of the CUSUM Test. There are three examined time periods, in which the stability is given for countries throughout all variables: the first Global Oil Real Price shock ($GORP$), the first Global Oil Inventories shock (GOI) and the first Global Oil Production shock (GOP). This underscores that all shocks that are directly linked to oil and occurred prior to and including February 2021 rely on stable variables. However, when shocking the same variables after February 2021, the stability of the variables is not given across all countries. Especially GOI shows strong tendencies of a time-varying element, which spreads through the countries over time until it is unstable by January 2024 at the latest. However, this inclination of instability for GOI is not given with GEA. Additionally, when GEA is analysed, the short-term interest rate (β_1) deviates from its normal behaviour.

It is essential to highlight that the CUSUM Test is a retrospective test to demonstrate if structural breaks occur in the training data. It does not provide predictability regarding future shocks. Hence, up to the point in time when the GEA shock triggers, there has already been a structural break caused by other external factors (Lütkepohl, 2005; Heij, 2004; Brown et al., 1975).

The last robustness test that each of the VAR models must pass is the test for system stability. The system must be stable to generate reliable results for the IRF, since predictability is only given if the response on a bounded input shock reflects in a bounded output. Otherwise, the model does not capture the dynamics and characteristics of the underlying data, and therefore the output of the IRF is not reliable (Lütkepohl, 2005; Pfaff & Stigler, 2024).

Table 4 illustrates that all trained VAR models are stable. However, the stability of a model is not yet a measurement for the goodness-of-fit for the empirical data. Although the model itself may be stable, serial correlation and heteroskedasticity in the residuals can lead to misleading outputs.

5.5 Results – Analysis & Interpretation

5.5.1 Overview

This section presents the overall results of the empirical analysis and will answer the question whether the specified VAR models that passed the statistical tests capture the system behaviour adequately, which would strongly indicate that these models can be employed to model the consequences of oil shocks on sovereign bonds.

To assess the models and answer the research questions, different approaches are possible. A standard practice is to use an FEVD, as this method crystallises the best cross-sectional explanation throughout the model variables (Lütkepohl, 2005). However, this thesis aims to remain closely aligned with practical application rather than focussing solely on theoretical outcomes, thus the IRF is compared to the real-world shock data. Nevertheless, some difficulties result from this test design. Normally, IRFs are visualised in line charts including their respective confidence intervals, resulting in a myriad of charts. Hence, the IRFs are displayed as Histmaps to paint a less granular, but more insightful picture of the variable interactions. By doing so, the confidence intervals are no longer retrievable, but due to the existing heteroskedasticity in the residuals, their informative value is limited and does not play a major role in the ongoing analysis.

To quantify the visual impressions, I decided to calculate the Spearman correlation. Instead of the widely utilised Pearson correlation, I applied the Spearman correlation, because of its ability to assess monotonic relationships rather than linear relationships. Thus, the rate between the movements need not necessarily be constant (Bluhm & Overbeck, 2007). However, the combined picture of visual and quantifiable analysis is needed, as the correlation alone can be deceptive.

I focussed my attention on the variables β_0 (long-term interest rate), β_1 (short-term interest rate) and β_2 (curvature of the yield curve). Since λ is only a control parameter for the rate of exponential decay, it is essential for the shape of the yield curve, but it is not directly associated with a financial quantity. Hence, the values for λ are provided for comprehensiveness, but they do not significantly influence the analysis.

5.5.2 Global Oil Real Price Shock Analysis

Figures 4 and 5 show the impulse responses grouped by the Nelson-Siegel parameters and countries for May 2020 and March 2023. Figure 4 is generated by applying a shock of 1.154σ on the country-specific VAR models. Throughout the variables, the chart demonstrates a cumulation of high responses – positive and negative – within the first three to four shock months. Within this short-term window after the shock, the model is most effective for β_0 and β_1. It successfully captures the negative shock responses for the long-term interest outlook for Spain, India and Singapore and the positive shock for the Netherlands. Additionally, β_1 correctly indicates a flattening of the yield curve for the United States and Singapore compared to a steepening for India and the Netherlands. Similarly, the curvature represented by β_2 has a good representation, especially for Spain and India. On the other hand, the accuracy of the models decays rather quickly. At the latest in the medium term (4-9 months), the performance starts to deteriorate throughout all variables by not capturing trends and stabilisation effects. These effects are further amplified when analysing the long-term effects. In the long run, the model output converges to zero, which must be the case to guarantee system stability. In contrast, the observed real data shows new peaks in the long term. Nevertheless, this appears to be more attributable to other external factors affecting the data than to a direct impact from the original shock that occurred in May 2020.

Overall, the correctness and quality of the models' outputs are highly dependent on country specifics. Whereas the model does a decent job in capturing the short-term interest rate (β_1) and the curvature (β_2) for the United States with a Spearman correlation of 0.308 and 0.441, it fails to model the long-term interest rate (β_0). None of the countries researched is fully reflected by the models. β_0 proves to be a good fit for the Netherlands with a score of 0.769, confined to the early phase of the observation period. In general, the scores for β_2 are all positive, yet they only suggest a weak-to-moderate correlation over the full period of 12 months.

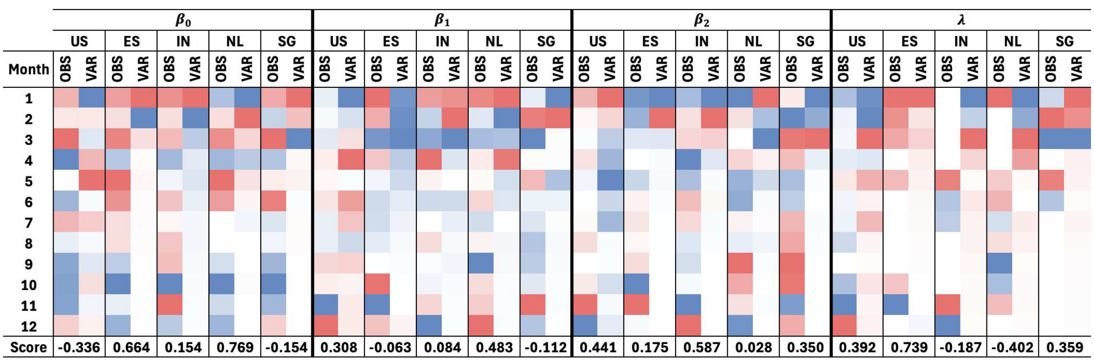

Figure 4: *Global Oil Real Price impulse response analysis for May 2020*

In Figure 5, a comparable trend emerges. In the short-term shock window, the model performs best by successfully predicting β_0 for Spain, India and Singapore. However, for a shock of 1.154σ in March 2023, the model does not reflect the Netherlands correctly, but it does for May 2020. Moreover, the fit for β_1 is rather poor for all variables, which stands in contrast to the curvature of the term structure (β_2), which is decently represented by the models for all countries.

Nonetheless, the same trend as for the first shock scenario occurs. The further the after-shock period progresses, the lower the quality of the model becomes. This observation in combination with other external effects influencing the countries' yield curves in the real world creates large dissipations in the mid- and long-term developments. Looking at the country score (Spearman correlation), the overall correlation effects are either marginal or adverse for countries that provided a decent correlation factor for the first *GORP* shock. An illustrative case is a Spearman correlation of -0.552 for the short-term interest rate of the United States.

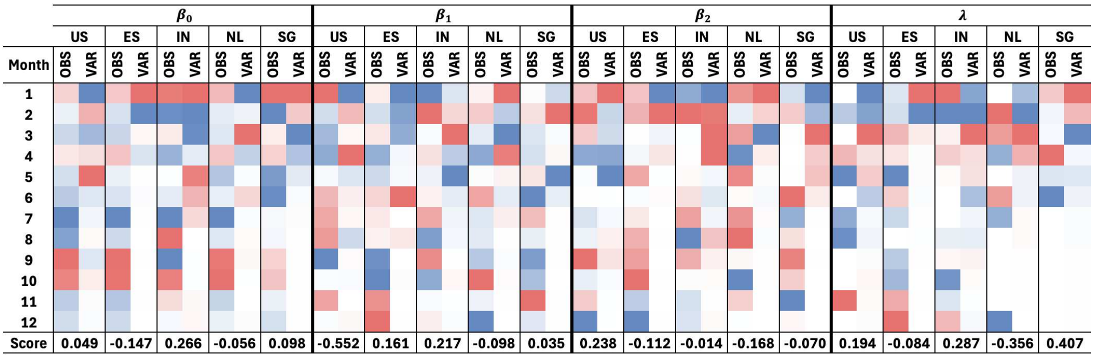

Figure 5: *Global Oil Real Price impulse response analysis for March 2023*

In summary, the model has an overall weak predictive power. There is some forecasting power predominantly in the short term, but this is highly dependent on the country, the variable and the timing of the shock. There are some notable aspects for the first shock scenario, in which the models perform decently, whereas the observed data for the second scenario is dominated by external influences in the mid- and long-term, creating a large amount of noise the VAR models cannot deal with.

5.5.3 Global Oil Inventories Shock Analysis

In this section, the attention is directed toward how a shock in Global Oil Inventories impacts the yield curve for each country. Figure 6 shows a shock of 1.925σ in November 2020 and Figure 7 displays the results of a 1.902σ shock in June 2022.

Figure 6 shows high similarities to the results yielded by the *GORP* analysis. Short-run effects are evident in the United States, Spain, India and Singapore. The VAR models for Spain, India and the Netherlands perform best in β_1 in the first months. Nevertheless, the capturing effect of those models decreases drastically as time progresses. Two findings are more prominently displayed in Figure 6 than in the previous charts. Firstly, the VAR models fail to adequately capture the dynamics for the curvature of the yield curve not only in the mid or long term, but within the first month of observation. This not only stands in contrast to previous shock analyses for β_2, but to all tested variables. This circumstance in conjunction with mid-to-strong negative correlation demonstrates a strong misspecification of the models. Secondly, some countries and variables show that the strongest shock effects do not occur at the shock dates. For example, the long-term interest rate of Singapore shows its highest observed negative shock in month 8 and highest observed positive shock in month 10. The same effect exists for the short-term interest rate of India and the term-structure curvature of Spain. Hence, the observed data is dominated by various other effects within the time period, which the VAR models cannot anticipate.

This assumption aligns with the Spearman analysis. With the exception of β_0 for India, all scores are either negative or close to 0, indicating a minor correlation. Particularly the curvature of the United States distinguishes itself from the *GORP* shocks with a highly negative correlation of -0.692 in comparison to the previous positive weak-to-medium correlation.

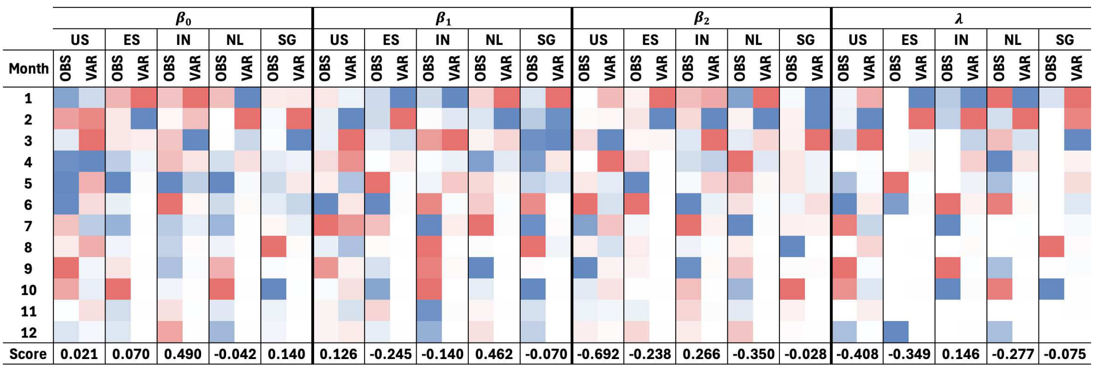

Figure 6: *Global Oil Inventories impulse response analysis for November 2020*

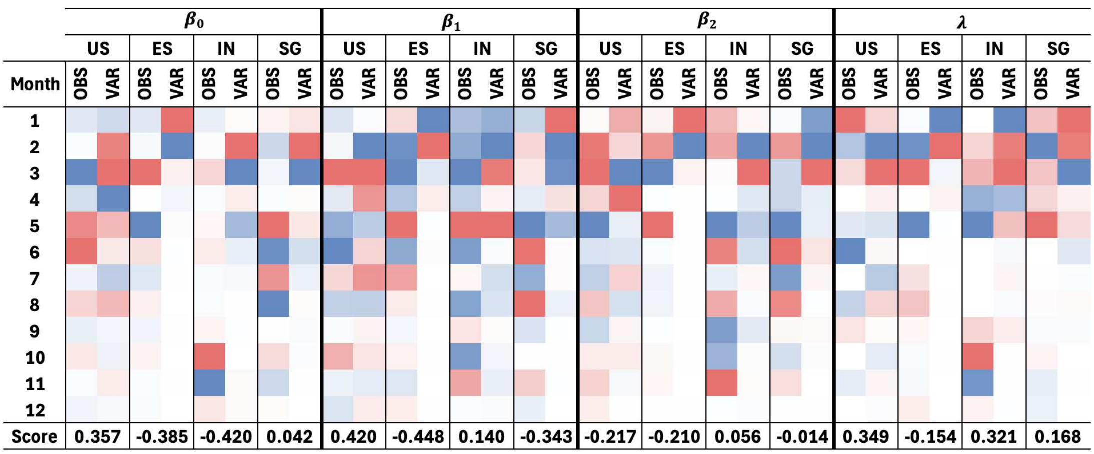

Figure 7: *Global Oil Inventories impulse response analysis for June 2022*

In Figure 7, the existence of external influences introducing noise to the observed data becomes evident. Across all variables (β_0, β_1, β_2), the most significant level change does not coincide with the timing of the tested shock in June 2022. This finding is even more substantial than for the first *GOI* shock and the *GORP* shocks. The interfering signal appears to emerge in the midterm of the observation window, with India being the only outlier.

The Spearman score is relatively high in absolute terms compared to the previously tested shocks. As an example, the United States has a moderately positive correlation for the long- and short-term interest rates with a score of 0.357 and 0.420, whereas Spain is negatively correlated throughout all variables. In conclusion, the time period analysed is dominated by external effects and although there are some variable-country combinations that capture a moderate effect, the results are unreliable and lack robustness across varying shocks.

5.5.4 Global Oil Production Shock Analysis

This section analyses the impact of a Global Oil Production shock on the yield curve. The first examined event applies a 1.769σ shock on the *GOP* variable in February 2021. The second shock has a scale of 1.616σ and strikes in January 2024. As the dataset's most recent timestamp is October 2024, the shock window in this specific case is reduced to ten months.

Figure 8 displays an initial positive response for India and Singapore on the long-term interest rate, whereas the United States and Spain reveal a short-term negative impact on the level (β_1). However, for the slope (β_0) and curvature (β_2), the impulse responses are inconsistent. As previously highlighted during the *GORP* and *GOI* investigations, the accuracy of the model slowly decays within the medium and long term. These observations are again solidified by the Spearman correlation, with two exceptions. Firstly, the country score suggests a moderate correlation for the curvature of Spain. The chart depicts the opposite result and highlights the importance of a combined visual and quantitative approach for the quality of impulse sensitivity. Secondly, the model for India performs exceptionally well in tracking the behaviour and country specifics. The VAR model for India has a weak-to-medium correlation for β_0, a moderate score for β_1 and a decent-to-high Spearman coefficient for β_2. Moreover, the visual evaluation supports these numerical findings. Hence, at this stage, this is the best performing model throughout all shocks. Besides that, Singapore shows signs of weak positive correlation for all variables.

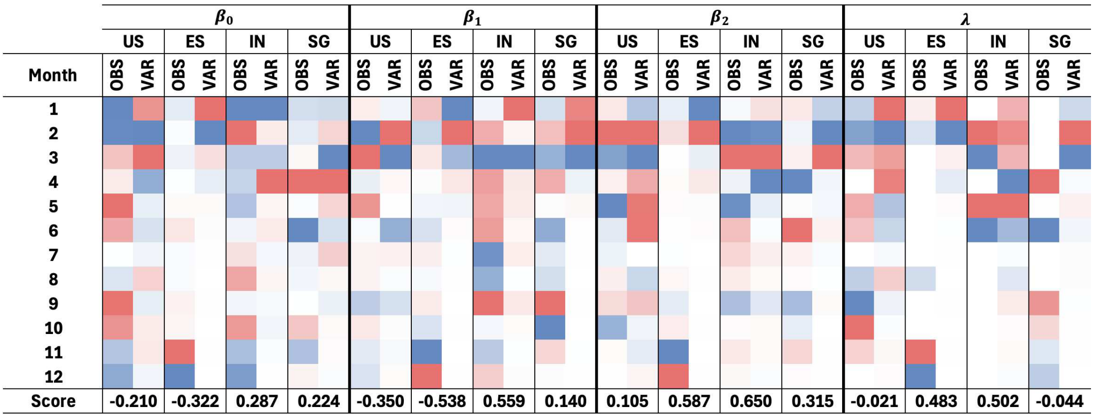

Figure 8: *Global Oil Production impulse response analysis for February 2021*

Particularly when assessing the mid- and long-term trends, external effects start to drastically impact the quality and reliability of the simulation. This holds true for the *GOP* shocks in 2021 and 2024. External influences initiate a performance downgrade of India during the January 2024 shock. Although the country score for the long-term rates increases to 0.467, the score for the level (β_0) decreases to 0.400, showing a moderate correlation. Yet, β_2 flips from a decent-to-high correlation to a weak negative correlation. In a similar manner, this underlines the importance of the time component for the trained VAR models.

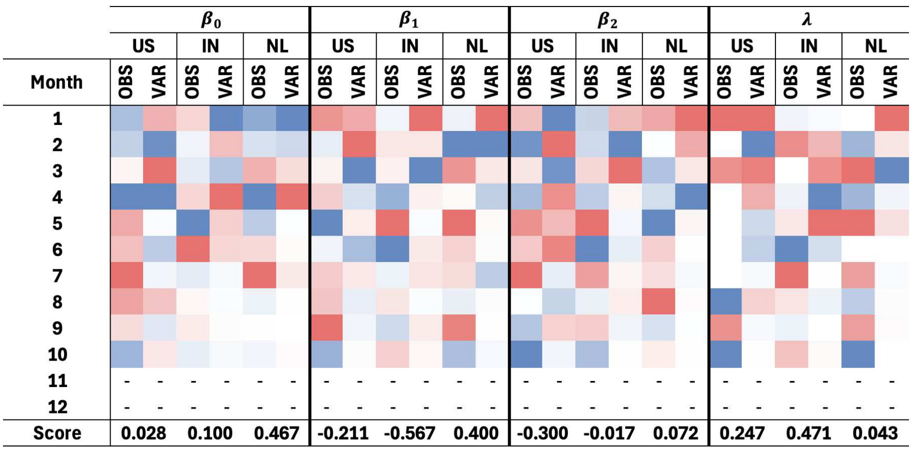

Figure 9: *Global Oil Production impulse response analysis for January 2024*

Due to the Netherlands' failed robustness tests, I excluded it from the first *GOP* shock analysis and hence, a comparison over time is not viable. This is particularly regrettable, as the VAR model for the Netherlands yields a decent characterising capability for β_0 and β_1 (Figure 9). Nonetheless, based on the behaviour of the IRF for the other shocks, it is legitimate to assume that the Netherlands is highly influenced by external shocks as well. Therefore, the timing of the shock plays a major role in the success of the impulse response measurement.

5.5.5 Global Economic Activity Shock Analysis

This section reflects the implications if the Global Economic Activity is shocked. Investigating the role of a GEA shock is particularly interesting, as it is semi-directly associated with oil. It represents the World Industrial Production, which has a strong bond to the oil market, but it is not a variable directly linked to the commodity itself. Therefore, it warrants attention if the IRFs for these shocks yield affirmative, diverging or even undermining results.

Figure 10 displays the IRFs of a 4.477σ shock in January 2020. I already eliminated most of the examined countries during the robustness tests and hence, the comparability between the shocks striking in January 2020 and in April 2022 is limited. The major focus lies on India, as it is the only nation that is represented within the simulations of both shocks (Figures 10 and 11). The long-term interest rate for India starts positively, but worsens after the first month, until it rebalances between the short-term and mid-term observation period. Nevertheless, it follows the same predominant pattern throughout all IRFs I tested in this research. The divergences between the observed data and the modelled data reinforce over time, with the observed data being dominated by external influences. Moreover, the short-term interest rate reveals a slight correlation of 0.231. Besides that, a similar pattern is apparent for all variables. Overall, the models show a poor fit for India and a weak inverse relationship for Singapore.

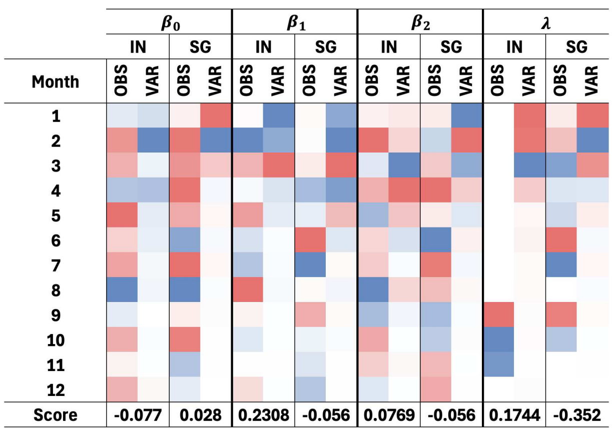

Figure 10: *Global Economic Activity impulse response analysis for January 2020*

However, when incorporating the second shock in April 2022 into the analysis, a heterogeneous picture is presented. While India shows a weak correlation for β_1 in January

2020, this explanatory power is deluded with a Spearman coefficient of 0.014 in April 2022. Already in the initial stage of the observation phase, there is a high fluctuation in the values, but as indicated by the previous shock, there is a balancing effect at the end of the short-term period for all variables and nations. Nonetheless, the capturing quality decreases with an increase of the observed time horizon. It is noteworthy to highlight an increased activity of the VAR values in the medium and long term. In Figure 10, the modelled values converge to zero relatively quickly compared to the shock in 2022, which serves as an indicator that there are outliers or a structural break within the training data, which might corrupt the results of the VAR models even further and thus shrinks their predictive power.

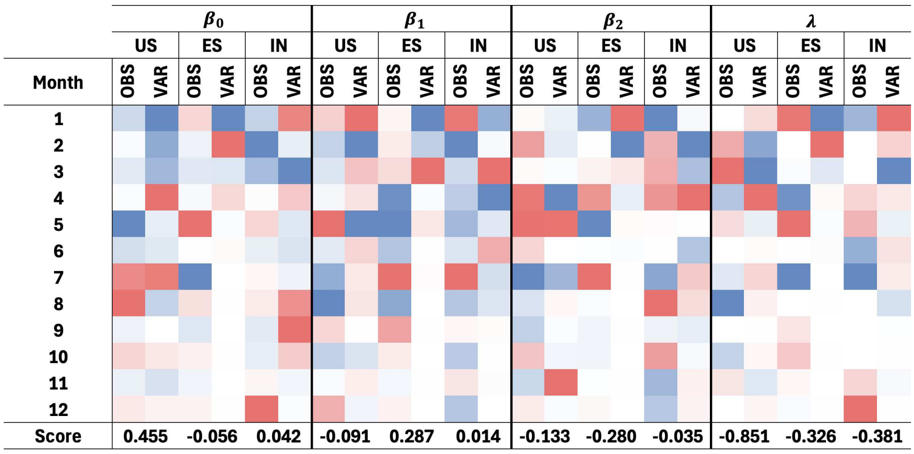

Figure 11: *Global Economic Activity impulse response analysis for April 2022*

Over the full period, β_0 performs exceptionally well for the United States. With a Spearman correlation of 0.455, the model captures the country-specific interactions of the variables decently. However, this effect is not observed within the short-term interest rates and the curvature of the yield curve. To condense, the models perform inadequately when a *GEA* shock is applied. Accordingly, the outcomes exhibit considerable sensitivity to country characteristics and the timing of the shock. This is consistent with patterns observed in earlier analyses.

6 Limitations & Future Research

6.1 Limitations

During this study, I detected a variety of limitations within the data, the methodology and the general study design. The first limitation are the contrasting transformation methodologies for oil and sovereign bonds data. This does not necessarily limit the findings, but the models' outputs are more complex to interpret, because the scale of the impulse variable is given in percent and the growth rate is shocked, but the response of the impulse is represented on an absolute level basis. Furthermore, the results shown in the IRF charts are the stepwise absolute changes of the variables, rather than the movement of the variables over time. Thus, the cumulative sum must be computed. Contingent upon the selected methodology, the analysis might yield contrasting results due to balancing effects of the absolute values during the summation process, which is especially relevant within the context of the Spearman correlation.

Additionally, the sovereign bonds data is not fed into the VAR model in an unprocessed manner, but through a Nelson-Siegel model. This offers the benefit of extracting the main features of the yield curve while reducing the dimensionality. However, as a consequence of this generalisation, it is no longer possible to derive insights at a more granular level regarding the maturity profile of sovereign bonds and thus an information loss occurs.

Moreover, the chosen events for the shocks are solely based on quantitative factors (rolling z-score), without including qualitative factors. This provides an unbiased framework to choose the event, but with the implication of having no differentiation of the shocks. It does not take into account why the shock occurred or what the shock-inducing macro-event was. Therefore, the assumption is that the reaction reflected in the response variable remains consistent, irrespective of the circumstances. This in conjunction with a constrained data history and an accompanying small training and test dataset restricts the window for the IRF analysis to not even five years.

Furthermore, the VAR model assumes its residuals to be homoskedastic. Therefore, a static variance is considered. In contrast, I proved the existence of heteroskedasticity and a subsequent time-varying component, which might, but does not necessarily overestimate the p-values of the Portmanteau Test. An indicator for this is the relatively poor overall performance in capturing the underlying systems' behaviour. Besides the existence of such structural breaks, multiple external and country-specific influences are present in the observed real-world data and these noise signals invalidate the models' outputs.

Equally important is to highlight the limit emerging from the large observation window in the aftermath of the shock events. Although the different terms within the observation window are discussed, the Spearman correlation is evaluated for the full timeframe. Therefore, the result for this might be corrupted by the declining quality of the IRF forecast over time. In addition, calculating the Spearman correlation on such a small sample size might yield unstable results, which must be taken into consideration primarily for the interpretation of predictive power.

6.2 Future Research

Building on the findings presented throughout this thesis, a broad scope of potential future research directions emerges. As the VAR models do not yield generalised results across countries, shocks and periods, further research must be conducted to identify the shortcomings of these models and to investigate the room for improvements. Over the course of the analysis, the suspicion was substantiated that the VAR models are misspecified and incorporate an omitted variable bias. Therefore, finding these control variables is of utmost importance and the first step toward building fully robust VAR models. Additionally, due to the complex underlying data and the country dependence, the idea of introducing country-specific controls must be taken into consideration.

Across the empirical analysis, I observed the existence of heteroskedasticity. To reduce this time-varying component bias and to optimise the resilience of the models to structural breaks and regime changes, future researchers are encouraged to experiment with Structural VAR models, Bayesian Structural VAR models and VECM models (Lütkepohl, 2005), which might go in tandem with an expanded amount of tested timestamps to mitigate time specific noise. Furthermore, to tackle intra-period dynamics, the usage of higher frequency data would be valuable.

On the condition that additional data for the oil-exporting countries becomes available, the original study design of a comparative analysis between oil-importing and oil-exporting countries should be revisited. Alternatively, different country selection processes should be explored. Although I decided to utilise the oil-importing and oil-exporting volume for country selection based on the idea of avoiding biases, applying different selection criteria could provide a broader and more generalised overview of oil shocks across a wider range of countries.

7 Conclusion

The existing literature has proven a strong relationship between oil shocks, macroeconomic variables and equity markets (Jones & Kaul, 1996; Sadorsky, 1999; Kilian & Park, 2009). In contrast, the impact of oil shocks on sovereign bonds and their associated term structure remains a question that has not yet been fully answered by science, as suggested by multiple studies (Kang et al., 2014; Ioannidisa & Ka, 2018). This thesis attempts to close the research gap by assessing a higher number of countries than previous studies, enhancing the analysed shock granularity and by putting a focus on practical applicability in a trading environment. Thus, this thesis considers itself a logical continuation of the research conducted by Ioannidisa and Ka (2018) by deepening the understanding of the subject and by providing a more robust and practically oriented framework.

Throughout this thesis, some consistent patterns across countries and shock scenarios emerged. The longer the duration between the initial shock and the measurement, the more the ability to correctly predict the response deteriorated. Whereas the VAR model converges to zero to guarantee system stability, the real yield curve is further driven by constantly occurring events, which are not necessarily directly linked to oil. Therefore, the long-term effect of oil shocks is not attributable to changes in the term structure, as these effects interfere with external influences.

Overall, the IRF quality is highly dependent on the country, the impulse and the response variable, the shock itself and the timing of the shock. Particularly the timing component plays a major role in the ability of the VAR model to capture the behaviour correctly. This observation in conjunction with big jumps in the correlation and partial reversals underlines that the results are largely driven by a time-varying component, whose existence is backed by the presence of heteroskedasticity in the residuals of the VAR models.

Altogether, none of the models manages to comprehensively identify the behaviour of the term structure in the event of a shock. Although there appear to be positive signs toward real-world applicability throughout the analysis especially for the short-term and long-term interest rates on the shorter end of the observation period, these results are not to be considered reliable due to a lack of robustness across shock events. Therefore, the findings by Ioannidisa and Ka (2018) regarding a rise in the long-term interest rate during an aggregate demand shock and an upward pressure during oil-specific demand shocks for the curvature of the yield curve could not be observed in this thesis.

As a result, the conducted research shows a trend indicating that oil shocks influence yield curve movements, which agrees with Ioannidisa and Ka (2018). Nevertheless, the used VAR models were not able to correctly predict the real world yield curve reactions to the simulated oil shocks, as the findings are not robust and require further substantial research. Hence, using a standard VAR model to analyse the impact of oil shocks on the term structure of sovereign bonds is not sufficient to execute profitable trades or implement a hedging strategy.

8 Bibliography

UNCTAD. (2025). A World of Debt. Retrieved from https://unctad.org/system/files/official-document/osgttinf2025d4_summary_en.pdf

IMF. (2024). 2024 Global Debt Monitor. Retrieved from https://www.imf.org/external/datamapper/GDD/2024%20Global%20Debt%20Monitor.pdf

United States Department of the Treasury. (2025). What is the national debt? Retrieved from https://fiscaldata.treasury.gov/americas-finance-guide/national-debt/

Eurostat. (2024). Structure of government debt. Retrieved from https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Structure_of_government_debt

Fang, X., Hardy, B., & Lewis, K. K. (2023). *Who holds sovereign debt and why it matters* (tech. rep. No. 1099). Bank for International Settlements. Retrieved from https://www.bis.org/publ/work1099.pdf

International Energy Agency. (2025). *Global Energy Review 2025*. International Energy Agency. Retrieved from https://iea.blob.core.windows.net/assets/5b169aa1-bc88-4c96-b828-aaa50406ba80/GlobalEnergyReview2025.pdf

Kilian. (2008). The economic effects of energy price shocks. *Journal of Economic Literature*, *46*(4), 871–909. doi:10.1257/jel.46.4.871

Kilian. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *The American Economic Review*, *99*(3), 1053–1069.

Casoli, C., Manera, M., & Valenti, D. (2024). Energy shocks in the Euro area: Disentangling the pass-through from oil and gas prices to inflation. *Journal of International Money and Finance*, *147*, 103154. doi:10.1016/j.jimonfin.2024.103154

U.S. Energy Information Administration. (2023). *International energy outlook 2023*. U.S. Energy Information Administration. Retrieved from https://www.eia.gov/outlooks/ieo/pdf/IEO2023_Narrative.pdf

United Nations. (2024). World population prospects 2024. Department of Economic and Social Affairs, Population Division. Retrieved from https://population.un.org/wpp/downloads?folder=Standard%20Projections&group=Most%20used

Energy Institute. (2025). Statistical review of world energy.

Baumeister, C., & Hamilton, J. D. (2019). Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks. *American Economic Review*.

Hamilton, J. D. (1983). Oil and the macroeconomy since world war ii. *Journal of Political Economy*, *91*(2), 228–248. doi:10.1086/261140

Hamilton, J. D. (2003). What is an oil shock? *Journal of Econometrics*, *113*(2), 363–398. doi:10.1016/s0304-4076(02)00207-5

Aastveit, K. A., Bjørnland, H. C., & Cross, J. L. (2023). Inflation expectations and the pass-through of oil prices. *Review of Economics and Statistics*, *105*(3), 733–743. doi:10.1162/rest_a_01073

Jones, C. M., & Kaul, G. (1996). Oil and the Stock Markets. *The Journal of Finance*, *51*(2).

Sadorsky, P. (1999). Oil price shocks and stock market activity. *Energy Economics*, *21*.

Kilian, & Park. (2009). The impact of oil price shocks on the u.s. stock market*. *International Economic Review*, *50*(4), 1267–1287. doi:10.1111/j.1468-2354.2009.00568.x

Kang, W., Ratti, R. A., & Yoon, K. H. (2014). The impact of oil price shocks on u.s. bond market returns. *Energy Economics*, *44*, 248–258. doi:10.1016/j.eneco.2014.04.009

Ioannidisa, C., & Ka, K. (2018). The impact of oil price shocks on the term structure of interest rates. *Energy Economics*.

Roncoroni, A., Fusai, G., & Cummins, M. (Eds.). (2015). *Handbook of multi-commodity markets and products: Structuring, trading and risk management*. Includes bibliographical references and index. - Description based on print version record. Chichester, United Kingdom: John Wiley and Sons.

Kilian, & Murphy, D. P. (2013). The role of inventories and speculative trading in the global market for crude oil. *Journal of Applied Econometrics*, *29*(3), 454–478. doi:10.1002/jae.2322

Baxter, M., & King, R. G. (1999). Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series. *The Review of Economics and Statistics*, *81*(4).

Christiano, L. J., & Fitzgerald, T. J. (2003). The band pass filter*. *International Economic Review*, *44*(2), 435–465. doi:10.1111/1468-2354.t01-1-00076

Cuñat, A., & Zymek, R. (2022). Bilateral Trade Imbalances. *IMF Working Papers*, *22/90*.

Crude Oil Industry Overview. (2023). Accessed: 30 July 2025. Retrieved from https://natural-resources.canada.ca/energy-sources/fossil-fuels/crude-oil-industry-overview

Wooldridge, J. M. (2010). *Introductory Econometrics: A modern approach* (4. ed., internat. student ed., repr.). Literaturverz. S. 830 - 834. [Mason, Ohio] [u.a.]: South-Western, Cengage Learning.

Nelson, C. R., & Siegel, A. F. (1987). Parsimonious modeling of yield curves. *The Journal of Business*, *60*(4), 473. doi:10.1086/296409

Diebold, F. X., & Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, *130*(2), 337–364. doi:10.1016/j.jeconom.2005.03.005

Fuller, W. A. (1995). *Introduction to statistical time series*. doi:10.1002/9780470316917

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, *74*(366), 427. doi:10.2307/2286348

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. doi:10.1007/978-3-540-27752-1

Stock, J. H. (n.d.). *Introduction to Econometrics* (Updated third edition) (M. W. Watson, Ed.). Pearson.

Ivanov, V., & Kilian, L. (2005). A Practitioner's Guide to Lag Order Selection For VAR Impulse Response Analysis. *Studies in Nonlinear Dynamics and Econometrics*, *9*(1). doi:10.2202/1558-3708.1219

Kilian. (2001). Impulse Response Analysis in Vector Autoregressions with Unknown Lag Order. *Journal of Forecasting*, *20*, 161–179.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, *50*(4), 987. doi:10.2307/1912773

Heij, C. (Ed.). (2004). *Econometric methods with applications in business and economics*. Includes bibliographical references and index. - Description based on print version record. Oxford: Oxford University Press.

Brown, R. L., Turbin, J., & Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society*, *37*(4), 149–192.

Pfaff, B., & Stigler, M. (2024). Package 'vars'. Retrieved from https://cran.r-project.org/web/packages/vars/vars.pdf

Hamilton, J. D. (1994). *Time series analysis*. Includes bibliographical references and indexes. Princeton, N.J: Princeton University Press.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*(4), 817. doi:10.2307/1912934

Bluhm, C., & Overbeck, L. (2007). *Structured Credit Portfolio Analysis, Baskets and CDOs* (Chapman & Hall, Eds.).

A Appendix

A.1 Oil Variable Overview

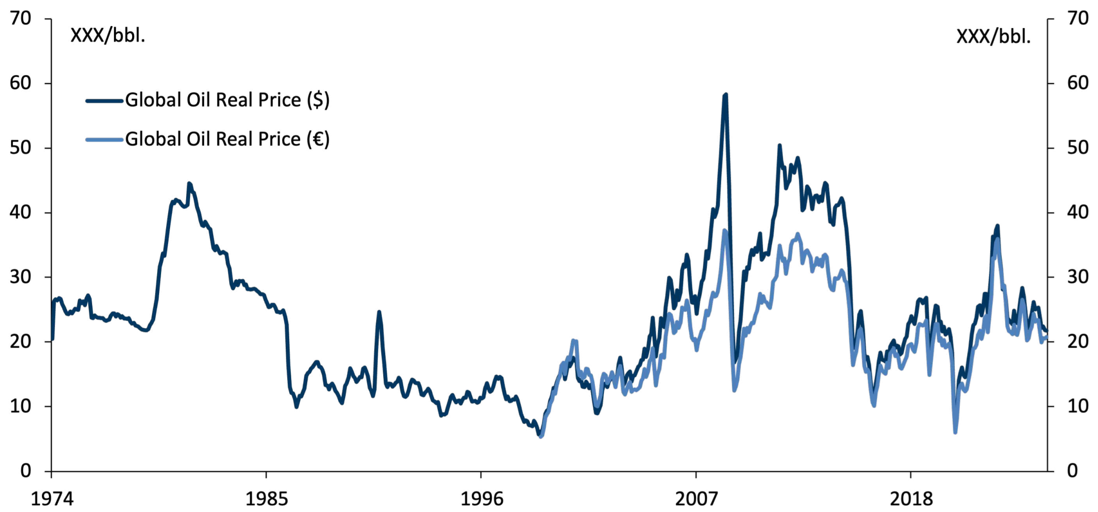

Figure 12: *Global Oil Real Price (XXX/bbl.): Inflation-adjusted cost of imported crude oil acquired by refiners.*

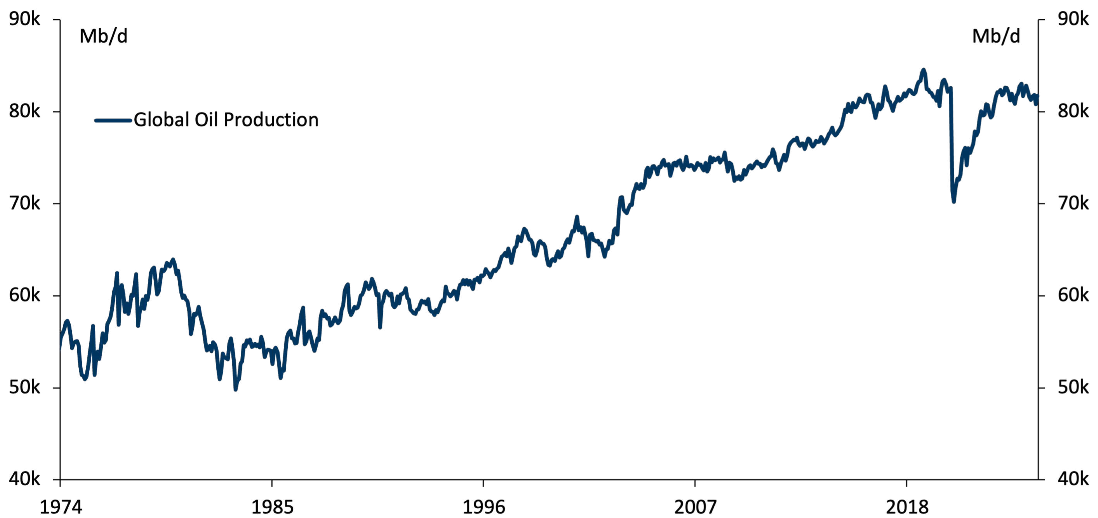

Figure 13: *Global Oil Production (Mb/d)*

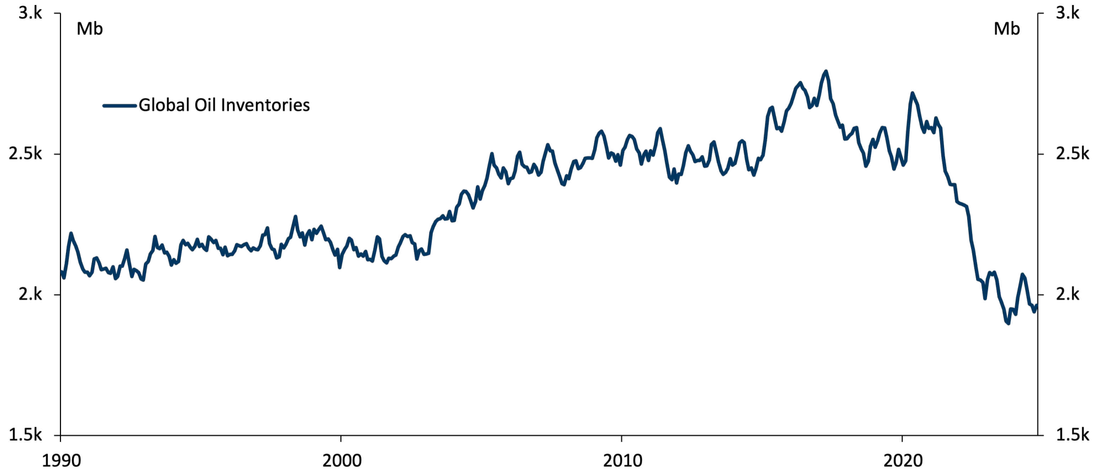

Figure 14: *Global Oil Inventories (MMb): Adjusted proxy variable based on US & OECD inventories.*

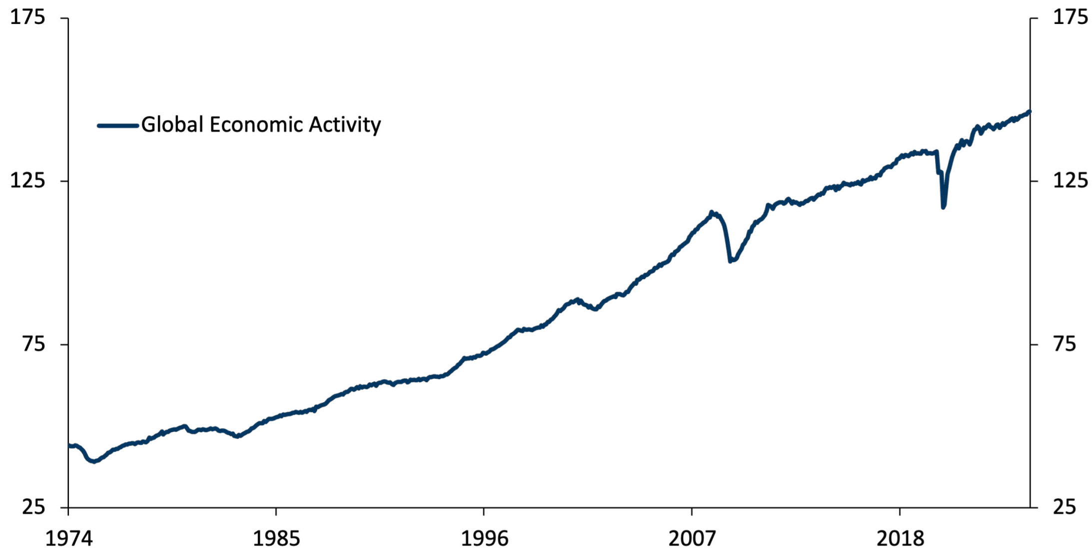

Figure 15: *Global Economic Activity*

A.2 Summary Statistics

	Global Oil Real Price	Global Oil Inventories	Global Oil Production	Global Economic Activity
Panel A: Raw				
	$/bbl	*MMb*	*Mb/d*	-
N	612	418	622	803
t	Jan-1974	Jan-1990	Jan-1973	Jan-1958
T	Dec-2024	Oct-2024	Oct-2024	Nov-2024
Mean	23.18	2336.30	66916.89	72.85
Std	10.64	206.23	9773.53	38.06
Min	5.71	1897.74	49812.05	16.69
Max	58.34	2795.14	84591.88	146.47
Skew	0.67	0.09	0.19	0.37
Kurt	2.74	1.88	1.65	1.84
Panel B: Transformed				
	f(.)	*f(.)*	*f(.)*	*f(.)*
N	611	417	621	802
t	Feb-1974	Feb-1990	Feb-1973	Feb-1958
T	Dec-2024	Oct-2024	Oct-2024	Nov-2024
Mean	0.01	-0.01	0.07	0.26
Std	8.21	1.28	1.61	0.74
Min	-52.42	-3.86	-14.46	-8.77
Max	46.67	4.20	6.50	4.51
Skew	-0.82	0.10	-2.41	-2.77
Kurt	11.45	3.15	20.69	36.27

Table 5: *Summary Statistics: Raw and transformed oil variables*

A.3 ADF Test Results

GORP	GOI	GOP	GEA
< 0.01 *	< 0.01 *	< 0.01 *	< 0.01 *

** < 0.01: H_0 is rejected with a significance level of < 0.01, indicating stationarity of the data.*

Table 6: *ADF test results: Oil variables*

	β0	β1	β2	λ
United States	< 0.01 *	< 0.01 *	< 0.01 *	< 0.01 *
China	< 0.01 *	< 0.01 *	< 0.01 *	< 0.01 *
Germany	< 0.01 *	< 0.01 *	< 0.01 *	< 0.01 *
Spain	< 0.01 *	< 0.01 *	< 0.01 *	< 0.01 *
India	< 0.01 *	< 0.01 *	< 0.01 *	< 0.01 *
Italy	< 0.01 *	< 0.01 *	< 0.01 *	< 0.01 *
Japan	< 0.01 *	< 0.01 *	< 0.01 *	< 0.01 *
South Korea	< 0.01 *	< 0.01 *	< 0.01 *	< 0.01 *
Netherlands	< 0.01 *	< 0.01 *	< 0.01 *	< 0.01 *
Singapore	< 0.01 *	< 0.01 *	< 0.01 *	< 0.01 *

** < 0.01: H0 is rejected with a significance level of < 0.01, indicating stationarity of the data.*

Table 7: *ADF test results: Nelson-Siegel modelled term structure*

A.4 Shock Event Analysis

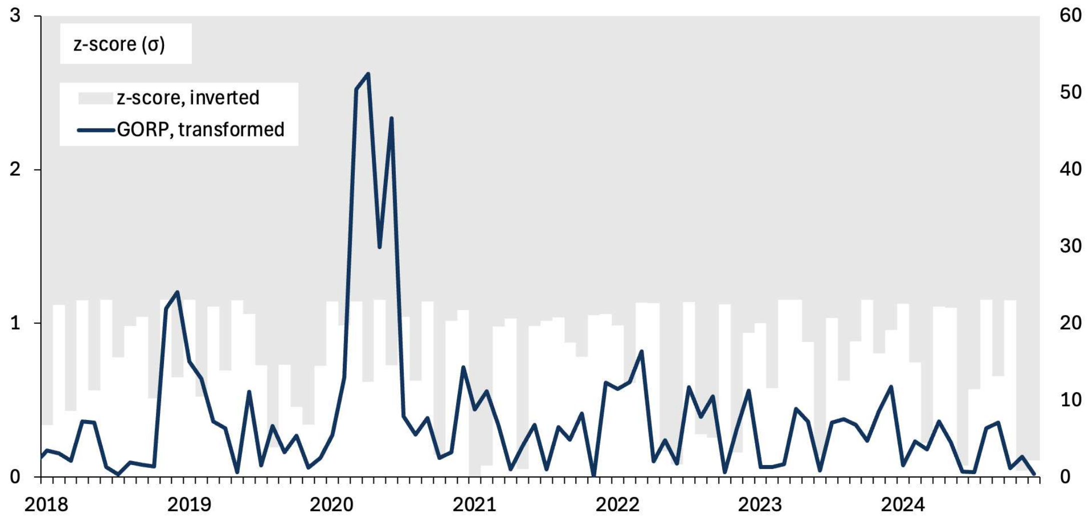

Figure 16: *Shock Event: Z-Score of Global Oil Real Price (GORP)*

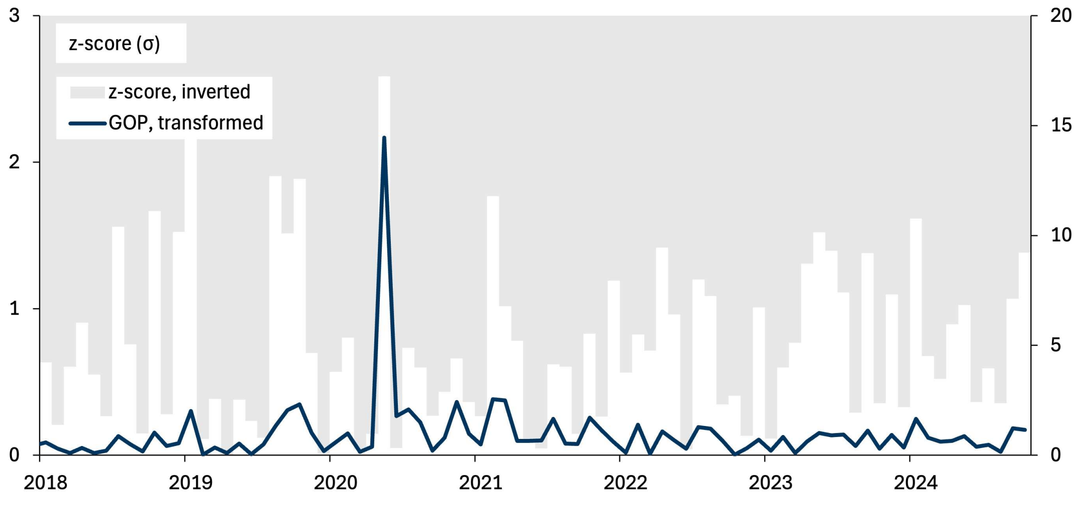

Figure 17: *Shock Event: Z-Score of Global Oil Production (GOP)*

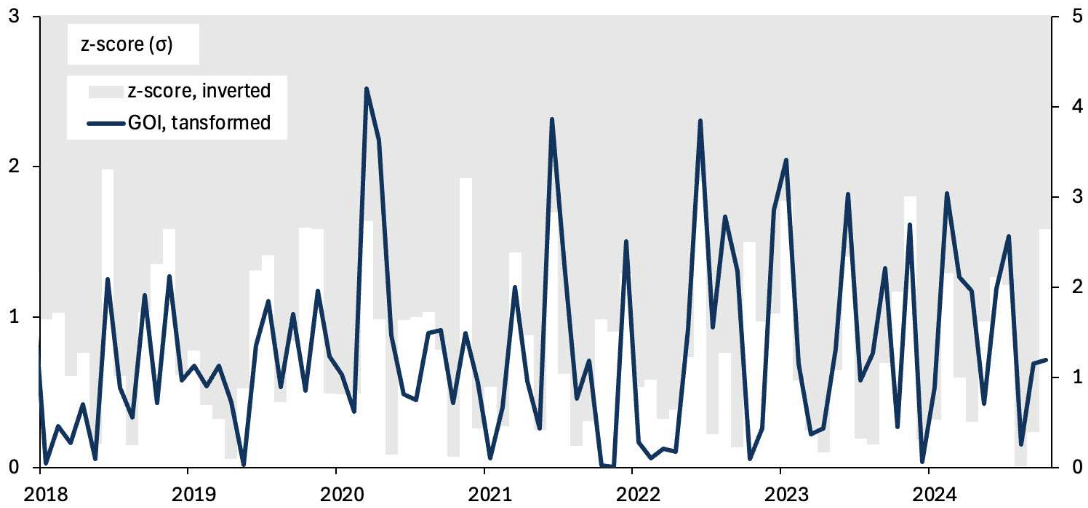

Figure 18: *Shock Event: Z-Score of Global Oil Inventories (GOI)*

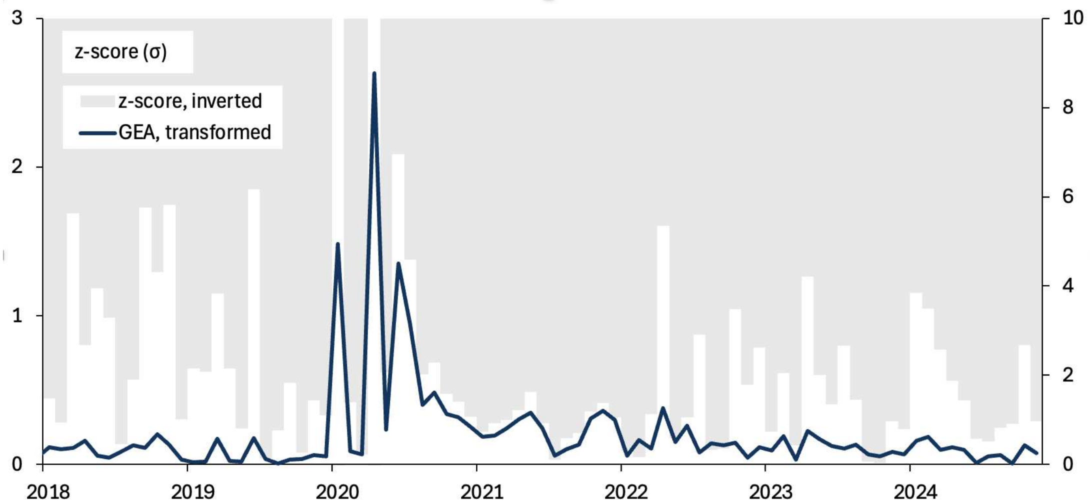

Figure 19: *Shock Event: Z-Score of Global Economic Activity (GEA)*